Susanna Nykyri

Equivalence and Translation Strategies in Multilingual Thesaurus Construction

Susanna Nykyri
Born 1970 in Turku, Finland
Master of Social Sciences 2000
at Tampere University
Researcher at the Unit of Information
Studies at the Åbo Akademi University

Photo: Meri Keränen

EQUIVALENCE AND TRANSLATION STRATEGIES IN MULTILINGUAL THESAURUS CONSTRUCTION

Equivalence and Translation Strategies in Multilingual Thesaurus Construction

Susanna Nykyri

Acknowledgements

I was introduced to the world of translation problems in the department of Translation Studies (Russian) at the Tampere University, and to the perspective of information studies and documentary languages by Mirja Iivonen, and later encouraged to write my master's thesis about equivalence issues related to multilingual thesauri by Kalervo Järvelin. Later Sami Borg and Maria Forsman from the Finnish Social Science Data Archive asked me to join a multilingual thesaurus project as a planning officer. I got generous help from Ken and Pam Miller from the United Kingdom Data Archive. I am thankful to these institutions and persons for directing me to step into this long and fascinating path of multilingual thesaurus construction!

I got a great opportunity to make my doctoral thesis in the department of Information studies at the Åbo Akademi University. The atmosphere has been warm, inspiring and fruitful. I am in gratefulness to professor Mariam Ginman and to my first supervisor Sara von Ungern-Sternberg as well as to all my colleagues at the department for making me feel so homely.

During the process I was first guided by Sara von Ungern-Sternberg, whose enthusiasm, openness and warmness meant a great deal for me. My project colleague Maria Forsman shared many stimulating coffee house moments and several nice travels, and the discussions with her strengthened the direction I had chosen. I was guided to the end of the process by professor Gunilla Widén, whose warmness, encouragement, belief in me and active feedback on drafts carried me through the heavy end. I thank also Jannica Heinström for active and stimulating comments on the first versions and all her encouragement. I have been lucky also to receive constructive feedback in the joyful research seminars at the department from my many colleagues. Their seminars have also guided me a great deal. I want to thank them also for the nice lunch hours, which have cut the day in a relaxing and strenghtening way and made the daily porridge a fun delicatesse. I also want to thank my room mate Reija Perttilä for sharing theoretical as well as practical issues.

I am grateful to the 29 respondents who took time and effort to participate in the research, and made it also fun to collect the research material. I am also grateful for encouragement at the early stage of the process and comments received for the research plan from Jaana Vuori, Harriet Silius and Inger Pietiläinen.

I got valuable comments and a constructive push to the final stage from my previewers Jaana Kekäläinen and Marianne Lykke Nielsen. Elizabeth Nyman kindly and efficiently checked the language.

The research would not have been possible without financing that provided me full-time engagement and also participation in several inspiring workshops and conferences. For that I am thankful to the Academy of Finland and on the targeted programme on the Production, Management and Use of Digital Information Resources,

CONTENT

ABSTRACT

In the present dissertation, multilingual thesauri were approached as cultural products and the focus was twofold: On the empirical level the focus was placed on the translatability of certain British-English social science indexing terms into the Finnish language and culture at a concept, a term and an indexing term level. On the theoretical level the focus was placed on the aim of translation and on the concept of equivalence. In accordance with modern communicative and dynamic translation theories the interest was on the human dimension. The study is qualitative.

In this study, equivalence was understood in a similar way to how dynamic, functional equivalence is commonly understood in translation studies. Translating was seen as a decision-making process, where a translator often has different kinds of possibilities to choose in order to fulfil the function of the translation. Accordingly, and as a starting point for the construction of the empirical part, the function of the source text was considered to be the same or similar to the function of the target text, that is, a functional thesaurus both in source and target context. Further, the study approached the challenges of multilingual thesaurus construction from the perspectives of semantics and pragmatics. In semantic analysis the focus was on what the words conventionally mean and in pragmatics on the 'invisible' meaning - or how we recognise what is meant even when it is not actually said (or written).

Languages and ideas expressed by languages are created mainly in accordance with expressional needs of the surrounding culture and thesauri were considered to reflect several subcultures and consequently the discourses which represent them. The research material consisted of different kinds of potential discourses: dictionaries, database records, and thesauri, Finnish versus British social science researches, Finnish versus British indexers, simulated indexing tasks with five articles and Finnish versus British thesaurus constructors. In practice, the professional background of the two last mentioned groups was rather similar. It became even more clear that all the material types had their own characteristics, although naturally not entirely separate from each other. It is further noteworthy that the different types and origins of research material were not used to represent true comparison pairs, and that the aim of triangulation of methods and material was to gain a holistic view.

The general research questions were:

1. Can differences be found between Finnish and British discourses regarding family roles as thesaurus terms, and if so, what kinds of differences and which are the implications for multilingual thesaurus construction?

2. What is the pragmatic indexing term equivalence?

The first question studied how the same topic (family roles) was represented in different contexts and by different users, and further focused on how the possible differences were handled in multilingual thesaurus construction. The second question was based on findings of the previous one, and answered to the final question as to what

kinds of factors should be considered when defining translation equivalence in multilingual thesaurus construction.

The study used multiple cases and several data collection and analysis methods aiming at theoretical replication and complementarity. The empirical material and analysis consisted of focused interviews (with Finnish and British social scientists, thesaurus constructors and indexers), simulated indexing tasks with Finnish and British indexers, semantic component analysis of dictionary definitions and translations, co-word analysis and datasets retrieved in databases, and discourse analysis of thesauri. As a terminological starting point a topic and case family roles was selected.

The results were clear:

1) It was possible to identify different discourses. There also existed sub-discourses. For example within the group of social scientists the orientation to qualitative versus quantitative research had an impact on the way they reacted to the studied words and discourses, and indexers placed more emphasis on the information seekers whereas thesaurus constructors approached the construction problems from a more material based solution. The differences between the different specialist groups i.e. the social scientists, the indexers and the thesaurus constructors were often greater than between the different geo-cultural groups i.e. Finnish versus British. The differences occurred as a result of different translation aims, diverging expectations for multilingual thesauri and variety of practices. For multilingual thesaurus construction this means severe challenges. The clearly ambiguous concept of multilingual thesaurus as well as different construction and translation strategies should be considered more precisely in order to shed light on focus and equivalence types, which are clearly not self-evident. The research also revealed the close connection between the aims of multilingual thesauri and the pragmatic indexing term equivalence.

2) The pragmatic indexing term equivalence is very much context-depended. Although thesaurus term equivalence is defined and standardised in the field of library and information science (LIS), it is not understood in one established way and the current LIS tools are inadequate to provide enough analytical tools for both constructing and studying different kinds of multilingual thesauri as well as their indexing term equivalence. The tools provided in translation science were more practical and theoretical, and especially the division of different meanings of a word provided a useful tool in analysing the pragmatic equivalence, which often differs from the ideal model represented in thesaurus construction literature.

The study thus showed that the variety of different discourses should be acknowledged, there is a need for operationalisation of new types of multilingual thesauri, and the factors influencing pragmatic indexing term equivalence should be discussed more precisely than is traditionally done.

1 Introduction

It's a common belief that the main challenge in cross-cultural communication[1] is to know the foreign culture, when in fact it is to know our own culture and how it effects our behaviour and thinking. (Adler 1997, 78-87)

In this chapter the aim and perspective of the study are introduced. First the background is illustrated, second the aims of the study are discussed, and third the structure of the study is described.

1.1 Background

We are living in an information society. Our everyday language includes expressions such as information flow, information over-flow, information warfare, globalisation, knowledge management, time management, efficiency, cost-effectiveness… To convey meaning we need to communicate, use language, and this is not always an easy task in a globally united yet linguistically and culturally separated world. At a societal level having information is not only power, but also a problem. - How should we store and find information? How should we separate relevant information from non-relevant information? Libraries and tools for information documentation and retrieval are not only living a new renaissance, but also encountering new problems. How should we provide relevant information in today's global life? What are the obstacles? How do we face them?

Languages, and the thinking they reflect, stem mainly from the cultural needs for expression. A controlled vocabulary, a thesaurus, can be seen as a cultural product. Global knowledge production presupposes communication across borders, cultures and languages between people. With the emergence of the Internet it was believed that content analysis and controlled vocabularies would become redundant because information could be retrieved on the net using simple word searches. It has become apparent, though, that the more digital libraries, link collections and portals are developed on the net, the more important content analysis and developing tools have become in order to aid these searches (Mulvany 1997; Forsman 2000; Hert & Jacob & Dawson 2000).

Each information search in a database covers at least five different ways to communicate: the authors, the indexers, the synthetic structure, the users and the search strategy (Buckland 1999), which all represent a type of discourse. An indexer's or a specialist's methods of expressing their ideas and thoughts on a certain social environment differ from each other and in indexing and in information retrieval this can cause problems (see also Bates 1986; Iivonen 1995).

---

[1] Author note: According to Fiske (1992, 14) the general definition for communication is 'social interaction that happens via transmission of messages'.

When analysing different discourses it is essential to be aware of the context in which the discourses take place. In this study, cultural context is understood accordingly to Nida (1975, 229) as "the part of the context which includes both the total culture within which a communication takes place and the specific nonlinguistic circumstances of the communication". At a general level, culture can be defined to refer to the customs, beliefs, and ways of life of a group of people. Culture has several subcultures such as those of indexers, thesaurus constructors and of social scientists, which are subsets of the main cultural group. These are the subsets studied.

The linguistic-cultural context studied is Finnish and British-English. English is *de facto* today the leading *lingua franca* (Chesterman 2003, 120), whereas Finnish is culturally very Finnish. English is also the working language of knowledge organizations such as CESSDA (Council of European Social Science Data Archives[2]) and ISKO (International Society for Knowledge Organization[3]). A lingua franca does not eliminate the problem of different discourses – the problem exists already within one language and culture area.

Discourses are in a constant flux. The rapid development of modern societies changes language and terminologies in the social sciences. Together with a widening globalisation, this increases the need for high quality documentation tools for locating up to date information from multilingual resources. As Jorna and Davies (2001) have stated, the most current forms of multilingual information access are inadequate to answer the needs of increasingly diverse user groups from different cultural and linguistic backgrounds, and a new form of multilingual thesaurus is therefore required.

Information seeking and retrieval across national borders are constantly on the rise. The success of creating and using international information resources still depends on common tools and on an understanding of the concepts used. Along with an expanding use of international databases, cross-cultural indexing and information seeking are becoming more common and tools like multilingual thesauri are needed. Not only have cross-cultural research projects, information resources and indexing become common, but cross-disciplinary research is also becoming more common. It has its own challenges from the viewpoint of knowledge management and information seeking. Equivalence and its different levels are the most important relationship to study in the comparative semantics and terminological studies (Nuopponen 2003, 78), and it is the case also in multilingual thesaurus construction, since the whole thesaurus, to be used operating with terms, is based on the nature of the adopted idea of equivalence.

The theme of this research – multilingual and multicultural thesaurus construction - is very topical because today many multilingual terminologies and thesauri are being developed, e.g. Finnish Library of Parliament Subject Headings (see also Landry 2004). The problems in their construction have not been studied much. Translation studies as a science is also very young and multicultural, and cross-cultural communication as a research topic has become more common in Finland only since the 1980s.

---

[2] About CESSDA see its homepage, URL: <http://www.nsd.uib.no/Cessda/>

[3] About ISKO see its homepage, URL: <http://www.isko.org/>

When considering multilingual and/or multicultural thesaurus construction the heart of the theoretical context lies in indexing, in information seeking, in the relation between language and culture and in equivalence. The theoretical framework is therefore, in its nature, cross-disciplinary and especially the theories from translation science are extended in order to include the new area of application, thesaurus construction. A reader of the study should be aware of the fact that translation theories are often formed on the basis of literature, which as a translation unit differs greatly from thesauri.

Within library and information science (LIS) Iivonen (1989) has shown how consistency is likely to be greater in the conceptual analysis step and less in the term selection step of indexing. When discussing the consistency of indexing it is also important to keep in mind that the aim of the documentary languages, e.g. thesauri, is to increase predictability, not consistency. Even bad indexing may be highly consistent. (Ibid.12; Fugmann 1994, 420-421)

The ISO standard 5964–1985 (Documentation -- Guidelines for the establishment and development of multilingual thesauri) states that the makers of a multilingual thesaurus face two additional kinds of problems which do not occur in monolingual work. The first types are related to management problems – decisions of the status of languages (exchange, dominant or secondary language(s)), and the second types to language problems – choice of an appropriate procedure when a term in one language expresses a concept which cannot be represented by an exactly equivalent term in the other language(s). (Ibid, 5)

According to Hudón (1997) the problems traditionally associated with multilingual thesauri are:

> "1) that of stretching a language to make it fit a foreign conceptual structure to the point where it becomes barely recognizable to its own speakers;
>
> 2) that of transferring a whole conceptual structure from one culture to another whether it is appropriate or not;
>
> 3) that of translating literally terms from the source language into meaningless expressions in the target language, etc." (Ibid.)

One way to create a multilingual thesaurus is translating an existing monolingual thesaurus (see ISO 5964-1985). Hudón (1997) points out that a monolingual thesaurus is always culturally biased, and a straight translation might lead to a form of "cultural imperialism".

In addition, Doerr (2001) states that even though semantic heterogeneity of terminological resources has frequently been referred to as a problem, a systematic analysis of its intellectual basis and structure has not been carried out. In his opinion translated thesauri are thesauri, "--- where each concept is optimally interpreted in words or multiple languages, to allow speakers of those languages to understand better and use concepts of this thesaurus more effectively." He also stresses that such translations are in general not established indexing terms of the target language (ibid.).

The art of translation meant for Horatius and Cicero careful interpretation of the source text and composition of source text according to the principle *non verbum de verbo, sed sensum exprime de sensu* (not word for word, but thought for thought). This means that a translator is always responsible to the target audience. (Bassnet 1995, 62)

This study concentrates on the examination of the function of the translation in thesaurus construction, and of the equivalence type which is driven by the focus. Therefore, the divisions of translation equivalents, which are based on the formal equivalence, are left outside the scope of the study. The functional, communicative translation theories are 1) Dynamic Equivalence, 2) Functional Equivalence and 3) Skopos-theory. In the study their influence can be seen in the perspective, hypotheses and in the research questions.

Research on multilingual thesauri is useful for all the aspects of production, management, and seeking of digital information resources. There is a need for basic research, which is based on theory and empirical findings. As in multilingual thesaurus construction so also in research of multilingual thesaurus construction the challenges exist on many levels – both between cultures and subcultures, languages, discourses and domains or disciplines.

1.2 Objectives of the dissertation

This study is a part of a knowledge organization (KO) approach in the field of Information Studies / Library and Information Science and in it part of a cultural and social issues in KO (López-Huertas 2008, 119).

The focus of this study is twofold:

- the translatability of British-English social science indexing terms into the Finnish language and culture at a concept, term and indexing term level, and;
- the aim of translation.

The interest is on the human dimension and translating is seen as a decision-making process.

This research aims at defining the nature of multicultural and multilingual terminology work and clarifying traditional standards and guidebooks for multilingual thesaurus construction and multicultural knowledge management – in general, to improve our understanding of translation as a significant activity in today's intercultural and interlingual global information society. In order to make a deeper analysis and to gain more detailed information about the phenomenon, the study is in its nature a qualitative case study.

A cross-domain approach will bring new perspectives to the research field and help to develop the research analysis method by borrowing expressive and useful concepts and tools especially from translation and communication studies. The project has also a

novelty value in describing a small language area's culture- and language-bound problems in the development and study of multilingual indexing tools and vocabularies.

The emphasis is on the Finnish language and indexing practices. The perspectives are linguistic and socio-cultural – a combination through which a broader understanding of the phenomena is being aimed at within the general framework of information science.

Edward T. Hall (1981) has stated, that

> "Years of study have convinced me that the real job is not to understand foreign culture but to understand our own. I am also convinced that all that one ever gets from studying foreign culture is a token understanding. The ultimate reason for such study is to learn more about how one's own system works. The best reason for exposing oneself to foreign ways is to generate a sense of vitality and awareness – an interest in life which can come only when one lives through the shock of contrast and difference." (Ibid. 1981, 30)

In the empirical part the aim is to identify different discourses and vocabularies, as well as ideas and principles in the focus of translation, existing in a particular information domain and to see how they are considered in information storage and in multilingual and multicultural thesaurus construction. The study is thus an empirical case study. The more theoretical goal is to illustrate "pragmatic indexing (term) equivalence".

The study uses multiple cases and several data collection and analysis methods aiming at theoretical replication and complementarity. The main data collection method is a series of focused interviews and the main analysis methods are discourse analysis and bibliometric co-word analysis.

As a terminological starting point a topic and case *family roles* is selected. It is considered to have a broader significance, first, since it represents a social categorisation, which is often difficult to translate into another culture and language and is considered to be a typical translation problem, and second, since it has a wider significance within social sciences (information about family roles are gathered often in sociological surveys as a background variable), and third, because the topic is part of everyday life and everybody has some kind of experience about it.

The samples studied are theoretical (not random) and thus they are assumed to give answers to the questions of this study. As a starting point and material the working version of ELSST (*European Language Social Science Thesaurus*) is used. (About ELSST see UKDA 2001; Forsman & Keränen 2002; Jääskeläinen & Forsman 2003; Jääskeläinen 2006.)

The study is based on the following general assumptions:

- Intercultural misunderstandings are often caused by

- Subconscious blinders. People are not aware of their own assumptions and their cultural basis.

- The lack of cultural self-awareness. It is a common belief that the main challenge in cross-cultural communication is to know the foreign culture, when in fact it is knowing our own culture and how it affects our behaviour and thinking that is more important.

- Projected similarity. In real life people differ more from each other than is assumed. Another person's situation also differs more than is usually assumed. Differences can then be expected, imagined and discovered as similarities.

- Cross-cultural misevaluation. We use our own culture as a standard of measurement.

(Hall 1981, 20-33; Adler 1997, 78-87. See also Keränen 2001, 83)

- Languages and ideas expressed by languages are created mainly in accordance with the expression needs of the surrounding culture (Suojanen 1993) and thus, a thesaurus can be seen as "a cultural product".

- The content of a thesaurus depends on the documentary needs of the data sources in the surrounding culture. Indexing is not necessarily consistent in one and the same culture. (Iivonen 1989)

- Finnish indexers are guided by the Finnish culture and its traditions. In Finland, the most important and widely used indexing term source is the General Finnish Thesaurus (Yleinen suomalainen asiasanasto, YSA[4]). Indexers usually find the Finnish equivalents for English terms in YSA and thus conform to the Finnish culture and the conventions in information retrieval and storage.

- Translation problems of concepts occur, for example, on the basis of the thesaurus' (hierarchical) structure, the differences between languages and cultures, the time factors which influence the content of concepts.[5]

- Indexing terms are part of a documentary language and because of this they can sometimes greatly differ from other discourses of natural language. For example, social scientists' and indexers' equivalents may differ from each other. When selecting an indexing term indexers are more likely to take into consideration the standards' and guidebooks' requirements for a good indexing term.

---

[4] General Finnish Thesaurus (Yleinen suomalainen asiasanasto, YSA) is a monocultural (Finnish) tool, which covers all the sciences and includes the general terminology of different fields. One of the main purposes of YSA is to be a general source vocabulary when developing special vocabularies. YSA is commonly used in Finland in public and scientific libraries and data archives in indexing and information seeking. YSA is included in VESA, which includes also special thesaurus for music and Swedish language equivalents. (Helsingin yliopiston kirjasto 1999)

5 There occurs time delay also in natural languages. A language does not change as fast as the surrounding culture, which adopts new things, concepts and phenomena e.g. from international culture. (See Suojanen 1993)

- Indexing term equivalence and concept equivalence are not necessarily the same. We can have equivalence between a Finnish and a British concept, but at the indexing term level equivalence may differ and vice versa.

Explanations and datasets are compared by **classification** of the terms and translations into different groups. Comparisons inside and between these groups are made due to e.g. their type of equivalence, nativeness in the Finnish language and culture-boundness. Problems can be caused by, for example

1 Society

The same kind of phenomenon or concept used in British society does not exist in the Finnish society. (*'homemaker'*)

2 Language

The concept exists in both cultures, but to find a term to describe it conforming to Finnish common language usage is problematic. (*'illegitimate births'*)

3 Euphemisms

Socially difficult concepts are described with euphemisms. (*'family disorganization'*)

4 Institutional differences

To express the concept conforming to Finnish indexing language practices a thesaurus constructor has to use factoring. (*'married women workers'*)

5 Citation loans, foreign words

There is not a Finnish native word to describe the concept and/or at the term level a citation loan or foreign word is commonly established (*'nationalism'*).

According to Hjørland (2002, 429) research on indexing, document representation and retrieval should be able to evaluate bad practices and improve them. This research is in its nature also evaluative – it evaluates certain practices and ideas and discusses their consequences.

1.3 Research questions

The general research questions are:

1) Can differences be found between Finnish and British discourses regarding family roles as thesaurus terms, and if so, what kinds of differences and which are the implications for multilingual thesaurus construction?

2) What is the pragmatic indexing term equivalence?

The first question studies how the same topic (family roles) is represented in different contexts and by different users, and further focuses on how the possible differences are handled in multilingual thesaurus construction. The second question is based on findings of the previous one, and answers to the final question as to what kinds of factors should be considered when defining translation equivalence in multilingual thesaurus construction.

The general research questions are operationalised into several sub-questions, such as:

- How are the studied concepts understood? To what extent are the differences due to institutional versus cultural differences? What is the semantic invariant?

- What are the studied terms about according to indexing and thesauri? How are the studied concepts used in indexing and why are they used in this way?

- How is equivalence understood? What do thesaurus constructors, indexers and social scientists aim at in their translations? Do the potential thesaurus users share same vision as thesaurus constructors?

The emphasis is on human effort and on the Finnish language and culture. What the social context of the studied concepts is in the source and in the target culture, is a background issue.

The empirical material and analysis consists of focused interviews (with Finnish and British social scientists, thesaurus constructors and indexers), simulated indexing tasks with Finnish and British indexers, semantic component analysis of dictionary definitions and translations, co-word analysis and datasets retrieved in databases, and discourse analysis of thesauri. (For more details see chapter 5 *The study* and 6 *Methods and material*.)

The language of the thesis is English, but the central methodological and theoretical terms as well as terminological examples are represented also with Finnish equivalents. The Finnish material is translated so that the reader of the thesis is not required to know Finnish. (The Finnish-Swedish study environment can be seen in some examples.)

1.4 Structure of the dissertation

This study is structured as follows. After the introduction the theoretical background is given in chapters 2-3. After the theoretical framework, in chapter 4, the linguistic and social background of the study is briefly illustrated, after which follow the research design and the methods and material used in this study. The empirical part is presented in chapters 7-9. Chapter 10 includes the discussion and conclusions of the study.

Chapter 2 presents the central concepts of the study and in doing so also introduces more closely the perspective of the study, which is pragmatic and sociolinguistic.

In chapter 3, the previous most relevant research related to language, culture and meaning (chapter 3.1), indexing (3.2), thesaurus construction (3.3) and translatability and equivalence (3.4) are discussed.

In chapter 4, some key elements of the linguistic and social context of the study are illustrated. This is not part of the theoretical framework, but (in accordance with the discourse analytical approach) provides contextual information to the research design of the study.

In chapter 5, the design of the empirical part of the study is illustrated. In chapter 6, the methods and material used in this study are presented. Results of the empirical study are reported in chapters 7-9 and finally chapter 10 summarises the study together with reflections and ideas for future studies.

2 The central concepts

Context may – depending on context – mean so many different factors, that bananas get terrified.[6] (Lehtonen 2000, 164)

Culture is a word that has so many meanings already that one more can do it no harm. (Hall 1981, 20)

It is initially very necessary to define the central terminology used in this study especially from the perspective of the research problem. Many of the concepts are very broad and there are already several alternative definitions within one discipline. Culture and equivalence are especially complicated ones. In this chapter, the aim is twofold; firstly it is to clarify how the central concepts are understood and used in this study and in pursuance of this objective to also introduce the perspective of the study. The central concepts are related to language and culture; indexing and thesaurus construction; and translation and equivalence.

Therefore, it is initially necessary to explain the definitions – or rather, to say what the definitions should be like.

A **definition** is a linguistic description of the concept. It outlines the extension of the concept and its objects. (TSK 1999) A good definition is written in clear and easily understandable language. It is linguistically irreproachable and conforms to practices of the common language and the special field concerned. In order for a definition to fulfil its purpose, one must consider the target group, and because of this a definition should only use words known from the common language, words that are obvious for its target group and/or special terms defined separately in the same vocabulary. (TSK 1989, 41) A definition can also be viewed as a scientific hypothesis:

> "A linguistic definition is a scientific hypothesis about the concept encoded in a given word. (see Robinson 1950: 41). Like other scientific hypothesis, it cannot be proved to be right, but it can be tested and proved wrong – in which case it is discarded, or revised and tested again. While the concept is not accessible to direct observation, it is manifested in a word's use. Accordingly, a definition can be tested against a word's range of use; this range of use may be very broad, but it has its boundaries, which are determined by the different components of the concept.
>
> The components of the concept determine which aspects of a word's use are variable and which are invariable. It is the purpose of a definition to capture the invariable aspects of a word's use, that is, its semantic invariant." (Wierzbicka 1996, 239)

According to Wierzbicka (1996, 240-241), dictionary descriptions often fail to capture a word's semantic invariant and instead of definitions they often provide rather random lists of quasi-synonyms. Such entries inevitably lead to vicious circles.

---

[6] (Original: "Konteksti voi – kontekstista riippuen – merkitä niin monenlaisia tekijöitä että heikkopäistä hirvittää." (Lehtonen 2000, 164)

Wierzbicka also criticises the use of words such as "or" and "usually" in definitions and considers them as unmistakable signs of the lexicographer's failure to find a semantic invariant, and mentions also a proliferation of different senses and subsenses in a dictionary entry as such. (Ibid, 241)

As stated in section 1.2 above, the study is a part of a **knowledge organisation** (KO) approach. According to Hjørland (2008), KO is broadly understood to be about the social division of mental labour, e.g. the organisation of universities and alike for research and higher education, the structure of disciplines and professions and the social organisation of media. In the narrow sense it is about activities such as document description, indexing and classification. As a field of study, it is concerned with the nature and quality of such knowledge organization processes (KOP) as well as the knowledge organising systems (KOS) used to organise documents, document representations, works and concepts. Library and Information Science (LIS) is the central discipline of KO in this narrow sense. (Hjørland 2008, 86)

2.1 Language and culture

At a general level, **culture** can be defined as referring to the customs, beliefs, and ways of life of a group of people. Culture has several **subcultures**, which are subsets of the main cultural group. Trompenaars (1993) compares culture with an onion, consisting of different layers:
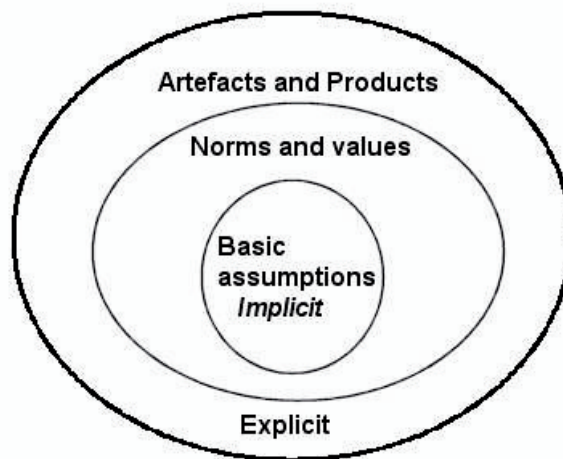
Figure 1: *A model of culture (Trompenaars 1993, 23)*

The outer level is the most visible one, 'explicit', and consists of explicit products. Trompenaars describes explicit culture as the observable and symbolic reality of the language, food, buildings, houses, monuments, agriculture, shrines, markets, fashions and aret. He also considers them as the symbols of a deeper level of culture. Prejudices mostly start on this level. (Ibid, 22) Trompenaars (ibid, 22) also points out, that "each

opinion we voice regarding explicit culture usually says more about where **we** come from than about the community we are judging".

The middle layer includes norms and values. The norms concern - and also to a large extent dictate – how one should behave in society. There may be a considerable gap between values and norms. Values are aspirations, which may never be actually achieved. The core layer is 'implicit', not visible. It is the heart of culture and it includes basic assumptions about life, which will have been handed down unconsciously from generation to generation. (Ibid.)

Tompenaar's idea of culture is similar to the *Iceberg Theory*. Hall (1981) notes, that the concept expresses how the most important part of the culture is completely hidden, and what can be seen is 'just the tip of the iceberg'. (Katan 1999, 29, see more in chapter 3.1 *On culture, language and meaning*.)

In this research **culture** refers to a **conceptual** level and can thus refer to

> 1. a (geographical) culture
>
> Culture can then be - on a very general level - defined as a framework for our lives, something that affects our values, attitudes and behaviour. Attention is paid to language and communication styles as a dimension of cultural differences between Finnish and British culture.
>
> 2. a subculture
>
> Different groups, for instance, Finnish and British indexers versus Finnish and British social scientists. Attention is paid to institutional differences.

In this study **cultural context** is understood according to Nida (1975, 229) as "the part of the context which includes both the total culture within which a communication takes place and the specific nonlinguistic circumstances of the communication". When analysing different discourses it is essential to be aware of the context in which the discourses take place. For instance, guidelines for thesaurus constructors presumably differ from the social scientists' language usage guidelines due to their function, environment and conditions of use.

Language is a set of characters, conventions, and rules, which is used for conveying information. **Linguistics** is the study of language (Hudson 1998, 4). The three relevant aspects of language for this study are pragmatics, semantics, and syntax. In this study, the two first aspects mentioned are particularly in focus - and the distinction between pragmatic and semantic focus is essential. (See also chapter 3.1 *On culture, language and meaning*.)

Language thus refers to an **expressional** (verbal) level. **Linguistic** differences can thus occur between different

> 1. **Languages**

Finnish versus British English. Their characteristic problems differ from each other. (See more in Chapter 4.1.2 *On Finnish and English*).

When speaking about languages the focus is on natural languages if not otherwise specified. **Natural language** is a "language which has developed through the normal processes as the speech form of a community of people; opposed to artificial language" (Nida 1975 231). (About documentary languages see the subchapter 2.2 *Indexing*.)

2. **Discourses**

At a general level discourse can be defined as referring to language use in social context (Pälli 2003, 22). It is also noteworthy to realise, that although discourse is a substantive, it also means doing something and is therefore an active process (Lehtonen 1994; Potter & Wetherell 1990; Pälli 2003).

Each information search in a database covers at least five different languages: the authors, the indexers, the synthetic structure, the users and the search strategy (Buckland, 1999), which all represent a type of **discourse**. An indexer's or a specialist's ways of expressing their ideas and thoughts on a certain social environment differ from each other and in indexing this can cause problems. For example, when speaking about lone mothers, a specialist may use a precise term (*single mother/yksinhuoltajaäiti*), a politician an eloquent term (*re-miss, mother/uusioneiti, äiti*)[7], a journalist an eye-catching one (*single moms/sinkkuäidit*) and an indexer a term accordingly to the thesaurus practices (*mothers --- divorced/ äidit --- eronneet*).

Tonkiss (1998) points out, that the important thing to think about here is not so much what sorts of language 'count' as discourse, but how discourse analysts approach language as data. Within discourse analysis **language** is viewed as the **topic** of the research. (Ibid, 247) (About discourses see also chapter 6.2.1 *Discourse analysis*.)

Different discourses and subcultures may also represent different **domains** (– or vice versa, different domains can be seen as different subcultures having their own discourse). Hjørland (2004, cited here Hjørland & Nicolaisen 2005) clarifies, that a "domain may be a scientific discipline or a scholarly field. It may also be a discourse community connected to a political party, a religion, a trade or a hobby". (Ibid.)

Hjørland (2002, 422) provides eleven approaches to domain analysis in information science: producing literature guides and subject gateways; producing special classifications and thesauri; research on indexing and retrieving specialities; empirical user studies; bibliometrical studies; historical studies; LSP (language for special purposes), discourse studies; studies of structures and institutions in scientific communication; and domain analysis in professional cognition and artificial intelligence.

When focusing on multilingual thesaurus construction and/or thesaurus research, languages can be divided into different types according to their role in thesaurus construction. The central concepts, from the perspective of multilingualism and

---

[7] In 2003, the Minister of Culture, Tanja Karpela, represented herself in her homepage as "re-miss" and a mother of two children. (Original Finnish version: "32-vuotias "uusioneiti" ja kahden lapsen äiti".) (Situation in 17.4.2003) URL: <http://www.tanjakarpela.net/kuka.html>

translatability issues, are then the source language, the target language, the exchange language and the dominant language.

According to ISO 5964 -1985 a **source language (SL)** is "that language which serves as a starting point when a preferred term is translated into its nearest equivalent term or terms in a second (or target) culture." **Target language (TL)** refers to the language into which a term first encountered in another language is translated. (Ibid, 3)

In this study, the **exchange language** is also understood accordingly to ISO 5964-1985 as a language used as a medium. Concerning its usage it can be stated:

> "The language used as medium for data exchange in those multilingual systems which, as a matter of policy, decide to use terms selected from only one language for this purpose. The exchange language may also be used for indexing and/or retrieval, and the multilingual thesaurus then functions principally as a means for translating the local languages of indexers and enquirers into, or out of, the exchange language. The different languages in such a system would still be recognized as having equal status (- - -) if equivalents are established reciprocally between the preferred terms in the exchange language and the preferred terms in all other languages." (Ibid, 2)

Furthermore, according to ISO 5964-1985 the **dominant language** represents a specific type of exchange language:

> "An exchange language (- - -) which is also used for indexing and retrieval in systems which, for policy reasons, do not give equal status to all the languages in the system. Every concept recognized in the system must necessarily be represented by a preferred term in the dominant language. In some cases, however, an equivalent expression may be lacking in one more of the other languages. These other languages are then known as *secondary languages*." (Ibid, 2)

As already indicated, there are thus different types of languages and they can be approached from many perspectives. According to Yule (1996, 114) **semantics** "is the study of the meaning of words, phrases and sentences. In semantic analysis, there is always an attempt to focus on what the words conventionally mean, rather than what a speaker might want the words to mean in a particular occasion." Hudson (1998, 176) defines it as "the study of meaning, especially those parts of meaning which do not derive from the circumstances in which words are used".

In addition, the study of 'intended speaker meaning', is called **pragmatics** (Yule 1996, 127).

> "In many ways, pragmatics is the study of 'invisible' meaning, or how we recognize what is meant even when it isn't actually said (or written). In order to that to happen, speakers (and writers) must be able to depend on a lot of shared assumptions and expectations. The investigation of those assumptions and expectations provides us with some insights into how more gets communicated than is said." (Ibid, 127)

Hudson (1998, 4) states that the parts of meaning that come directly from the meanings of the words and constructions themselves are the province of semantics, whereas the parts which come from the context in which the sentence is uttered are handled by pragmatics. He (ibid, 175, 176) further defines **pragmatics** as the study of language in relation to the way it is used, especially when the study focuses on meaning,

and **semantics** as the study of meaning, especially those parts of meaning which do not derive from the circumstances in which words are used. Concerning the roots of semiotics Slater (1998) states, that

> "Semiotics starts from Saussure's structural linguistics - - - . As against earlier approaches to language, Saussure argued that meaning should be studied as a system of **signs**. Words do not derive their meaning either from the psychological intentions of individual speakers, or from the things the words describe. Rather their meanings arise from their place in a system of signs and their relations of difference or sequence with other terms in system. The meaning of a word 'cat' does not arise as an expression of an individual's thoughts or as a label for an object (there is nothing about four-legged animals that miaow that produces the word 'cat' in English) but rather arises from the relation of the word to others such as 'dog', 'lion', and so on. Different languages divide up the same object world in different ways, using different systems of terms (*decoupage*). These different systems of signs are neither natural objects nor intentional acts but rather *social facts* - - - " (Ibid, 238)

When the focus is especially placed on pragmatics, the **context** becomes more acutely relevant. According to Lehtonen (1998, 158-176) texts are not only products of a certain context, but also active producers and co-builders of context. As a context of a text may act, for example, some global societal structures like World War or co-texts of the texts like *Blade Runner* movie directed by Ridley Scott (ibid, 164). Discourses decode the limit between a text and a context. The border between these two is not ready-made, but becomes shaped in the negotiation between the text and its readers. The border between text and context is thus in constant flux and unstable and in addition functions on several levels. Because of this the definition of this border cannot serve as a starting point for a study, but the definition can be achieved during the research process. Texts are always produced by a certain context, but they also reshape their contexts. (Ibid, 166) Hjørland (1997, 17) discusses what is a 'document' in LIS and in that context states, that documents contain "texts" (in a generalised meaning, including also pictures) written by other people, and that the study of them must build on some kind of theory of communication and signs, i.e. semiotics.

Yule (1996) states, that there are different kinds of context and separates linguistic context from physical context. A **linguistic context** is also known as a **co-text**, and the co-text of a word is the set of other words used in the same phrase or sentence. It has a strong effect on what we think the word means. **Physical context** refers to the physical location of the word, which influence the interpretation. Our understanding of what we read and hear is greatly tied to the physical context and especially to the time and place, in which we encounter linguistic expressions. (Ibid, 129)

Yule (ibid.) has as an example the homonym 'bank' as in a wall of a building, in which context the meaning is unambiguous. In thesaurus environment the linguistic and physical context is extremely important. However, even if a term has a scope note, it does not make the separation between different concepts always simple within the social sciences. For example in YSA[8] "avotyö" [Swedish equivalent: öppet arbete, 'open

---

[8] In the study the online version is used, see chapter 6.3.2 *Thesauri*.

work', 'non-institutional work', 'out-patient work'] can mean two different things and it is explained in a scope note. 'Avotyö' may mean work for disabled people in normal workplaces, which is based on a care relationship. It can also mean work done in child welfare and around family support centers.

Chang and Lee (2001) have studied the concept of context and its relation to information behaviour in dissertation research process. They asked what is actually meant by context and describe its relation to information seeking based on empirical studies of doctoral students information seeking and literature studies. They found that context is many-sided, stratified, dynamic, and consists of several situations. (Ibid.)

In the empirical part of the study co-text is emphasised more, but also the physical context is presented especially with regard to interview situations. Co-text of the studied terms will be specified in more detail to the informants during the interview process.

2.2 Indexing

In this study **indexing** is understood in accordance with ISO standard 5963-1985 as the "act of describing or identifying a document in terms of its subject content". (Ibid, 2) Further, **subject** refers to "any concept or combination of concepts representing a theme in a document" (ibid, 1) and **document** refers to "any item, printed or otherwise, which is amenable to cataloguing or indexing" (ibid, 1).

Indexing consists of the three following stages: 1) examining the document and establishing its subject content; 2) identifying the principal concepts present in the subject, and; 3) expressing these concepts in the terms of the indexing language. In practice these stages tend to overlap. (ISO 5963-1985, 2)

According to ISO 5964 **indexing language** means "[a] controlled set of terms selected from natural language and used to represent, in summary form, the subjects of documents" (ibid, 3). Thesauri are constructed to be used in information storage and retrieval. But in accordance with thesaurus function terms belong to documentary language and they may therefore differ from natural language usage. Indexing languages

---

Original: " avotyö

Huomautus: Avotyö voi tarkoittaa kahta asiaa. 1) Avotyö on vammaisten henkilöiden huoltosuhteeseen perustuvaa työtä normaaleilla työpaikoilla. 2) Avotyö tarkoittaa lastensuojelun ja perhetukikeskusten piirissä tehtävää työtä

Kuuluu ryhmiin: [73] Sosiaalipolitiikka. Yhteiskuntapolitiikka. Sosiaaliturva. Terveydenhuolto. Asuminen

[78] Työelämä. Työsuojelu. Työterveys

Lähde: Koivumäki, E. Työkeskukseen vai avoimille työmarkkinoille

Ruotsinkielinen asiasana: öppet arbete"

(YSA 2008, retrieved in 29.8.08, thesaurus last updated 29.08.2008)

can be categorised as 1. controlled indexing languages, 2. free indexing languages and 3. natural-language indexing (Rowley 1988, 52).

When studying index(ing) terms, descriptors and keywords, one is studying **metadata.**[9]

> "Metadata is data about data. It describes attributes and contents of an original document or work. --- standard bibliographic information, summaries, indexing terms, and abstracts are all surrogates for the original material, hence metadata. --- The metadata, if well chosen, should describe the central topics of a document." (Milstead & Feldman 1999)

This study focuses mainly on metadata (present indexing practices) and thus keywords (keywords, descriptors, indexing terms) are used. For the study's purposes it is essential that the words used in searches appear in certain fields (titles, indexing terms, author keywords etc). Iannella (1998) reminds us that metadata is structured data about data and states, that "this structure is the crucial element that gives metadata the edge over full-text indexing. The benefit is that the structure alludes to the semantics of the metadata".

Due to the ambiguity of natural language, different **documentary languages** (controlled vocabularies) has been developed in indexing. Documentary languages are artificial and controlled. Certain institutions develop them and inform their users about this development. The idea of a documentary language is to provide a common language for indexers and information seekers, so that expressions used in information storage and retrieval would correspond better. In documentary languages there can be separation of classification languages and indexing languages. (See e.g. Alaterä&Halttunen 2002, 71)

A **controlled vocabulary** is a list of preferred terms and non-preferred terms. Controlled vocabularies can be divided, for example, into general and domain-based or into alphabetical or hierarchical ones. A controlled vocabulary is often related into a specific domain, document collection or information retrieval system. So a controlled vocabulary may be just a list of words in alphabetical order, but in many controlled vocabularies also the relationships between the terms are represented, in which case it should be called a thesaurus. Today, in practice, the terms "controlled vocabulary" and "thesaurus" are used synonymously. (Alaterä & Halttunen 2002, 75) However, in common language usage and in linguistics a thesaurus is often seen as a dictionary of synonyms (see e.g. Karlsson 2000, 83).

A thesaurus is a kind of kind of controlled vocabulary. Soergel gives a formal definition of **thesauri**:

> "A thesaurus in the field of information storage and retrieval is a list of terms and/or of other signs (or symbols) indicating relationships among these elements, provided that the following criteria hold:

---

[9] Author note: The standards ISO 5963-1985 and ISO 5964-1985 have no definition for metadata.

(a) the list contains a significant proportion of non-preferred terms and/or of preferred terms not used as descriptors;

(b) terminological control is intended. (Soergel 1974, 38-39)

What is especially noteworthy is the emphasis on non-preferred terms. - It is via the non-preferred terms that one can express (and anticipate) the variety of possible choices and guide users to select and use in information searches the descriptors that are used in indexing.

Miller (2003a) defines thesaurus

"--- as a multi-aspectual lexicosemantic model of a terms' system reflecting a certain conceptual reality, which is used as a means of information processing and retrieval. Such point of view on thesaurus provides the satisfactory interpretation of reality and determines the normal relationships between the object and subject of information processing and retrieval." (Ibid, 2807)

So for Miller (2003a) there is no universal conceptual system, and it is seen also in thesaurus construction. He states, that every conceptual system is subjective because:

"a) to a considerably extent it depends on the complex of knowledge and world outlook of its bearers (which assumes the simultaneous coexistence of a multitude of conceptual systems differently reflecting one and the same reality);

b) it is greatly exposed to changes "in space and time"." (Ibid, 2803)

When controlled vocabulary refers to just a list of words (in one or several languages), it is not far from the idea of a vocabulary or a dictionary. In this study thesaurus is used to refer only to those controlled vocabularies, which include also semantic relationships between its terms.

Thesauri and subject indexing consist of words, which are called **index terms** or **indexing terms**, and also referred to as **descriptors** or **keywords**.

"Several individual words may be used as 'word', known as a 'descriptor', identifying a concept." (Townley & Gee 1980, 19)

In the *International Encyclopedia of Information and Library Science* (1997) keyword is undersood as a "word which succinctly and accurately describes the subject, or an aspect of the subject, discussed in a document", and is further defined in the following:

"In a permuted title index, the word from the title of a document considered to be most indicative of the subject matter, i.e. the keyword, can be used as an indexing term, permuted by one of a number of methods: KWIC (Keyword in Context), KWAC (Keyword and Context) or KWOC (Keyword out of Context)." (Ibid, 240)

In SFS Standard 5471 - *Guidelines for the establishment and maintenance of Finnish language thesauri* – "asiasana" has as its English equivalents ***descriptor, indexing term*** and ***index term***. It is defined as an indexing language word, which is

selected to describe a certain concept.[10] **Indexing language** ("indeksointikieli")[11] is defined as an index language, where terms represent natural language words and **index language** ("dokumentaatiokieli")[12] as a language used to describe the content of documents. (SFS 5471, 2) Thus, the distinctive characteristic is that indexing language represents terms from a natural language and index language does not necessarily. Index language is a more general term and indexing language is more narrow.

In ISO 5964-standards (1985, 3) **indexing terms** is defined as "the representation of a concept, preferably in the form of a noun or noun phrase". An indexing term can thus consist of more than one word, in which case it is known as *a compound term*. In a controlled indexing vocabulary a term is designated either as *a preferred term* or as *a non-preferred term*. (ISO 5964-1985, 3)

A **preferred term** is "[a] term used consistently when indexing to represent a given concept, sometimes known as "descriptor""(ISO 5964-1985, 3 and also ISO 5963 – 1985, 1), whereas **non-preferred term** is:

> "The synonym or quasi-synonym of a preferred term. A non-preferred term is not assigned to documents, but is provided as an entry point in a thesaurus or alphabetical index, the user being directed by an instruction (for example USE or SEE) to the appropriate preferred term; sometimes known as "non-descriptor". (ISO 5964-1985, 3; ISO 5963-1985, 1)

In thesaurus construction guides and standards no distinction between **multilingual thesauri** versus **multicultural thesauri** is made, e.g.:

> "Multilingual thesaurus = A thesaurus (see 3.16[13]) containing terms selected from more than one natural language. It displays not only the interrelationships between terms, but also equivalent terms in each of the languages covered." (ISO 1985, 3)

Miller (2003a) touches upon this idea of multiculturalism when speaking of the benefits of multilingual thesauri and referring to them as a "switching tool":

> "Now I want to dwell on one more problem of thesaurus construction that became more intensively debated only during recent years but had been already topical for many years – multilingual thesauri, their barest necessity, and construction specifity. [---] The present-day growth of discussions' intensity is obviously engendered by interactions between Web multilingual materials and Net-users' searching needs.
>
> But it is very important to take into consideration that such multi- or bilingual thesauri can be very useful not only in the WWW environment but also in other online

---

[10] Original: "**Asiasana** (*descriptor, indexing term, index term, preferred term*): Indeksointikielen sana, joka on valittu kuvaamaan tiettyä käsitettä." (SFS 5471, 2)

[11] Original: "**Indeksointikieli** (*indexing language*): dokumentaatiokieli, jonka termit ovat luonnollisesta kielestä valittuja sanoja". (SFS 5471, 2)

[12] Original: "**Dokumentaatiokieli** (*index language*): dokumenttien sisällön kuvailuun käytettävä kieli," SFS 5471, 2

[13] "3.16 thesaurus: The vocabulary of a controlled indexing language (see 3.7), formally organized so that the *a priori* relationships between concepts (for example as "broader" and "narrower") are made explicit." (ISO 1985, 3)

conditions that have one or another specificity (for instance, uncommon language environment). The inclusion of common-language synonymous (pr quasi-synonymous) terms in a format of a descriptor gives many additional options in indexing as well as in searching. We can build in such way not only bilingual or multilingual thesauri, but also receive a "switching" tool between databases." (Miller 2003a, 2806-7)

In this study a distinction is made between **multilingual** versus **multicultural** and also multilingual or multicultural versus **international thesaurus**. One can even say that a monolingual thesaurus may be more multicultural than a multilingual one. Furthermore, a multicultural and/or international thesaurus may be monolingual. The *content* of the thesaurus is more important than the number of its languages.

If we consider, for example, how the term "nationalism" is represented in the General Finnish Thesaurus YSA[14] and in the British HASSET thesaurus (Humanities and Social Science Electronic Thesaurus)[15] where the distinction, also at the content level, is clear.

In the YSA "nationalism" (*nationalismi*) has broader term ideologies (*ideologiat*) and it belongs to the thematic group [65]: Political science, Politics, International politics. Related terms are also represented such as ethnocentrism (*etnosentrismi*), Fennomania (*fennomania*), patriotism (*isänmaallisuus*), nationalism, idea of nationality (*kansallisuusaate*), nationality questions (*kansallisuuskysymykset*), Panslavism (*panslavismi*), Svekomania (*ruotsinmielisyys*), Slavophiles (*slavofiilit*) and Finnishness (*suomalaisuus*). In HASSET "nationalism" is understood as both an ideology and as a term of politics, but on a deeper contextual level it differs greatly from YSA:

---

[14] The YSA (Yleinen suomalainen asiasanasto) is a general thesaurus in Finnish and it covers all fields of research. The Thesaurus is maintained by the National Bibliographic Services. YSA has been used for indexing Finnish publications since 1987 in public and scientific libraries and data archives. (YSA 1999) Web thesaurus version (Finnish/Swedish) VESA – verkkosanasto. The thesaurus last updated 21.1.2005, retrieved in 25.1.2005, URL: http://vesa.lib.helsinki.fi/index.html

[15] "HASSET was developed to allow subject access to the UK Data Archive's own collections of data. Terms found in the thesaurus are used to search the Archive's online catalogue." UKDA 2002 (25.1.2005): HASSET - Version 3. URL: http://www.data-archive.ac.uk/search/hassetSearch.asp

Figure 2: *NATIONALISM in HASSET (retrieved in 25.1.2005)*

If we translated the YSA into English so as to be used in England there would be irrelevant terms such as *Panslavism* and *Finnishness* and in addition to that relevant terms such as *Scottish nationalism* and *Plaid Cymru*[16] would be missing.

The fundamental art of multilingual thesaurus construction lies in content design, and after that in the languages and in the translation. That is also why in this study the emphasis is more on pragmatic equivalence than on formal or linguistic equivalence.

In this study a **multilingual thesaurus** is considered as a tool for information retrieval and documentation, where indexing terms have equivalents in one or more foreign languages. They are not necessarily multicultural in a sense that they may be bound to the surrounding culture, e.g. the British. **Multicultural thesaurus** refers to the multicultural content and using environment of the thesaurus studied – it is not designed only one cultural group in mind. **International thesaurus** is designed to serve the needs of many cultures without cultural-specific content, and it may be monolingual or multilingual. (See chapter 10.3.1 *Different types of multilingual thesauri*.)

This study examines translatability aspects on three levels: concept, term and indexing term level.

The interest is on the **concepts** the words embody. According to (ISO 5963-1985, 1) a concept is a unit of thought and

> "The semantic content of a concept can be re-expressed by a combination of other and different concepts, which may vary from one language or culture to another." (Ibid.)

In this study the idea is appropriated and proved by e.g. cognitive psychology, that conception is also independent from language, and concepts may be communicated via e.g. pictures (Saariluoma 1992, 71). In linguistics there are also opposite views, and for example Temmerman (2000, 6) states, that concepts cannot be communicated – and probably cannot even be conceived without language. (See also Nuopponen 2003, 3-4)

---

[16] In HASSET as synonyms "The party of Wales" and "Welsh Nationalist Party" are given.

Traditionally, semantics refers to meanings and pragmatics to language usage (see previous subchapter 2.1 *Language and culture*). According to Wierzbicka (1991)

> "Since the meanings conveyed in natural language are inherently subjective and anthropocentric, they cannot be neatly divided into 'referential' and 'pragmatic', or 'denotational' and 'attitudinal'. What is needed, therefore, is a unified semantic framework, equally suitable for describing the meaning of 'cultural kinds' (such as *cup* and *mug* in English, or *sake* in Japanese), 'natural kinds' (such as *cat* and *dog* in English, or *nezumi* 'rat/mouse' in Japanese), interactional verbs (such as *promise*), *vow*, or *pledge* in English, or *materit'sja* 'mother-swear' in Russian), and so on. All such meanings are culture-specific, subjective, and anthropocentric (see Wierzbicka 1985a,b, 1987), 'referential' and 'pragmatic' at the same time. ---" (Wierzbicka 1991, 17)

There are several aspects of meanings. According to Nida (1975) **central meaning** refers to that "meaning of a word, which is generally understood when the word is given with no context; also called unmarked meaning". We can also separate **conceptual** versus **associative** meaning. Conceptual meaning is similar to denotative or central meaning. Associative meaning is similar to connotative meaning. (Ibid. 228) Yule (1996) explains, that

> "When linguistics investigate the meaning of words in a language, they are normally interested in characterizing the **conceptual** meaning and less concerned in characterizing with the **associative** or stylistic meaning of words. Conceptual meaning covers those basic, essential components of meaning which are conveyed by the literal use of a word. Some of the basic components of a word like *needle* in English might include 'thin, sharp, steel, instrument'. However, you may have 'associations', or 'connotations', attached to a word like *needle* which lead you to think of 'painful' whenever you encounter the word. This 'association' is not treated as part of the conceptual meaning of *needle*. In a similar way, you may associate the expression *low-calorie,* when used to describe a product, with 'good for you', but we would not want to include this association within the basic conceptual meaning of the expression. Poets and advertises are, of course, very interested in using terms in such a way that their associative meanings are evoked, and some linguistics do investigate this aspect of language use. ---" (Ibid, 114-115)

In this study, meanings are studied both from the perspective of their definitions and of their usage. The emphasis is on practice. Denotative, i.e. central or conceptual meanings, and connotative levels (i.e. associative or stylistic meanings) are seen more as aspects rather than as separate units of a word. The emphasis is on the denotative level of the word, but also the connotative level is studied in order to get a deeper and broader understanding of the words usage in practice. (For more concerning denotation and connotation see chapter 3.1 *On culture, language and meaning*.)

Word meanings are related to aboutness. Ingwersen and Järvelin (2005) discuss and define **aboutness** in their book *The Turn. Integration of Information Seeking and Retrieval in Context*, as

> "Fundamentally, the concept refers to 'what' an *information object,* text, image, etc. is about (i.e. the topic it discusses), and the 'who' deciding the 'what'. Several definitions exist. In this book, aboutness is not an inherent feature of a document, but

dependent on the *cognitive actor*(s) who determines the 'what' during the acts of interpretation and representation in a time-space continuum." (Ibid, 381)

In this study aboutness is understood similarly. The emphasis is on concepts, but also on different users and contexts. We can thus, therefore, separate e.g. **author aboutness** and **indexer aboutness**. The first mentioned is:

> "The *aboutness* determined by the authior(s) of *information objects* through natural language and other means of representation. For instance, the full document as well as author-generated abstract reflects author aboutness." (Ingwersen & Järvelin 2005, 381)

The second mentioned is:

> "The aboutness determined by an indexer or algorithmic indexing device, implying an analysis of an *information object*, which results in the addition to or/and a transformation of original features and concepts into those accepted by the indexer or indexing device. The use of controlled vocabularies or a thesaurus typically results in indexer aboutness." (Ingwersen & Järvelin 2005, 384)

Note: In this study the focus is not on documents, but on words and terms representing concepts, and on discourses.

2.3 Translations and equivalence

Vehmas-Lehto (1999, 12) has given a careful definition of **translating**: when translating one expresses something with the means of target language that has been expressed earlier with the means of source language[17]. The definition displays that one is not translating *language*, but the *content* of a translation, the *meaning*. (Vehmas-Lehto 1999, 12)

There is more than just equivalence and non-equivalence. The variety (and also richness in some cases) of equivalence distinction is great. Sometimes it refers more to the nature of equivalence, sometimes more to the degree.

In *Sanastotyön käsikirja* (Handbook of Terminology, TSK 1989) **equivalence** is divided into exact equivalence, (FI: *täydellinen käsitevastaavuus*), partial equivalence (FI: *osittainen vastaavuus*), broader equivalence (FI: *laajempi vastaavuus*) and narrower equivalence (FI: *suppeampi vastaavuus*). Exact equivalence is considered to occur when the concepts are identical and the terms related to it refer to the same common concept. In partial equivalence the contents or domains of the concepts differ from each other. It is also possible that in other languages one concept needs to be expressed with several concepts (i.e. with a group of concepts) or that a concept exists only in one language. If one concept is represented with several concepts in another language, it is question of a broader and narrower equivalence between different language versions. (TSK 1989, 101-103)

---

[17] Original: "**Kääntämisen** voisi varovasti määritellä seuraavasti: käännettäessä ilmaistaan tulokielen keinoin se, mikä on jo ilmaistu lähtökielen keinoin." Vehmas-Lehto 1999, 12

Translation equivalent (FI: *käännösvastine,* SV*: översättningsekvivalent*) is an equivalent, which is used to express the source concept, when there is no settled term in target language. Translation equivalent is often an ad hoc expression and its usability depends on the context. In some cases translation equivalent can be replaced by a **near equivalent** (FI: *lähivastine*, SV: *närekvivalent, närmaste motsvarighet*). Near equivalent is a linguistic expression, which refers to a similar concept in target language and thus not to the same concept as the term does in the source language. Due to linguistic and cultural differences in the construction of concepts and concept systems there might be cases, where no specific equivalent can be used. (TSK 1999, 180)

Standard for multilingual thesaurus construction (ISO 5964-1985, 9) lists several degrees of equivalence: exact equivalence, inexact equivalence, partial equivalence, single-to-multiple-equivalence or non-equivalence. (See more detailed in chapter 3.3.3.2 *Equivalence in multilingual thesaurus construction*).

As previously seen, equivalence can be understood in many different ways. In the study a functional and dynamic view is borrowed from translation science concerning equivalence (see especially chapter 3.4.1.1 *Function of translation*). An equivalent term should thus first and foremost represent the referred concept by its informative content and be usable from the perspective of its user. Second, but still of importance, comes the demand for coherence with the practice of the field and linguistic and terminological acceptability. **In this study translation equivalence does not refer to any specific kind of equivalence, but is used as a general term referring to any kind of equivalence, e.g. to exact or single-to-multiple-equivalence.**

In accordance with different approaches to language (see subchapter 2.1 *Language and culture*) the study of equivalence in this study is also divided into two different approaches, namely into **semantic vs. pragmatic equivalence**. Semantic equivalence refers more to the theoretical and literal equivalence, which can often be seen in dictionaries etc. Pragmatic equivalence refers more to the practices and considers whether the given equivalents can be used for similar purposes also in practice, that is in real (here: communicative) language usage - as in this special context – in indexing and in information retrieval. If there is no equivalence in the other language it can be seen as a **semantic primitive** equivalence (about semantic primitives see Wierzbicka 1991, 12-24).

- **Summary and Conclusions**

The definitions provided in the chapter illustrated the common concepts as well as the perspective of the study. Languages and cultures are considered to consist of several subsets. Words are also viewed as (potentially) having different kinds of meanings. The focus in the study is on different meanings and discourses, which are explored in various sub-cultural contexts from the perspective of multilingual thesaurus construction.

3 Previous research

In this chapter previous research relevant to the empirical part is discussed. As a theme multilingual thesaurus construction and equivalence is a very broad and multi-disciplinary. The aim is to illustrate the central theoretical background as well as theoretical and conceptual tools for further empirical analysis. First culture, language and meaning from the context of equivalence and thesaurus construction are discussed. Secondly the understanding of indexing in the study is discussed. Third thesaurus construction principles and equivalence in thesaurus construction are discussed. Finally translatability and equivalence is discussed. The idea is to move on from a more general background into more specific questions.

3.1 On culture, language and meaning

In this chapter, the relationship between words, meaning, language and culture is discussed. (Finnish and English as languages and cultures are illustrated from the viewpoint of this study in chapter 4.1.2 *On Finnish and English*.) Different types of meaning are related to different types of text functions (Vehmas-Lehto 1999, 73). It is important to keep in mind whether we are translating an informative text type such as a thesaurus or an operative text such as a commercial text or an expressive text type such as a poem. In this study the focus is on thesauri and therefore only issues relevant to that text type are considered.

As stated before, the social sciences are connected not only to the development of science but also to the development of their surrounding culture and society. In a social sciences thesaurus this phenomenon is seen more clearly than in, for example, thesauri of technology or medicine. Language is not static (see e.g. Aitchison 1991; TSK 1989; Varantola 1990; Wierzbicka 1997; Lehtonen 2000; Katan 2004) and therefore the language and documentation of social sciences is tied (not only, but also) with culture and time.

In chapter 2.1 *Language and culture* it was mentioned how culture consists of different layers (see Tompenaars 1993, 23) and pointed out, that culture hides much more than it reveals (Hall 1981). Strangely what it hides, it hides most effectively from its own participants. (Ibid, 30) Kramsch (1993) talks about cultural imagination and myths:

> "On the reality of facts and events that constitute a nation's history and culture is superimposed a cultural imagination that is no less real. This cultural imagination or public consciousness has been formed by centuries of literary texts and other artistic productions, as well as by a certain public discourse in the press and other media. ---
>
> The teaching of culture is all the more difficult as myth and reality both contradict and reinforce one another. --- And yet, myths cannot be discarded, for they affect the way

learners of a foreign language see others in the mirror of themselves, despite all evidence to the contrary from 'objectively' transmitted facts.

Thus the teacher of culture is faced with a kaleidoscope of at least four different reflections of facts and events, that replicate on a larger scale the hermeneutic configuration ---" (Kramsch 1993, 207)

This means the reality and different reflections can be seen as a kaleidoscope:
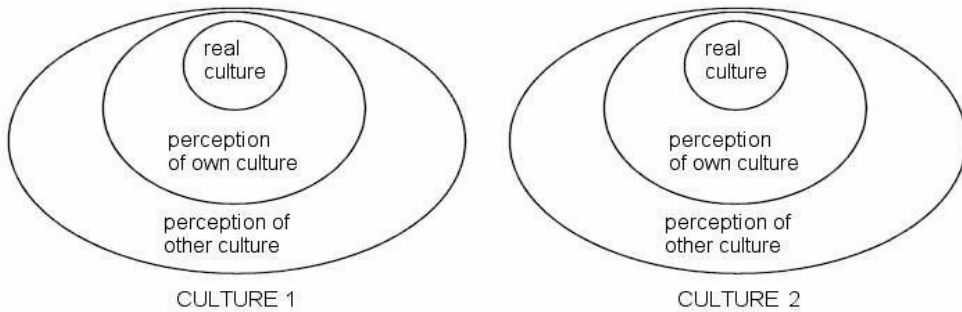
Figure 3: *Perception Rings (Kramsch 1993, 208)*

We identify ourselves and others as members of cultures. - According to a translation theorist David Katan (2004) identification can happen at the level of a continent (America, Europe, Asia), a country, a region and so on. However, cultures are also cross geographical and have political boarders and are thus sometimes more usefully categorised ethnically, linguistically or religiously. According to Katan, to be a member of a culture, one will need to share beliefs at every level of culture. Below is his sample set of beliefs congruent with being British:

Belief at the level of...	Belief in...
Value:	fair play
	democracy
	compromise
	privacy
Strategy:	internalization of feelings
Behaviour:	queuing
Environment:	an Englishman's home is his castle
	a little bit of dirt never did anyone any harm
	there's nothing like a good cup of tea (Ibid, 84)

A set of beliefs congruent with an Italian identity would be value: *the family* (la famiglia), *mother is not to be touched* (a mamma non si tocca), *is always a mother – whatever* (é sempre la mamma), *respect* (il rispetto), *aesthetics* (l'estetica). Strategy would be: to *know one more than the devil* (saperne una piú del'diavolo), *to make a good impression* (fare bella figura), *the art of making the best out of any situation* (l'arte

d'arrangiarsi). Behaviour would be: *eating well* (mangier bene) and the environment: *the house shines like a mirror* (la casa brilla come uno specchio) and *good cooking* (la buona cucina). (Ibid, 86)

We can see that cultures are an organic part of societies occurring on different levels and language is an essential part of cultures, and the broad perspective is a necessity in order to understand the detailed level of the study. The idea of language merely as a tool for communication becomes in this context far too narrow. According to Hiidenmaa (2000) language should be considered more broadly: it constructs the world and guides values and choices. Since language is not just a harmless package for our thoughts but on the contrary a true instrument to exercise of power, it is worthwhile to examine its essence and way of action in schools and work tasks. It helps us to understand also other languages essence and the societal role of language and the power that guides language's life. (Ibid, 50-51)

Following the ideas of Hall (1981) Brake, Walker and Walker (1995) have succesfully developed a taxonomy of orientations to different cultures, and according to Katan (2004, 230-233) their model is the most comprehensive in the literature (to date). Brake et al. list in their *Iceberg of Cultural Orientations* model as orientations action, communication, environment, time, space, power, individualism, competitiveness, structure and thinking. (Brake & al. 1995, 39-40) When considering thesauri as cultural products, these orientations are represented also in their content.

music, art
food and drink,
dress, architecture
institutions
visible behaviour

Hall's triad:
Technical

Language

Formal

Appropriacy
rituals
customs
ways/styles (of discourse,
dress,..)

Orientations
Action
Communication
Enviroment
Time
Space
Power
individualism
Competitiveness
Structure
Thinking

Informal
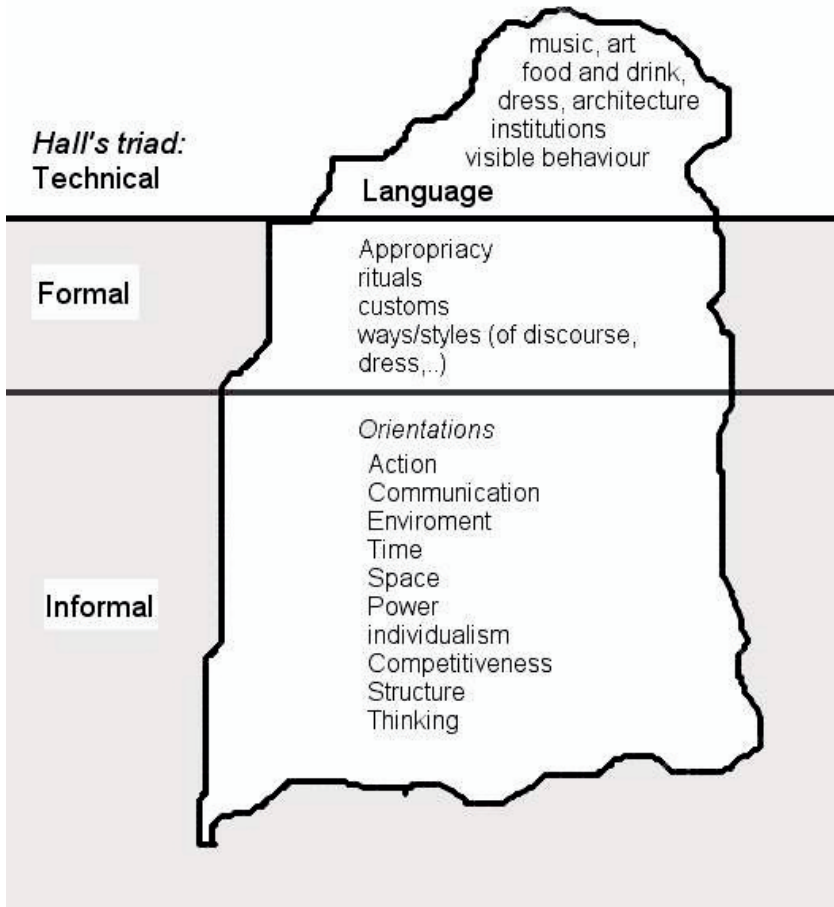
Figure 4: *Brake's Iceberg of Cultural Orientations (Brake & al. 1995, 39)*

As cultural products, the content of thesauri reflect the different layers of culture – some more visible and observable than others, concrete systems as well as more hidden values. (Hall's triad is discussed more in detail from the perspective of translation in chapter 3.4.1.3 *Domestication or foreignisation – or internationalisation*.)

According to Suojanen (1993) the world's information culture has commonly approved a "neo-language" or "an euphemistic language" in public and official discourse – technology, politics, economics, and religion – especially when the context of a phenomenon feels 'heavy', 'scary', 'threatening'. With this euphemistic language we try to guide the thinking of familiar things from a new perspective by the choice of words or a new compound term. The media especially talks e.g. about "the poor" with words such as "low-paid". (Suojanen 1993, 23) A specialist may thus in some situations use euphemisms and in some cases more unambiguous terms, whereas indexers may aim to conform to guidelines and controlled vocabularies in some institutions more tightly than in others. ISO and SFS standards about monolingual thesaurus construction (ISO 2788-1986, 12; SFS 5471-1988, 172) define what kind of term is a good indexing term. They may sometimes greatly differ from other guidelines and common language

usage and are not always easy to follow. The time dimension is also considered to be one of the major factors in communication styles. (About thesaurus construction guidelines see chapter 3.3 *Thesaurus construction*.)

Wierzbicka (1996, 3) states that language is an instrument for conveying meaning, and it is the structure of the instrument which reflects its function. For Wierzbicka, to study language without reference to meaning is like studying road signs from the point of view of their physical properties (how much they weight etc), or similar to studying the structure of the eye without any reference to seeing. (Ibid.)

Therefore, it is no use to study only how long or linguistically correct etc the terms in a multilingual thesaurus are. We need to understand what meanings the terms embody and in which kinds of context they are supposed to be used. Hall (1981) states, that in the study of languages nothing can be safely assumed, since no two languages are alike. For example English and Navajo differ greatly and therefore force the speaker into two different images of reality. (Ibid, 101)

The same applies to discourses – Finnish indexer may see the same descriptor in a different light than a Finnish socials scientist. How stable versus flexible is the meaning?

According to Häkkinen (1994, 171) many kinds of theories and tools are developed to describe meaning. According to the so-called conceptual meaning theory ("*konseptualistinen merkitysteoria*") meaning is an inner mental picture, which is produced by a certain language for the language user and which somehow describes all the objects of the sign. Problematic and even opposite opinions that arouse questions are, for example, to which point peoples mental pictures are congruent with each other and how directly do they correspond to the world outside the language. Some researchers think the relation is iconic, and thus mental pictures and also the semantic systems imitate faithfully the world outside the language. In addition, the classifications found in the concept systems are similar to those in the world outside the language. The world outside the language dictates what kinds of semantic categories are found in the language. (Ibid, 171-172)

The so-called cognitive, psychology-based semantic theory ("*kognitiivinen, psykologispohjainen merkitysteoria*") states that there is no universal, depending on the realities of the world outside the language, way to analyse or structure. Each language community has their own way to perceive the world. Linguistic semantics is considered to be only a part of a human beings conception in general. The conceptual categories rather are formed on the basis of prototypes, i.e. of as typical cases as possible, than on the basis of clearly defined categories. Therefore, when categorising entities with the help of linguistic concepts a person evaluates, which prototype the entity reminds them of the most and names it according to that. If the entity does not remind them clearly of any prototype it may cause classification problems. The cognitive categories may also overlap. For example, there are no clear definitions for concepts such as *hill*, *elevation* and *height*. - One can refer to the same object by using different terms depending on what characteristic is emphasised. For example, instead of the neutral hill one can use hillock and in that way emphasise the small size and roundness. The meaning of a word

is not thus in direct relation with the object. The meaning is not formed entirely on the basis of the relation between a word and an object. (Häkkinen 1994, 172) The study leans towards the psychology-based semantic theory.

Lykke Nielsen (2005) lists a range of sources of vocabulary ambiguities that should be considered when designing and deciding on the conceptual content of conceptual tools: subject crossover and overlaps, disciplinary focus, interdisciplinary borrowings, "fashionable" terminology, semantic shifts over time, cultural differences, semantics, acronyms, and abbreviations. According to Lykke Nielsen (2002, 16) thesaurus compilers also need to know and must make up their mind whether it is possible to assign a firm definition to any word or whether words have fuzzy meanings. The selected standpoint has impact on our view of the role of thesauri and the semantic information, needed in the thesaurus. (Ibid.)

In the fixed meaning approach (which dates back to Aristotle) in order to capture the meaning of a word one must perform an analysis, which isolates the essential characteristics of the concept, and establish a set of conditions, which are absolutely necessary to the meaning of the word. When the conditions are taken together they are sufficient to encapsulate the essential, universal meaning of the word. A major problem with this approach is deciding which attributes are included on the list, and sometimes a true expertise is needed in order to find the fixed, correct meaning of a word. (Lykke Nielsen 2002, 16-17)

According to the fuzzy-meaning approach (which Wittgenstein represents) word meanings are fluid and words have fuzzy edges in the sense that there is no clear point at which one word ends and another begins. There are number of words such as *square*, which appear to have a fixed meaning, but the majority of words do not have. It may be difficult to specify a hard-core meaning (cf. central meaning in Nida 1975), the words may have a fuzzy boundary, i.e. overlapping meanings with other word(s), and a single word may apply to a family of items, which all overlap in meaning but do not share any one common characteristics. (Lykke Nielsen 2002, 18-19) (Ibid, 19)

Words and meanings may be translatable, but they may represent or carry different world views and values in different cultures. For example, in some Indian cultures ownership is seen as a sign of unhappiness or misfortune – a poor rich person might not have enough friends to share his or her things (Lehtonen 1998, 17). The semiotic Ogden and Richard triangle is a model of how linguistic symbols are related to the objects they represent (Hjørland 2005) and it can be used to illustrate the idea discussed here before:

THOUGHT OR REFERENCE
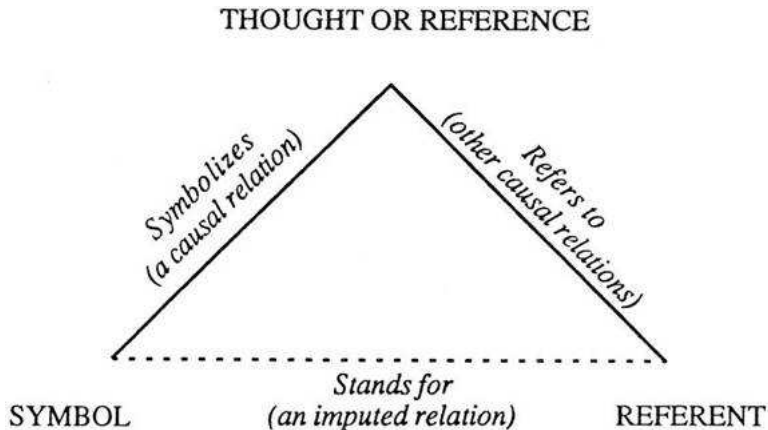
Figure 5: *Ogden & Richards triangle of meaning (cited here Hjørland 2005)*

The Ogden & Richards (1923) triangle of meaning (represented above) implies that the referent of an expression (a word or another sign or symbol) is relative to different language users (Hjørland 2005, Fiske 2000).

> "A sign, or representamen, is something which stands to somebody for something in some respect or capacity. It addresses somebody, that is, creates in the mind of that person an equivalent sign, or perhaps a more developed sign. That sign which it creates I call the interpretant of the first sign. The sign stands for something, its object [or referent]. It stands for that object, not in all respects, but in reference to a sort of idea, which I have sometimes called the ground of the representamen." (Peirce, 1931-1958, 2, 228, cited here Hjørland 2005, see also Fiske 2000, 64-65)

It is commonly stated that in different languages words can have different meanings. What are different meanings that words can embody? Vehmas-Lehto (1999, 74) lists four different types of meanings, which is based on divisions made by Švejcer[18]. The different types of meanings are: denotative meaning, connotative meaning, pragmatic meaning and interlingual meaning. (Ibid.)

The two first mentioned are rather commonly known and broadly used divisions. Hörmann (1986) describes the connotation to refer to what is associatively suggested by the word and denotation points to the object which is meant by the word. The denotation of *moon* is "earth's natural satellite" and the connotation of *moon* would be e.g. cold, distant, lonely, longing. (Hörmann 1986, 142, see also e.g. Häkkinen 1994, 173)

In different languages the denotative meaning may be the same, but the connotative meaning may differ. Vehmas-Lehto (1999, 75) gives as an example 'a snail'. In French 'escargot' has a connotation delicious, but in German 'Schnecke' connotates as ugly, slimy and in Finnish 'etana' is closest to slow.

Words can have different connotations, but correspondingly a same connotation can embody different words in different language communities. When in Finland the stereotypic image of an angry wife is a woman holding a rolling pin, in Russia it is a

---

[18] Švejcer 1973 = Švejcer, Aleksandr. *Perevod i lingvistika*. Voenizdat, Moskva 1973.

woman with an iron. As a management method in Finland the 'cane and carrot' can be used, and in Russia 'whip and gingerbread'. (Vehmas-Lehto 1999, 75)

Pragmatic meaning is determined by the relationship between an expression and participants in a communication situation. The relationship between expression and participants in communication varies due to two reasons. Firstly, language users differ from each other as to their age, origin of birth, societal position, level of education, profession etc. For example the use of foreign words ("*sivistyssana*") such as 'acute' and 'discourse' correspond with the level of education. Secondly, situations and reasons of language use vary. An audience may be small or large, known or unknown, communication may happen in a literal or oral form, the level of formality may vary from intimate to formal, the style vary from dialectical language to a neutral common language etc. (Ibid, 75-76) – The terminological case of this research, 'family roles'', can represent, for example, professional jargon as well as ordinary language or conversational or even ironic language depending on its language use context.

Interlingual meaning is also called grammatical meaning. It refers to meaning, which is connected to the relations between language units. The meaning is expressed by e.g. grammatical affixes, finite forms, prepositions, word order etc. Grammatical meaning is usually not important when translating, if not aiming at formal equivalence. (Vehmas-Lehto 1999, 74-76) In thesauri grammatical meanings are not a problem, because keywords are given in a basic form, independently from sentences. However, grammatical meanings are highly important in a different way (and not common for e.g. in fiction translations) – with regard to hierarchical relationships.

Katan (2004, 60) gives some examples of words, where the meanings differ radically according to belief:

Word	Meaning according to belief
The IRA	terrorist
	freedom fighter
Capitalism	freedom to manage property for
	profit exploitation of man by man (Ibid, 60

Causes of external events are also dependent on belief, e.g.

corruption / sleaze	basic human condition
	the need to do business and cement relations
	the legislation which is impossible to follow
	lack of enforcement (ibid, 60)

To Katan (ibid.) the problems that arise in translation of meaning are here twofold.

"First, meaning is not inbuilt but interpreted according to individual and culture-bound beliefs. Second, there is the widespread belief that translation of the denotative

meaning automatically assures a good translation. However, as Séguinot (1995:60)[19] – citing here Barthes[20] – points out, "the motivating force is generally carried by the more hidden messages, the connotative meaning ... And connotations are notoriously culture-spesific". " (Ibid, 84)

Meschonnic (2003, 342-343) has criticised the idea of separating connotation and denotation. According to him, polysemy is inseparably language and culture and this proposition means that one can no longer separate denotation from connotation nor value from meaning. Furthermore, it means that a translation that pretends to be no more than linguistic is a cultural translation that misrecognises itself as such. In pratice it is privileging, for theoretical and historical reasons, and decentring. (Ibid.)

In this study, words are still considered to have both a denotation and a connotation. It is admitted, that (following the ideas previously presented by Wierzbicka and Meschonnic) they cannot be merely separated. – It is further assumed, that we cannot achieve 'a pure denotation' - there will always be also connotation in social science contexts both in scientific and common language as well as in a documentary language context. When speaking in the empirical part (chapters 7-10) about connotations and denotations the author wants to emphasise different aspects or nuances of a word, and not considers them as 'pure' meanings.

According to Vehmas-Lehto (1999) pragmatic meaning is determined by the relationship between the expression and the participants of the communication situation. The relationship can vary due to two reasons. Firstly, the language users differ in age, origin of birth, societal status, level of education, occupation etc. (For example the use of foreign words correlates with the level of education.) Secondly, the language use situations and reasons for language use vary, because communication may occur between any combination – from a large group of strangers down to a small and familiar setting; communication may be literal or oral, the level of formality may vary; the style can be elevated or colloquial etc. The pragmatic meaning is close to the style – a word can be archaic, formal, informal, popular, belong to language of advertising etc. Pragmatic meaning is also related to text conventions. (Ibid, 75-76)

Vehmas-Lehto (1999) states that although it is necessary to consider pragmatic matters in monolingual communication in bilingual communication consideration is even more important, since the risk for distortion is greater. (Ibid, 76) In this study translatability concerns mostly the pragmatic level of meaning and this is discussed in more detail in Chapter 3.4 *Translatability and equivalence*.

As previously discussed, it is generally stated that language and culture work in close cooperation. What exactly is their relation though? Is this an equal partenershisp or is one leading or guiding the other? Which comes first? Are translation problems due to differences in languages or differences between cultures? Nida (1964, 130) states that

---

[19] Séguinot Candace (1995) 'Translation and Advertising: Going Global' and 'Debate', in *Cultural Functions of Translation*, C.Schäffner and H. Kelly-Holmes (eds), Clevedon: Multilingual Matters, 55-86.

[20] Roland Barthes (1915 – 1980) was a French literary critic, literary and social theorist, philosopher, and semiotician.

differences between cultures may cause more severe complications for the translator than do the differences in language structure. What is the role of discourses?

Social scientists have studied the relation between language usage and power and they generally do not see a language as a neutral expression of realities, cultural artefacts etc. Actors similar to policy makers use language according to their purposes and the traditional way of expressing categories, phenomena, matters etc. should be called into question. (See e.g. Mälkiä & Stenvall 1997)

According to Fairclough (1992) discourses do not just reflect or represent social entities and relationships, they construct or 'constitute' them. Furthermore, different discourses constitute the entities (e.g'mental illness', 'citizenship' or 'literacy') in various ways, and position people in different ways as social subjects (e.g. as doctors or patients). It is these social effects of discourse that are focused upon in discourse analysis. (Ibid, 3)

Language users create discourse communities, which resemble sub-cultural groups. Porter (1986, 38-39) defines a discourse community as "a group of individuals bound by a common interest who communicate through approved channels and whose discourse is regulated". Hjørland (1997) sees a discourse community as a scientific, scholarly or professional domain with unique structures of communication and publication, unique types of documents, specific terminology, information structures, information systems, literature and relevant criteria. In his book *Information Seeking and Subject Representation* the concepts of disciplines, subdisciplines, discourse communities, and epistemic communities are used interchangeably.

> "The ideals about scientific communication stated are mostly met on the subdisciplinary level where the problems are most precise and the communication most effective. Disciplines are perhaps often too big to serve as epistemic communities, but to a lesser degree the same principles also prevail on this level. Most formal planning of information systems and databases are done at the disciplinary level. (Hjørland 1997, 51-52)

Much that has been discussed earlier can be found in a summarised form in four theses about language, communality and power, formulated by Mälkiä (1997, 18-20, translation made by the author). They also represent the central underlying assumptions and motivations behind this study.

In his (Mälkiä 1997) first thesis, Mälkiä states that language, social interaction, communality, identity and power are much more strongly connected together than we have traditionally thought. As a structurer and an interpreter of reality, as a mediator of purposeful communication, and as a converting/hiding constructor of a social interdependency language is not neutral, not an occurrence that mirrors reality passively. In a social context, language and language use is often connected with the creation, maintenance and strengthening of power- and interaction relations between individuals and groups, and inside groups. That is why on the background and as organically connected to language and linguistic, expressions are always another world: the complex and encroaching to everywhere network of social and power relations. This network

creates the context, where linguistic expressions should be proportioned and against which they should be studied. (Ibid.)

From this standpoint, language and linguistic expressions are above all social. In a similar way it also becomes obvious, that language has influence and power. The power of language penetrates everywhere in social interaction, is often unintentional and subconscious, insidious, cultural and represents mental power shaping our thinking and acting. (Ibid.)

The second thesis states that to study the interaction between language, communality and power is one of the most central tasks in social, linguistic and cultural studies today. During the past years this kind of need has been emphasised by the changed societal stature of knowledge, a change typical in all the Western societies. Along with this development the intellectual operations – such as formation, revision, transmission and utilisation of knowledge – has advanced to an increasingly central power that guides the social development. Knowledge and know-how have progressively become key factors in production. Because knowledge is mainly bound to the language - linguistically created, expressed and exchanged - it is additionally obvious, that the more important knowledge becomes, the more important it becomes to understand the linguistic conditions, possibilities and limitations of its production, revision, analyse and transmission. (Ibid.)

The third thesis states that to study language, communality, and power, requires strengthening the interdisciplinary research tradition. Social-linguistic or linguistic-social sociological and cultural studies crossing the traditional research boundaries become especially important. During the past years there has been a stronger development of this kind of research tradition. (Ibid.)

According to the fourth thesis, when considering the relations between language, communality and power we need, in addition to the research describing, explaining and understanding the phenomenon, to pay more attention in the future to supporting research traditions, which question, liberate, are critical and constructive. If the current language and linguistic practices direct scientific thinking too much, research becomes lazy and inefficient; lazy language produces only lazy thinking, and lazy thinking starts easily support already earlier adopted way of thinking and current power-relations. (Ibid.)

As discussed previously, it is important to consider which terms we use to communicate. Thesaurus construction is in its nature terminology work[21]. According to The Finnish Terminology Centre (TSK 1999) terminology work is to collect, analyse, store, modify and/or represent concepts and terms.

Terminology work can be normative or descriptive or both (their combination). According to Varantola (1990), in the domain of technology terminology work is usually, by nature, normative. The aim is then to increase the contractual unambiguity

---

[21] *Terminology work* - in Finnish *sanastotyö* or *terminologinen työ*, in English also *terminology management*. TSK 1999.

between concepts and terms and thus to assure the trouble-free flow of information. Terminology work in the domain of technology is thus not descriptive and does not aim to illustrate and describe the variety of meanings in the use of concepts. (Varantola 1990, 7)

In thesaurus construction, the primarily function is to standardise the use of terms in information storage and retrieval, and to provide a common language for information seekers and indexers. It is also noteworthy, that thesaurus terms often represent several concepts and thesauri also include non-preferred terms to guide the user and to describe the variety of terms and expressions referring to or close to the preferred term (see chapter 3.3. *Thesaurus construction*). When being descriptive the aim is still not to map all the possible variants, but to anticipate the ones that most probably could cause misunderstandings and/or unwanted differences in indexing and in information retrieval. The linguistic and cultural diversity, as well as different meanings, are then included in a thesaurus in a focused and narrow way instead of reflecting the true variety of language use.

Terminology work in thesaurus construction is thus primarily also normative, since the aim is to guide both indexers and information seekers to use certain terms. It is also normative in the sense that there are standards and rules according to which decisions and selections are (or should be) made. Genderism and non-discriminatory language use also applies to thesaurus constructors. What is said and what is not said gives valuable information about the culture and the language. For example, in Finnish we have established terms for mothers with a career, but not for fathers with career. We talk about care-giving fathers (*hoivaisä*), but not about care-giving mothers. We have an established expression *pehmoisä* ("soft-dad"), but not for *pehmoäiti* ("soft-mum"). Multinational marriages have become more common and a new term *transnational family* has been established. (About these example terms and others see e.g. HS 2004, Juvakka & Viljamaa 2002)

• **Summary and Conclusions**

In the study, cultures are considered organic parts of societies occurring on different levels. Furthermore, language is considered an essential part of culture.

Language is seen as an instrument for conveying meaning, and the meanings in a social science thesaurus are often regarded as fuzzy. The meanings are often considered as translatable, but as potentially carrying different world views and values in other cultures.

The language of a social science thesaurus is seen as tied to e.g. time and culture. Culture is considered to consist of different layers, both visible and concealed, and the invisible to be hidden most effectively from its own participants.

Furthermore, languages are viewed as consisting of different discourses. Discourses do not just reflect or represent social entities and relationships, but they

construct or constitute them. Thesaurus construction is considered to be terminology work and primarily as normative (instead of descriptive).

3.2. Indexing

This chapter presents an overview of indexing as a notion, indexing as a research topic, the indexing process, the tools and languages of indexing, and indexing inconsistency. The viewpoint is not merely theoretical - some of the current practices are also briefly illustrated and reflected to the ideas and theories represented in the literature.

Anderson (1997, 336) characterises indexing, cataloguing and classification as the methods for the organisation of knowledge. These methods are often confused, although they have separate intentions and functions. Indexing aims at providing access to document collections, single documents and parts of documents without regard to the location of these documents, whereas cataloguing is the indexing of documents in particular collections and classification is the arrangement of index (or catalogue) categories and entries in accordance with their relationship to each other. The fundamental method for organising knowledge is indexing. Cataloguing and classification depend on indexing and in fact represent special cases of indexing. (Ibid, 336-337)

According to Soergel (1985, 64) "*indexing* is commonly used for the intellectual process of subject indexing: establishing relationships between an entity and appropriate subjects or, put differently, assigning subject descriptors to an entity", whereas "*cataloguing*" is commonly used for the intellectual process of indexing in most library situations". Iivonen (1989, 4) notes, how in library and information science (LIS) indexing and classification are often conceived as parallel processes, not as overlapping processes which can be seen also in the names of many books and writings (e.g. *D.J. Foskett, Classification and indexing in the social sciences 1974, Vickery, Classification and indexing in science 1975*). (Ibid.) In this study, the focus is on the problems in keywords and descriptors, and subsequently abstracting, classification and problems associated with the other aspects of indexing and/or related to it are not concidered in detail.

As a research topic indexing is very broad and complex. It can be studied, for example, from the viewpoint of an indexer (about indexing inconsistency see e.g. Iivonen 1989 and 1995; Saarti 1999; Mai 2001), an information seeker (about thesaurus' effectiveness in information retrieval see e.g. Kekäläinen 1999, associative relationships in information retrieval, see Tudhope, Alani & Jones 2001), automation (see e.g. Lahtinen 2000) and of a constructor of an indexing system (about semantic problems see e.g. Doerr 2000, about word association thesaurus see Lykke Nielsen 2001). In this study the emphasis is placed on human effort and the viewpoint is that from the thesaurus constructor - but users are also strongly involved.

In Finland indexing is not studied to such a great extent as the other aspects and topics of IR, although its aim is to help information retrieval and it can be considered as an integral part of the IR field. The value of indexing thus also depends on its impact on information seeking (Rowley 1988, 48).

> "The objective of any index is to be able to retrieve the records or documents that have been stored and organized by the indexing process. --- it is important to remember that an index is of any value only if information and documents can be satisfactorily traced. The indexing process creates a description of a document or information, usually in some recognized and accepted style or format. Successful retrieval hangs on the searcher being able to reconstruct that document description when he wants to locate the document again. Indexing, and later searching, centre upon the matching of document profiles. Two points cannot be too strongly emphasized:
>
> the searcher needs to appreciate indexing principles in order to enhance the possibility that he will bring the appropriate document or information profile to the index
>
> the indexer must consider the index's potential audience."
>
> (Rowley 1988, 48)

Furthemore, indexing is usually not seen as an easy undertaking. According to Rowley (1988) all indexes must cater for the variety of potential approaches to a subject. This leads to two categories of problem. First, the indexer must label or name the subject effectively, and second, the indexer must find some avenue by which to signal the relationships between subjects to that search strategies may evolve. (Ibid, 49)

Within the first category problems arise related to synonyms (e.g. *salt versus sodium chloride*), near synonyms (e.g. *wages versus income*), homographs (e.g. *bear*), plural versus the singular forms of nouns, compound words (e.g. *merchant ships*) and composite subjects (e.g. *an encyclopaedia of bibliography versus a bibliography of encyclopaedias*). The second category covers problems around the specification of relationships (semantic and syntactic). (Rowley 1988, 50) The typical problems, especially in Finnish, are ambiguity, inflectional forms, compounds and derived words (Järvelin 1995, 165-169).

3.2.1 Indexing process, languages and tools

Vickery (1965) identifies three stages in the operation of indexing. In the first stage "the text is scanned to select a set of words, phrases or sentences which collectively represent its subject". Secondly, "a decision is taken as to which of these subject descriptions are worth recording as being relevant to the purpose of the retrieval system". Thirdly, "the relevant subject descriptions are transferred into the standard descriptor language used in the system". This is consistent with the ISO standard 5963– 1985 *Documentation – Methods for examining documents, determining their subjects, and selecting indexing terms*. Rowley (1988, 50, 52) expresses the same linear process similarly – *Familiarization* → *Analysis* → *Conversion of concepts to index terms* – and

states, that if an indexer is familiar with a given indexing language (s)he may be capable of accomplishing the stages simultaneously.

Soergel (1985, 63) states that "*indexing*" in its most encompassing sense refers to the total process of integrating new entities and relationships into a data base. The process is two-pronged: 1) the intellectual process – assigning an identifier to a new entity; determining any text associated with the new entity; and determining relationships; 2) the clerical process – inputting the data generated into the database and integrating them into its structure. The first process (intellectual process) is also called indexing in its narrow sense. (Ibid, 63-64)

Indexing activity can be separated, e.g. into pre-coordinate and post-coordinate indexing[22]. In pre-coordinate indexing the coordination of separate concepts is done at the time of indexing (input). Post-coordinate indexing means that the coordination of separate concepts is done at the time of the inquiry (output). Many index languages of the thesaurus type provide terms for post-coordination but some of the terms themselves are compounds, which is reminiscent of pre-coordination. This balance between the two styles presents one of the most difficult problems in the design of thesauri. (Gilchrist 1971, 2)

In a pre-coordinate indexing system a document about family roles in Finland could be indexed with descriptors "family roles – Finland". In a post-coordinate index system, the document would be assigned with the terms "family roles" and "Finland", but no relationship would be assigned to them, in which case the user can retrieve documents that include one or both of the terms.

Hjørland (2006) states that using post-coordinate indexing, single words may be combined, which increases recall. Respectively, the drawback is particularly the increased possibility of false drops. In online databases indexing phrases are similar to a kind of precoordinate indexing. The advantage of using precoordinate indexing in electronic searching is the elimination of false drops, whereas the drawback is the decrease in recall. (Hjørland 2006. See also Miller 2003a, 2801)

As stated before, indexing is a complicated task. Mai (2001) has used semiotics as a framework for understanding the "interpretative" nature of the subject indexing process. About its analysis Mai (2001) states, that

> "One obvious conclusion that may be derived from applying Peirce's semiotics to the subject indexing process is to demonstrate how fundamentally interpretative and, therefore, variable, the entire process is. To portray the process in this way and to make a point of saying that it is a useful conclusion should not be seen as an attempt to demean the process as something that will not yield itself to precision and exactness. Rather it is a way of showing how inexplicably profound and human the process is. Indeed, it is the profoundly human nature of the subject indexing process that makes it so impervious to analysis solely by quantitative empirical methods on the other hand,

---

[22] Soergel (1974, 115) suggests to use terms post-combination and pre-combination instead of post-coordination and pre-coordination. According to Vickery (1975, 5) post-coordinate systems are also called as 'manipulative' indexes.

and so demanding of the need for qualitative and humanistic approaches to understanding it on the other." (Ibid, 619-620)

In the study, indexing is studied from the semantic and human point of view with the help of discourse analysis. Nevertheless, the focus also remains foremost on translatability issues.

According to Rowley (1988, 52) "an indexing language is the language used to describe subject or other aspects of information or documents in an index". The three main categories of indexing languages are 1. controlled indexing languages, 2. free indexing languages, and 3. natural-language indexing. (Rowley 1988, 52)

The first category includes subject heading lists and thesauri. The indexer selects and assigns terms to documents in accordance with this prescribed list of terms. Controlled indexing languages are usually used in systems with a human assignment of indexing terms. The second category – free indexing languages – is more difficult to delimit. When using free indexing languages an indexer selects the word(s) or term(s) that suits the subject. Free-language indexing is most common in a computer-indexing environment. When using natural language indexing, the indexer selects words and terms from the document. The last mentioned type could thus also be seen as one type of free-language indexing. (Rowley 1988, 52-53. See also the flowchart of the indexing operation using a thesaurus in ISO 5963-1985, 5.)

Alaterä and Halttunen (2002) notice that the rules of constructing indexing language vary slightly, but the rules often follow certain common guidelines. Selected terms should be established in practice, all the terms should be presented in a specific form (for example adjectives and verbs are not included) and that in an indexing language a concept is covered by only one term (in natural language same concept can be expressed with several different expressions). (Ibid, 75)

According to Rowley (1988) terms from an indexing language may be assigned in one of three ways:

"1 analytically, or by a human indexer who analyses subject content and selects and assigns index terms which accurately reflect subject content;

2 clerically, or by a human indexer selecting terms clerically according to some prescribed procedure. For example all significant words in the title and first two lines of the abstract of a document may form the basis for the indexing of that document. The human indexer works mechanically and rapidly; he should require no insight into the document content;

3 automatically, or by a computer selecting terms from titles, abstracts or the full text of the document." (Ibid, 57)

Anderson (1997) also separates two fundamental approaches to the description of the content, features and possible applications of messages for the purposes of information retrieval – human intellectual analysis and computer algorithm.

"Human indexing is often erroneously referred to as 'manual' indexing, but it is performed by the mind, not the hands. Increasingly, these two approaches to the analysis of the messages are being combined in order to take advantage of the

strengths of each approach, and to counterbalance their weaknesses as well." (Ibid, 338)

The Finnish general thesaurus YSA represents the first category – there is a controlled indexing language and in the Finnish LINNEA and ARTO and also British COPAC databases, indexing is done analytically. In addition, in LINNEA and ARTO a specific controlled vocabulary (YSA) is used.

In the previous discussion the emphasis has often been on the terms, but there are also other levels and perspectives. The conceptual level has been especially neglected. As Iivonen and Kivimäki (1998, 91) state "unfortunately most studies of indexing focus on terms, not concepts, and so do most rules for indexing".

Nowadays different combinations are quite common (see an example below of a combination of the first and third ways). In addition, the distinction between descriptors, identifiers and keywords is not always as clear in common language usage and database manuals as it may seem to be in the LIS literature.

For example, in the Sociological Abstracts database the indexing terms (descriptors) are produced by authors (Social Scisearch[®)]) and by indexers with the help of a certain thesaurus (Sociological Abstracts). In Social Scisearch[®] the keywords (DE) are supplied by the authors within their papers. They are copied directly from the original document and thus not edited or modified by an indexer etc. Keywords Plus (ID) are created automatically from the references of an article, without any "human effort".

> "Authors sometimes provide a list of keywords or terms that they feel best represent the content of their paper. These keywords are contained in the ISI record (1991 data forward, depending on the database) for each article and are searchable. In addition, ISI generates KeyWords Plus for many articles. KeyWords Plus are words or phrases that frequently appear in the titles of an article's references, but do not necessarily appear in the title of the article itself. KeyWords Plus may be present for articles that have no author keywords, or may include important terms not listed among the title, abstract, or author keywords." (ISI Citation Databases Help 2000)

The meaning of ID and DE differ from the above definition in a CSA database, which also contains an LISA (Library and Information Science Abstracts). In a CSA database Keyword Search (KW=) is used to search simultaneously in title, abstract and descriptors (CSA 2002a). Descriptor Search (DE=) is based, by contrast, on the indexing terms taken from a thesaurus or controlled vocabulary, whereas Identifiers field (ID=) contains subject terms not included in the controlled vocabulary but considered by the indexer to be highly relevant to the record. It may thus include different synonyms, abbreviations, acronyms, company names, trademarks etc. (CSA 2002b)

3.2.2 Challenges

Chen (2003) has stated, that

> "Scientists who are seeking information outside their own research area are unlikely to be aware of the current terms being used in less familiar areas and the categories under which this material is classified in indexing services. One may overlook potentially relevant material if it has been classified in unexpected categories.---" (ibid, 7)

The imperfection of using controlled vocabularies is well known, but the benefits of those, e.g. thesauri, are also commonly established. Tudhope summarises this as follows,

> "Studies support the use of thesauri in online retrieval and the potential for combining free text and controlled vocabulary approaches (e.g. Fidel 1991[23]). There are various research challenges, however, including the 'vocabulary problem' - differences in choice of index term at different times by indexers and searchers (Chen et al. 1997[24]). Indexer and searcher may be operating at different levels of specificity, and at different times an indexer(s) may make different choices from a set of possible term options. While conventional narrower term expansion may help in some situations, a more systematic approach to thesaurus term expansion has the potential to improve recall in such situations." (Tudhope et al. 2001)

Järvelin (1995) has summarised the problems connected to natural languages from the perspective of information seeking in text databases. First, language is not common and natural language is only partly common to the members of the culture. Several different sub-cultures and discourses exist in societies. In addition, language is not static as it is developing all the time. Language is also ambiguous and additionally each person has his/her own style of interpreting the message. There exists synonymy, periphrases, anaphoric expressions, compounds, prefixes and suffixes. It is also noteworthy that in most languages words inflect. Derived words and misspellings are problematic, too. The problems that are especially typical of Finnish are ambiguity, inflectional forms, compounds and derived words. (Ibid, 165-169)

Järvelin (1995, 180-184) has also summarised proportionally the advantages and disadvantages of the use of index terms. The strong aspects of using index terms are:

- the search of broad concepts is simple, if there exists a proper index term in documentary language ; user saves much effort, when there is no need to input alternative expressions for the concept ;

---

[23] Fidel, R. 1991: "Searchers' selection of search keys (I-III)". Journal of American Society for Information Science, 42(7), 490-527.

[24] Chen, H., Ng, T., Martinez, J. and Schatz, B. 1997: "A concept space approach to addressing the vocabulary problem in scientific information retrieval: an experiment on the Worm Community System". Journal of the American Society for Information Science, 48(1), 17-31.

- when using the structure of documentary language one can often in a simple way choose the whole hierarchical index term group to be searched at one time, i.e. user can choose a broad term with all its narrower subterms ;

- documentary languages often also include terms which describe the type of a document, but which are not usually present in the document title or in the abstract (for example in medicine *clinical* versus *theoretical*, *in vitro* vs. *in vivo*) ; these can be used efficiently to limit the searched documents ;

- precoupling decreases incorrect connections between terms and unrelevant search results (for example search terms *plants* Λ *poisons* is an ambiguous expression, but precoupled terms *poisonous plants* and *herbicides* are not ambiguous ;

- index terms given to documents are considered by an indexer. They embody the meaning produced by documentary language. It depends on the competence of an indexer and on the organising of indexing work as to how properly the given index term describes the topic of the document. (Ibid, 180-181)

  Respectively the disadvantages are:

- Index terms are not flexible : their meanings do not develop as fast as the language used in documents ;

- index terms may not be part of common language and therefore they may seem artificial and foreign:

- documentary languages do not include the most recent words or their relationships appearing in documents ;

- the most recent version should always be available in the documentary languages – and in addition all the previous ones, so that the user can search for older documents ; each version is used only in part of a database ;

- the same documentary language is usually not used in several databases. (Ibid, 181)

Thesauri faces similar problems and in some cases aim to report e.g. the different variations in time. In CSA, the thesaurus of Sociological Indexing Terms (English, visited 2000, 2002, 2009) the user is guided through to the history of preferred terms – which forms were used before, what the non-preferred and related terms were also during that time, this gives a good picture of the changes in the used terminology over time and of the updating policy of CSA:

> Homemakers
>
>> Formerly (1964-1985) DC 214425, Homemaker/Homemakers.
>
> Use For
>
>> Homemaking (1965-1985)
>>
>> Housewife/Housewives (1963-1985)

Homemaking

DC 214435.

Housewife/Housewives

DC 216392.

Related Terms

Family Life

Formerly (1963-1985) DC 171700, Family Living/Family Life.

Housework

Formerly (1983-1985) DC 216395.

Sexual Division of Labor

Added, 1986.

Wives

Formerly (1963-1985) DC 489880, Wife/Wives.

Workers [+]

Formerly (1963-1985) DC 492500, Worker/Workers.

In the chapter, the focus is (accordingly to the focus of the study) on vocabulary indexing tools such as thesauri. What kinds of meanings the word thesaurus may embody to the different informants of the study? Thesaurus as a term seemed already to be an illustrative sample from the discourse point of view, as to how differently the same, concrete object, can be seen within one field (see also Keränen 2000).

Buckland (1999) notes that in common language usage "vocabulary" has a good connotation, but in the LIS field the connotation is clearly negative:

"Vocabulary" is a regular, respectable word in everyday discourse, but it has had an unsatisfactory position in Library and Information Science. Here it seems to be somewhat alien, something exotic, that has wandered away from its ordinary habitat into another environment. It seems, somehow, an immigrant, useful, but somehow suspect. When used in Library and Information Science, "vocabulary" is commonly and awkwardly qualified. One speaks of "natural language vocabulary" and of a "controlled vocabulary." These phrases imply possibilities of improper "unnatural language vocabulary" and, perhaps, dangerous "uncontrolled vocabulary." It is as if the concept of vocabulary has been is only half accepted in our field. Suppose that we were adopt it and naturalize it. What would we find to say about it?" (Ibid.)

As an answer Buckland (ibid.) makes three claims concerning vocabulary:

"1. An Economic Claim. Vocabulary is central to the cost-effectiveness of digital libraries, and, therefore, to returns on investment. There a massive investment world-wide in making repositories accessible over networks and also a major investment in providing indexing, categorizing, and other metadata. A situation of increasing difficulty arises for users of the repositories because the number and proportion of

network-accessible repositories with unfamiliar metadata vocabularies is increasing. Decreasing effectiveness in selection is the predictable result. (We use "selection" as a general term to include searching, filtering and retrieval.) Therefore, any technique that can assist in the use of unfamiliar metadata, either by making that unfamiliar metadata more familiar or by mitigating the consequences of it being unfamiliar, could provide enormous leverage in improving the rate of return on the enormous investments that have been made in establishing repositories and their metadata. ---

2. Issues of Identity are Central to Information Science. In a world in which the politics of identity is central, issues of identity are also central to information science, and they are so for reasons relating to the role of vocabulary.

3. Vocabulary is a Central Component in Digital Libraries. The claim here is, firstly, that all filtering and retrieval systems can be modeled in terms of a series of transformations of sets (or "collections") from one state to another and, secondly, that "vocabulary" is an appropriate term for the variety or range of values in any given set (or collection)." (Ibid.)

In International Encyclopedia of Information and Library Science vocabulary (1997) "vocabulary control" is introduced in a positive sense:

"Because of the enormous variability in the human use of language, some sort of vocabulary management is essential in the overall information retrieval process. If the knowledge organization system itself does not provide it, then the user will have to provide it by trying to think of the various ways in which a particular concept might be expressed.

The purpose of vocabulary management systems is to provide links among terms that are synonymous or are considered to be equivalent in the context of the knowledge organization system (e.g. 'lawyer', 'attorney', 'barrister', 'solicitor') and to point from any initial term to other terms that are narrower in scope or share some other type of important relationship. Some systems also provide pointers to broader terms as well." (Ibid, 348)

Accordingly, thesaurus is defined as:

"A lexicon in which words are grouped by concept, thus providing a grouping or classification of synonyms or near-synonyms, and a set of equivalent classes of terminology. Thesauri of the most commonly used terms in various fields have been published so as to permit a harmonizing of indexing terminology in these fields." (Ibid, 443)

However, the negative connotations of vocabularies (e.g. thesauri) can be understood by considering what has previously been said about natural language, in the same context of indexing:

"Used in reference to **indexing**, where there is a choice between adopting a structured and consistently applied set of terms, or relying on the (natural) language which actually occurs in the documents that are indexed. The natural language alternative has the advantage of responding to changes in terminology and offering the user access via current understanding of a topic." (International Encyclopedia of Information and Library Science 1997, 314)

Nevertheless, many recent studies support the usefulness of thesauri. They are also considered useful in full-text databases and in the Internet environment (e.g. Doerr 2000, Bates 1998, Milstead 1998), but the tone of those studies is often defensive, almost manifestative. The arguments, negative connotations and common assumptions seem to affect in the background.

Miller (2003a) concludes his encyclopedic article about thesaurus construction:

> "And now I want to put the question that emerges from time to time: for what is a thesaurus needed and why is the methodology of its construction is so important?
>
> In my opinion, a methodologically accurate constructed thesaurus essentially provides certain significant advantages to the whole information retrieval system, which it supports.
>
> Firstly, a thesaurus, constructed on the systems principle, qualitatively increases "hospitality" of the whole information retrieval system to new concepts because the process where every new concept receives its own proper place in the conceptual network.
>
> Secondly (and this is inseparable from the first), such a thesaurus construction philosophy sharply reduces even the possibility of divergences of conceptual interpretation of one and the same descriptor as "in space" (by two different indexers as well as by indexer and user) or "in time" (by one and the same indexer during different periods).
>
> Thirdly, such philosophy significally reduces the duration of information retrieval, which (as R. Fidel and L. Su [---][25] clearly demonstrated is the most important quality of an information retrieval system for users." (Ibid, 2807)

Miller (2003b) states that similar to the Internet, which can be seen as an independent entity, a thesaurus should also be independent from the concrete documents and based on conceptual totality.

> "--- Such a thesaurus must:
>
> include the maximum number of terms and their synonyms,
>
> be based on objective relations between terms,
>
> be multilingual, and
>
> be "hospitable" for new terms." (Ibid, 2815)

(Thesauri are discussed more from the constructors' point of view, later in chapter 3.3 *Thesaurus construction*.)

---

[25] Fidel, R. Who needs controlled vocabulary? Spec. Libr. **1992**: *83* (1), 1-9.

Su, L.T. Is relevance an adequate criterion for retrieval system evaluation: An empiric inquiry into the user's evaluation. Proc. ASIS Annu. Meet. **1993**, *30*, 93-103.

3.2.3 Aboutness and indexing inconsistency

This chapter concentrates on the earlier research results of indexing inconsistency. It is important to make a difference between indexing consistency and interindexer consistency, because the factors influencing it are partly different. When discussing indexing inconsistency it is essential to consider also aboutness, which is seen as a closely related concept and an item that might have indexing inconcistency.

In LIS field aboutness is often discussed as "what is this document or seeking task about" (see e.g. Järvelin 1995, 29). In this study, the perspective is slightly different. As Leininger (2000, 4) states, "indexing effectiveness concerns the ability of these terms to facilitate comprehensive, accurate retrieval" (see also Hjørland 1997, 62-63, Hutchins 1975, 115). This study concentrates on translatability issues within thesaurus construction. When constructing a multilingual thesaurus we cannot avoid questioning what the index(ing) term/descriptor is about. Does it truly refer to the same concept in different cultures and subcultures? Is it used consistently in indexing? Or is it used in other cultures? - If not, why not? - Is there no need for a certain concept or is the topic covered from another perspective using different concepts?

In this study, aboutness is thus dealt with from two perspectives and at two levels, which leads us to following problems:

> 1. **indexer aboutness** as a document description problem at the (monolingual) indexing term level and;
>
> 2. **concept aboutness** as a multilingual thesaurus construction problem at the (multicultural) conceptual level.

In indexing and also in thesaurus construction 'aboutness' is an important element of analysis. 'Aboutness' also carries several meanings depending on the perspective and its user. Ingwersen's (2002) typology of aboutness operates with the following categories:

> "Author aboutness, i.e. the *content as is*;
>
> Indexer aboutness, i.e. the *interpretation of contents* (subject matter) with a purpose;
>
> Request aboutness, i.e. the user or intermediary interpretation or *understanding of the information need* as represented by the request;
>
> User aboutness, i.e., *user interpretation* of objects, e.g., as *relevance feedback* during interactive IR or the use of information and giving credit in the form of references." (Ingwersen 2002, 289)

As in the case of different meanings, when discussing 'aboutness', divisions into lower categories also make the phenomenon and terms more transparent to its different users and describe efficiently whose perspective is being considered (cf. Venuti 1995 and 1998).

According to Ingwersen (2002) indexer aboutness is based on human interpretation of information objects (- books, articles etc.) and can be viewed from several perspectives.

> "Indexer aboutness can be directed towards the subject matter and meaning of the information object or towards its future potential intellectual use or user grouping(s). The former purpose is the common issue of indexing while the latter calls for tremendous predictive power of the domain expert indexer. ---" (Ibid, 291)[26]

Indexing consistency/inconsistency refers to the degree of similarity in indexing of the same documents in different situations (see Iivonen 1989, 53). The consistency can be examined by comparing the indexing results between several indexers or from the viewpoint of interindexer consistency, that is between several indexing situations, where indexing is done by the same indexer(s) (Meriläinen 1985, 59).

According to Leininger (2000, 4), "indexing consistency addresses the degree to which indexers, when assigning terms to a chosen records, will choose the same terms to reflect that record". This study touches on the indexing consistency in this above-mentioned sense. The aim is finding out how (consistently) different indexers representing different cultures index the same documents. The consistence in this study refers to the degree of conceptual equivalence (cultural differences) and indexing term equivalence (also institutional differences) and the emphasis diverges thus from the studies reported here. Relevant aspects are also indexing quality and indexing effectiveness. Leininger (ibid.) defines indexing quality as the degree to which chosen index terms accurately reflect the content of a given record. Furthermore, indexing effectiveness concerns the ability of these terms to facilitate comprehensive, accurate retrieval (ibid.). Leininger also states that

> "Though all three of these aspects of the indexing and retrieval process are certainly not synonymous, the characteristics of each are contingent upon, and reflected in, others. The quality of indexing, though of undeniable importance, is subjective and can be difficult to quantify. The effectiveness of indexing is subject to variables such as searches style, subject knowledge, and the powers and limitations of the interface and related tools being employed. Therefore, indexing consistency has long been used as a means to provide insight into the potential quality and effectiveness of indexing. It stands to reason, and has been demonstrated by Leonard (1975[27]), that interindexer consistency is positively correlated with retrieval effectiveness." (Ibid, 4)

When discussing the consistency (or inconsistency) of indexing it is also important to keep in mind, that the aim of the documentary languages, e.g. thesauri, is to increase

---

[26] See also Lancaster, F. W. (1998) *Indexing and Abstracting in Theory and Practice*. London: UK, Library Associate Publishing.

Ingwersen, P. (1996) Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory. *Journal of Documentation*, 52(1), 3-50.

Jacob, E. K., and D. Shaw. (1998) Sociocognitive perspectives of representation. In M. E. Williams, ed. *Annual Review of Information Science and Technology*, 33, 131-85.

[27] Leonard, L.E. (1975) *Inter-indexer consistency and retrieval effectiveness: measurement of relationships*. Unpublished Dissertation, University of Illinois, Graduate School of Library Science.

predictability, not consistency. Even bad indexing may be highly consistent (Iivonen 1989, 12; Fugmann 1994, 420-421).

The metadata, if well chosen, should describe the central topics of a document (Milstead & Feldman 1999). There are several studies about human inconsistency in the selection of index and search terms. These "human factors" are often why the usefulness of the controlled vocabularies is questioned. However, research has shown that people tend to agree on the major topic of a document or the key term in a search query. The inconsistency is greater for subthemes (Iivonen 1989; Persson 1991; Weinberg 1996).

Leonard (1975, cited here Leininger 2000) has found, that interindexer consistency is positively correlated with retrieval effectiveness. According to Iivonen (1989, 174-175) indexing is more consistent when using a standardised vocabulary. The same kind of phenomena can also be seen in information seeking. Iivonen (1995) has studied the consistency in the formulation of query statements and shown how concept-consistency is higher than term-consistency and how the time also affects the consistency. She found that there is inconsistency in the formulation of query statements. Various searches select different search terms from the same search request. Furthermore, the same searcher formulates different query statements from the same search request on various occasions. There are more differences in the selection of search terms than in the selection of search concepts. This result is remarkable. In information retrieval we have to operate with search terms. The search concept cannot be expressed without form, and it is in the area of form where there is a lot of inconsistency and variety. This variety must be recognised and acknowledged when we design information retrieval systems. (Iivonen 1995, 287, 294)

There is also inconsistency between different databases representing the same field when compared quantitatively. For example, von Ungern-Sternberg (1994) found, how Derwent Biotechnology Abstracts and Pascal Biotechnologie have used twice as many index terms per article than the Current Biotechnology Abstracts and Life Sciences Abstracts.

When studying consistency of the abstracts and subject indexing of novels (between different actors) Saarti (1999, 163) found, that public library professionals were clearly more consistent than library clients, and that among professionals experience increased consistency.

Iivonen and Kivimäki (1998) have studied the selection and representation of concepts in indexing of the same documents in two databases of library and information studies. They compared the indexing of 49 documents in KINF and LISA databases. They focused on the types of concepts presented in indexing, the degree of concept consistency in indexing, and similarities and differences in the indexing of concepts. They addressed the following research questions:

- Which types of concepts are most often presented in indexing in KINF and LISA?

- What is the degree of concept consistency when the same document is indexed in KINF and LISA? The concept consistency means the degree of agreement in the selection of concepts on various occasions.

- What kind of differences occur in the indexing of concepts in KINF and LISA? (Iivonen & Kivimäki 1998, 92)

The study showed, that there were clear differences in indexing of concepts in KINF and LISA, although the number of concepts per document used in indexing was rather similar to LISA having on average 3.6 concepts and KINF 4.8 concepts per document. 2.2 of these were common concepts represented in both databases. This result was similar to the earlier studies of Iivonen (1989)[28]. When taking into account both databases simultaneously, approximately six concepts per document were indexed altogether. The concept consistency remained still rather low the mean value being 34 percent. (Iivonen & Kivimäki 1998, 93-4, 100)

From the viewpoint of this study this statement is particularly interesting

"We can assume that the main content of a document will be indexed rather consistently although the indexing will take place in various environments for various clienteles. We can regard the indexing as a special kind of expertise where indexers know how to compress the content of a document into a few main concepts. --- " (Iivonen & Kivimäki 1998, 100)

In this study, it is assumed that languages and ideas expressed by languages are created mainly in accordance with expression needs of the surrounding culture (Suojanen 1993, 16) and thus, a thesaurus can be seen as "a cultural product" (see also 1.2 *Objectives of the dissertation*). The content of a thesaurus depends on the documentary needs of the data sources in the surrounding culture. Finnish indexers are guided by the Finnish culture and its traditions. In Finland, the most important and widely used indexing term source is the Finnish general thesaurus (*Yleinen suomalainen asiasanasto, YSA*). Indexers usually find the Finnish equivalents for English terms in YSA and thus conform to the Finnish culture and the conventions in information retrieval and storage. Translation problems of Finnish and non-Finnish concepts occur for instance on the basis of the thesaurus' structure, the differences between languages and cultures, and the time factors which influence the content of concepts. If the main content of the documents are being indexed, these above-mentioned factors – if existing - should be seen in the results.

It is noteworthy, that inconsistency in our field is not only a problem in information storage, but it is seen in all the aspects of information storage and retrieval – in production, use, retrieval and storage. Variations in concept versus term level are a considerable challenge for our field, and here also style plays a crucial role. As Ingwersen (1994, 106) has said, every document type has its independent style, which vary from domain to domain.

---

[28] In Iivonen's study ten indexers indexed the same ten documents, that is monographs, and the mean value was 3.5. See Appendix 2, tables 2.1-2.11 in Iivonen 1989, 69, 75) and Iivonen & Kivimäki 1998, 93.

Zhangs, Anghelescus and Yuans (2005) have studied the impact of domain knowledge, search behaviour and search effectiveness to each other. The results show that as the level of domain knowledge increases, the user tends to do more searches and to use more terms in queries. Nevertheless, the search effectiveness stayed the same for all participants. (Ibid.)

Lehtokangas and Järvelin (2001, 535, 546) have studied consistency of textual expression in newspaper articles and found, that the mean value of consistency calculated on the basis of words was 65 % (for short news wires 83 % and for long articles only 47 %) , whereas at the concept level consistency was considerably higher, ranging from 92 % to 97 % between short and long articles. They conclude, that an "effective theory of information retrieval (IR) should take into account, in addition to statistical properties of document texts, their linguistic and conceptual properties, as documents are purposive units of human communication". (Ibid.)

- **Summary and Conclusions**

In this study indexing is studied from the semantic and human point of view with the help of discourse analysis and the focus is foremost on translatability issues. Indexing is understood as an intellectual process, in which identifiers are assigned to a new entity, i.e. keywords or indexing terms into a record for describing the content of the indexed document. Ideally, indexing should cater for a variety of potential approaches to the subject.

Indexing can be separeted into pre-coordinated and into post-coordinated indexing. In the first mentioned style the keywords are already coordinated as separate concepts at the time of indexing, and in the last mentioned style the coordination of separate concepts is done at the time of search.

Indexing consistency refers to the degree of similarity in indexing of the same documents in different situations. The indexing consistency is greater at the main content and concept level than at the selection of separate indexing terms. (Iivonen 1989; Iivonen & Kivimäki 1998).

In LIS, aboutness is often discussed as what the document or seeking task is about. In this study, the perspective is slightly different. When constructing a multilingual thesaurus, we cannot avoid questioning what is the index(ing) term/descriptor about? Does it truly refer to the same concept in different cultures and subcultures? Is it used consistently in indexing? Or is it used at all in other cultures? - If not, why is this so? - Is there no need for a certain concept or is the topic covered from another perspective using different concepts? In this study aboutness is viewed from the perspective of semantics and pragmatics.

3.3 Thesaurus construction

Lykke Nielsen (2004, 57-74) has divided the issues and literature related to thesaurus construction into the concept of thesaurus, general writings about thesaurus construction, approaches to thesaurus construction, collection of terms and concepts, formation and definition of concepts, organisation of concepts and into automatic thesaurus construction. Although the study is not about thesaurus construction in general, nor about thesaurus construction as a process, some general and contextual background information is needed about how a thesaurus is constructed *in principle*, whereas other aspects of thesaurus construction are not considered relevant for the frame work of this study. In general thesaurus construction guidelines, the emphasis is on standards, given by ISO and SFS, since they are the authorities in the (Finnish) field, and therefore a relevant source for comparisons with empiricism. Concerning equivalence issues Finnish guidebooks are also studied.

According to Hjørland (2002, 425) "thesauri are mainly domain specific vocabularies, and the methodology of designing them can also be seen as one (implicit) form of domain analysis." He further concludes, that

> "--- indexing and retrieval information is always specific. Main stream IS has, however, largely ignored the way different domains may put different demands on systems for organising and retrieving documents. A stronger focus on different domains may make our field more realistic and our masters more relevant in different environments. Such research might benefit by co-operating with, among others:
>
> producing special classifications and thesauri:
>
> bibliometrical studies;
>
> epistemological and critical studies; and
>
> terminological studies, LSP (language for special purposes), discourse studies." (Ibid, 430)

According to Aitchison et al. (1997) the primary purpose of a thesaurus is for information retrieval, which may be achieved in various ways. Furthemore, the primary use may be achieved by using the thesaurus in the indexing (intellectual or automatic) of a database, and/or in its searching, by various ways:

- thesaurus used both in indexing and in searching
- thesaurus used in indexing, but not in searching
- thesaurus used in searching, but not in indexing. (Ibid, 1)

This view also represents how the thesaurus is seen in this study.

The Finnish SFS 5471 –standard (1988) *Guidelines for the establishment and maintenance of Finnish language thesauri* is commonly and broadly accepted in practice and many Finnish thesauri are reported to be constructed according to it, such as the YSA (Helsingin yliopiston kirjasto 1999), Kielitieteen asiasanasto (Kotimaisten kielten tutkimus tutkimuskeskus 2003), Puolustushallinnon asiasanasto

(Puolustusvoimat 2000) and Hoidokki – Hoitotieteellinen asiasanasto (SHKS 2005). The Finnish standard follows the international ISO 2788 –standard *Documentation - Guidelines for the establishment and development of monolingual thesauri*.

The other central standard is ISO 5964 (1985) *Guidelines for the Establishment and Development of Multilingual Thesauri*. Concerning its motivation and usefulness, which is also in line with the motivation for this study, it is stated that:

> "--- Information systems are expanding across language boundaries, leading to a notable increase in the provision of indexing and retrieval tools which are either language-independent --- or multilingual. Aids of this kind are essential if retrieval of documents indexed in more than one language is not to depend on the acquisition and use of a single, dominant language. Indexers or searchers should, where possible, be able to work in their mother tongues, or at least in a language with which they are already familiar. Within this context it is considered that multilingual thesauri have a significant part to play in improving the bibliographic control of literature on a global scale.
>
> The standardization procedures for the construction of a multilingual thesaurus is seen as a primary step in achieving compatibility between thesauri produced by indexing agencies using terms selected from different natural languages. The recording of these procedures will also enable indexers engaged in this task to benefit from the experience of others, and to work in a logical and consistent fashion, using recommended practices which have been established in the course of discussions at an international level." (Ibid, 1)

The ISO 5964-1985 standard should be used in conjunction with ISO 2788 (the Finnish equivalent of which is the SFS 5471) and the standard for multilingual thesaurus construction should be regarded as an extension of the scope of the monolingual guidelines. It is also noteworthy, that although the examples given in the standard represent three major languages (English, French and German), it does not imply that these languages are regarded as dominant in the field of documentation by ISO. Furthermore, it is also admitted, that the procedures described in the guidelines may not be entirely appropriate for all languages, but the procedures, considered in the standard, relate to problems which may be encountered in any language. (Ibid, 1)

First the general principles of thesaurus construction, provided in standards, are briefly introduced. Second, the relationships in thesauri are given. Third, the special case – construction of multilingual thesauri, is discussed.

3.3.1 General principles

As discussed before, thesauri are controlled vocabularies. The ISO 5964-1985 standard represents two principal means for achieving vocabulary control, employed in thesauri:

> "terms are deliberately restricted in scope to selected meanings. Unlike the terms in a dictionary, which may be accompanied by a number of different definitions reflecting common usage, each term in a thesaurus is generally restricted to whichever single

meaning best serves the needs of an indexing system. The structure of a thesaurus, notably its display of hierarchical relationships, frequently indicates the intended meaning of a term. If this technique is not sufficiently explicit, a definition or scope note should be appended to the term. This should state the chosen meaning, and may also indicate other meanings which are recognized in natural language but which have been deliberately excluded for indexing purposes;

when the same concept can be expressed by two or more synonyms, one of these terms is usually selected as the preferred term (---) which is then used consistently in indexing. Reference to the preferred term should be made from any synonym which might also function as a user's access point. To assist the user of a printed thesaurus, it is recommended that preferred terms should be distinguished typographically from non-preferred terms." (Ibid, 5)

In the ISO standard 2788-1986 *Documentation – Guidelines for the establishment and development of monolingual thesauri* two different kinds of methods of compilation are introduced – the deductive method and the inductive method, which are both likely to be employed during the construction of a thesaurus. When applying the deductive approach:

" --- terms are extracted from documents during the preliminary stage of indexing, but no attempt is made to control the vocabulary, nor to determinate relationships between terms, until a sufficient number of terms has been collected. All terms are then reviewed by a group of specialists, preferably consisting of both indexers and subject experts. These should first identify terms which represent categories, and the remaining terms should then be allocated to these categories on the basis of their logical relationships, so that the categories tend to be established on a broader-to-narrower term basis. Vocabulary control should be applied at the stage where categories are established --- " (Ibid, 29)

The inductive method is introduced as:

"With the inductive method, new terms are admitted into the thesaurus as soon as they have been encountered in documents. Vocabulary control is applied from the outset, and each term, as it is admitted, is designated as a member of one or more broader categories constructed on an ad hoc basis at an early stage. The thesaurus is therefore established on a narrower-to-broader term basis. Thesaurus construction is regarded from the outset as a continuous operation, and although assistance from technical experts should be sought as necessary, they need not be involved as members of a formal editorial group or committee." (Ibid, 29)

In the Finnish national standard regarding the construction and maintenance of a Finnish language thesaurus (SFS 5471-1988, 8) the construction of a thesaurus is divided into five sections: tentative work; subject fields; actors of the work; conclusion of the structure and parts of thesaurus; stages and timetables.

In the tentative working stage, the already existing thesauri and controlled vocabularies are explored and their usability for possible exploitation is studied. In addition, the need for the new thesaurus should be justified and also it should be accertained whether it is necessary to construct a structured list of indexing terms or if references SEE and SEE ALSO would be sufficient. (Ibid, 8)

In the following, the key areas or subjects fields of the thesaurus will be defined and their scope and the specificity of indexing terms will be evaluated. The constructors will be named – a person answerable for construction, or a working group and the group in charge. The work of the working group will be organised and the experts, term collectors, test users of thesaurus and evaluators will be find out, after which the plan will be checked in order to construct the list of indexing terms into thesaurus form, and then the decisions about thematic parts of thesaurus and about division and update of the thesaurus will be made. (Ibid, 8)

In the last stage, the time table for thesaurus work is established, and the actual terminological work is completed, that is, the choice of terminological sources, the thesaurus construction and test-indexing. In the actual working stage the selection criteria of terms and term forms are confirmed by following the principles. The actual working stage also includes: refining the terms, language check of indexing terms, reconciling preferred and non-preferred terms to the chosen indexing terms and deciding on the term relationships; grouping the terms into larger unities; negotiating between the working group and experts; documenting the decisions and checking the indexing terms according to them; test-indexing in the working group. (Ibid, 8) The construction of a thesaurus demands a great variety of expertise – about the subject field of the planned thesaurus, thesaurus construction principles – and about indexing.

In the SFS 5471 -standard terminological sources, as well as other information, are the thesauri of the field and of the fields close to it, controlled vocabularies and classifications systems with their indexes, the study books of the field and of the fields close to it, reference books, professional magazines and journals, standards etc. index publications of the field and their indexes, the indexing terms provided by the experts and future users of the thesaurus, indexing terms used in the indexing of the field's documents. (Ibid, 8)

(Construction of multilingual includes also other kinds of decision-makings, see chapter 3.3.3 *Multilingual thesauri*.)

3.3.2 Terms and relationships in thesauri

Thesaurus construction is guided by international and national standards. National standards are constructed in accordance with international standards (ISO 5964-1985, 1). The thesaurus construction standards stipulate the use of three basic inter-term relationships, which are the equivalence relationship, the hierarchical relationship and the associative relationship. (Aitchison & al. 1997, 47; ISO 2788-1986, 13) The relationships are regarded as language-independent and also culture-independent, and therefore they apply equally to both monolingual and multilingual thesauri (ISO 5964-1985, 25, see more in Hudón 2001, 74-75).

Equivalence relationship (author note: as a basic relationship within one language in a thesaurus – not to be mixed here with translation equivalence between different languages) refers to synonyms and quasi-synonyms. In indexing all the members of an

equivalence set are regarded as referring to the same concept. One of the terms in the set is designed as the preferred term, and the other(s) as non-preferred. When all the languages in a multilingual thesaurus have equal status, then every preferred term in one language should be matched by a corresponding term in the other language(s). Since languages frequently differ in the number of synonyms which refer to a given concept, an exception for one-to-one demand are non-preferred terms, for which it is not necessary to establish correspondences in different languages. (ISO 5964-1985, 26)

Hierarchical relationships cover two different kinds of logical relationships – the generic relationship and the hierarchical whole-part relationship, which applies only to certain classes of terms, such as geographical entities and parts of the body. It is the basic relationship which most obviously distinguishes a systematic thesaurus from other organised lists of terms. (ISO 5964, 26)

ISO 5964-1985 sees, as a general rule, that any hierarchy which the users of one language regard as logically acceptable should appear to be equally valid when its terms have been translated into another language (ibid, 26). Hierarchical relationships can cause translation problems – or reveal them:

> "It can further be asserted that if a hierarchy is accepted as "natural" by the users of one language, and this hierarchy is not, for cultural reasons, regarded as equally valid by the users of another language, then it may not be possible to achieve acceptable equivalences at the level of the individual terms. Such fundamental differences between the categorical systems of different language users would tend to indicate that the terms in these languages either refer to different concepts, or they express the same basic concepts from such different viewpoints that the hierarchy expressed in the source language cannot be translated, as it stands, into the target language. Faced with this situation, the term which appears to belong to different categories, depending upon the language or culture, should be treated as polyhierarchical term (---) : that is, it should be admitted as a subordinate term in each of these different hierarchies, and both hierarchies should then be translated into the other language."

> "Translation of a hierarchy from source to target language will sometimes reveal a gap in a hierarchy in one of the languages. If the target language already possesses an acceptable equivalent to the "extra" term in the source language, and provided that this term is regarded as useful by either of the language communities, then the "missing" term should be added to the hierarchy in the target language." (Ibid, 26)

> "- If an acceptable equivalent does not exist, this should be treated in accordance with the recommendations given about equivalence degrees and their solutions" (ibid, 27).

The associative relationship is considered to be the hardest one to define (ibid, 27):

> "It refers to the link between two terms which do not form an equivalence set, nor can they be organized as a hierarchy, yet they are associated in common usage to such an extent that the user who refers to one of these terms in an index or thesaurus should be directed to the other term. It is necessary to exercise a measure of control over terms linked by this relationship. It can be stipulated, for example, that one of the terms should form a necessary component in any definition or explanation of the other." (Ibid, 27)

According to ISO 5964-1985 associative relationships in multilingual thesauri should be represented one-to-one between different language versions, and due to cultural reasons, there should be generosity in providing them, but at the same time they should not serve the needs of one (cultural) group.

> "Before the associative relationship which has been recognized in one language is transferred to another, it should be examined to determine the extent of its validity. If it appears to apply to only one group of language users, it should generally be excluded. In that case, the terms should also be re-considered to ensure that they do, in fact, refer to the same concept. Despite this injunction, a multilingual thesaurus should usually contain a richer variety of associative relationships than a monolingual thesaurus in the same field, since it will benefit from the viewpoints of different language users." (Ibid, 27)

When selecting or evaluating preferred terms to a thesaurus it is noteworthy, that according to the SFS 5471-1988 the selected descriptor should be as clear, short and unambiguous as possible, and its meaning must be the same as that generally used in the field of thesaurus. When focus on multilingualism the guidelines towards loan words and translation of loan words is also interesting. The ISO 2788-1986 states:

> "Terms from other languages are sometimes encountered as "loan words", i.e. foreign terms which are accepted as newly-coined terms. If these terms are well-established, they should be incorporated into the thesaurus. Occasionally a loan word and a putative translation co-exist. If the loan word is more widely accepted, it should be treated as the preferred term, but if the translation becomes well-established, this should be preferred. Reciprocal references should be made between the preferred and non-preferred terms.

> Example:

> X-RAYS

> **UF** roentgen rays

> Roentgen rays   **USE**   X-RAYS" (Ibid, 7-8)

Here, one should also keep in mind, even in the case of a citation loan that the same concept must exist in both languages. YSA (1.4.2009), for example, includes terms such as *aerobic, afasia* (aphasia), *balalaika*, *bingo*, *bulimia*, *calvados*, *rock, punk, reggae* etc since they are established concepts and words in Finnish common and/or specific language. Sometimes the citation loans are adapted into the Finnish language system by different transliteration, (e.g. *cyberpunk USE kyberpunk*). In translation science is often repeated, that these kinds of terms may represent false friends – terms which look the same but have different meanings.

In the guidelines (see e.g. ISO 2788-1986, 30) it is stated that the correctness of terms should be verified before they are admitted to thesaurus:

> "The following types of authority should be checked when terms are accepted as candidates for inclusion:

> standard technical dictionaries and encyclopaedias;

> existing thesauri;

classification schemes.

Subject specialist, especially those with some knowledge of indexing and documentation, may also be consulted." (Ibid.)

Equivalence relationship is according to ISO 2788-1986

"--- relationship between preferred and non-preferred terms where two or more terms are regarded, for indexing purposes, as referring to the same concept. In a printed thesaurus, preferred and non-preferred terms should be distinguished typographically whenever possible. Reciprocity is expressed by the following conventions:

USE, written as a prefix to the preferred term;

UF (use for), written as prefix to the non-preferred term. (Ibid, 13)

The general equivalence relationship covers two types of terms, namely synonyms and quasi-synonyms. Terms considered as synonymous are those, "whose meanings can be regarded as the same in a wide range of contexts, so that they are virtually interchangeable". They are more common in a controlled indexing language, where meanings are more restricted, on purpose, than in the natural language. In practice, there are various types of synonyms: terms of different origin (e.g. polyglot vs. multilingual); popular names and scientific names (aspirin vs. acetylsalicylic acid); common nouns and trade names (vacuums flasks vs. thermos flasks); variant names for emergent concepts (hovercraft vs. air cushion vehicles); current or favoured terms versus outdated or deprecated terms (developing countries vs. underdeveloped countries); variant spellings (geese vs. goose); terms originating from different cultures sharing a common language (lifts vs. elevators); abbreviations and full names (PVC vs. polyvinyl chloride); and the factored and unfactored form of a compound term (coal + mining vs. coal mining). In the case of synonym, preferred terms should be selected to serve the needs of the majority of users, bearing in mind the general guidelines for indexing term (form) selection. Predictability should be taken into account, and the selected principle (e.g. popular names rather than scientific names) should be used consistently. (Ibid, 14)

Quasi-synonyms are defined as "terms whose meanings are generally regarded as different in ordinary usage, but they are treated as though they are synonyms for indexing purposes" (ibid, 14). It is also pointed out, that quasi-synonyms are likely to be least common in an indexing language, which covers several disciplines, and terms should be treated as quasi-synonyms only in fringe subject areas –they should not be used as a means for reducing the number of preferred terms in an indexing language. (Ibid, 14)

In the context of equivalence relationships upward posting as a third case is discussed. Upward posting "refers to a technique in which the name of a class and also the names of its members are treated as an equivalence set, the broader term then functioning as the preferred term". This technique is sometimes used to reduce the number of terms in an indexing language, but it should generally be avoided. "If employed, it should be applied only to terms in the fringe area of the subject field covered by the thesaurus". (Ibid, 14-15)

According to ISO 2788-1986 -standard **hierarchical relationship** is the one, that most distinguishes a systematic thesaurus from an unstructured list of terms, such as

those from a glossary or dictionary. "It is based on degrees or levels of superordination and subordination, where the superordinate term represents a class or whole, and subordinate terms refer to its members or parts". It is expressed with the abbreviation **BT** (i.e. broader term), written as a prefix to the superordinate term and **NT** (i.e. narrower term), written as prefix to the superordinate term. The hierarchical relationships cover three logically different situations – the generic, the hierarchical whole-part and the instance relationship. It should be pointed out, that every subordinate term should refer to the same basic kind of concept as its superordinate term, i.e. both the broader and narrower term should represent a thing, or an action, or a property, etc. (Ibid, 15)

ISO 2788-1986 states, that **associative relationship** is easier to define in terms of negative rather than positive characteristics. Furthermore,

> "It covers relationships between pairs of terms which are not members of an equivalence set, or can be they be organized as a hierarchy in which on term is subordinated to another, yet they are mentally associated to such an extent that the link between them should be made explicit in the thesaurus, on the grounds that it would reveal alternative terms which might be used for indexing or retrieval. This relationships is reciprocal, and is indicated by the abbreviation "**RT**" (related term), or its equivalent in other languages. (Ibid, 17)

ISO standard (ibid) puts stress on the rule that thesaurus constructors should exercise strict control over the choice of terms linked in this way, and avoid subjective judgements.

> "As a general guideline it can be stated that one of the terms should be strongly implied, according to the frames of reference shared by the users of an index, whenever the other is employed as an indexing term. More specifically, it will frequently be found that one of the terms is a necessary component in any explanation or definition of the other, to the extent that the term "birds", for example, forms a necessary part of the explanation of "ornithology". (Ibid, 17)

Using associative relationship, two kinds of terms can be linked - those that belong to the same category and those belonging to different categories. Terms belonging to the same category "relates to siblings with overlapping meanings, such as "ships" and "boats", where each of the terms is amenable to an exact definition, (consequently they do not form an equivalence set), yet they are sometimes used loosely and almost interchangeable, so that the user seeking documents on one of the terms should be reminded of the other. Terms belonging to different categories are of various kinds, but they should satisfy the requirements that one of the terms should be strongly implied by the other. As representative examples of typical relational situations encountered in practice is offered: a discipline or a field of study and objects or phenomena studied (forestry RT forests); an operation or process and its agent or instrument (temperature control RT thermostats); an action and the product of the action (weaving RT cloth); an action and its patient (harvesting RT crops); concepts related to their properties (poisons RT toxicity); concepts related to their origins (Dutch RT Netherlands); concepts linked by causal dependence (bereavement RT death); a thing and its counter agent (plants RT herbicides); a concept and its unit of measurement (electric current RT ampere); and

syncategorematic phrases and their embedded nound (model ships RT ships). (Ibid, 17-19)

Not only do hierarchical and equivalence relationships give valuable information especially from the translatability's and (sub)cultural context's point of view but also the associative relationships provide valuable information about the informants' conception of the studied concepts; they are thus also in focus but mainly as a contextual and semantic information. Recognition of associative relationships may be very cultural- and group-bound action, and that is the reason why the creation of an associative thesaurus (Term Association Thesauri, TA-Thesauri) is considered to be especially demanding, although worth the effort (see e.g. Iivonen1989, 176). Associative relationships are important (Kristensen & Järvelin 1990, 15) or equally or more important than hierarchic relationships (Iivonen 1989, 176-178; Tudhope et al. 2001) for expansion of the query and improving recall. Furthermore, as Hudón (2001) has stated,

> "A critical function of the thesaurus is that of helping users in making sense. The defining function of thesaurus relationships is well established: Relationships "define," admittedly not always very clearly, by providing a context of sorts that determines the place of a concept in its semantic environment, and by supplying more or less explicit information on the intension and extension of this concept. Relationships also serve as navigational aids to arrive at the intended destination or at the most interesting destination on a semantic map. In a multilingual context, both the defining and navigational functions of thesaural relations acquire even more importance." (Ibid, 69)

Since related terms are recommended for use in the expansion of the query and improving recall and also often seen as culturally biased, they can give very valuable information about the content and context of the terms in a specific context and are thus also very important for the study of multilingual thesaurus construction. It is of importance to see how they are represented within the search results of this study, that is, in the selected general and social science thesauri (about material see esp. 6.3.2 *Thesauri* and about results esp. 9.1 *Semantic lexical networks in thesauri*). The hierarchical relationships also provide similar kind of information; but otherwise, the focus is on equivalence relation (due to focus of the study).

3.3.3 Multilingual thesauri

> "**multilingual thesaurus**: A thesaurus (---) containing terms selected from more than one natural language. It displays not only the interrelationships between terms, but also equivalent terms in each of the language covered." (ISO 5964-1985, 3)

In this subchapter, the general aims of multilingual thesauri is discussed as well as the methods of construction, which are provided in the standards, and also, how equivalence is considered - possibly defined – and operationalised - in thesaurus construction literature. Accordingly to the perspective of the study (- translatability into *Finnish* language and culture) the emphasis is on the literature that guides Finnish thesaurus construction.

Hudón (2001) has illustrated the history of multilingual thesauri as target-language biased, and not including a true multicultural design, and incorrectly reflecting an idea of cultural neutrality:

> "The first multilingual thesauri were developed rapidly, preserving the standardized structure of the monolingual thesaurus, and using the same design principles. Much emphasis was initially put on compatibility of structures across languages. Strong compatability resulted from full correspondence of concepts and relations, while weak compatability resulted from full correspondence between concepts but not necessarily between conceptual relations. Multilingual thesauri were often built by translating an existing monolingual thesaurus, most likely one in the English language. Equal status of languages did not appear to be major concern, and little feedback from the target language – which could potentially lead to a modification of the source language – was allowed. A most unfortunate consequence of this approach was that cross-language equivalences were forced where they did not exist (e.g., one source term, no target term), or were ignored or eliminated where they did exist (e.g., one source term, two or more target terms), and questionable relational structures were established. Although many individuals advocated more flexibility in relational structures and true cultural representativeness, the social, cultural, and political considerations related to any manipulation of natural language were brushed aside, and existing practices were little affected. Although thesaurus designers and users who were not native English speakers believed that the linguistic problems had to be more important than the organizational ones in the construction of a multilingual thesaurus, no official proposal for an alternate model was ever made. The particularities of the multilingual thesaurus with regard to issues of conceptual overlap and differences in relational structures have not been studied in depth, as though multilingual thesauri were language-neutral and culture-neutral objects." (Ibid, 68-69)

3.3.3.1 General principles and aims of multilingual thesauri

ISO standard 5964 *Guidelines for the Establishment and Development of Multilingual Thesauri* recognises three approaches to the construction of multilingual thesauri:

> 1.Ab initio construction, i.e. the establishment of a new multilingual vocabulary without direct reference to the terms or structure of an existing thesaurus;
>
> 2.Translation of an existing monolingual thesaurus;
>
> 3.Reconciliation and merging of existing thesauri into two or more working languages. (ISO 5964-1985, 6)

When constructing a multilingual thesaurus there are naturally more matters to decide, than in a monolingual thesaurus work. In the context of management issues ISO standard 5964-1985 notes that one has to define the status of every language used in the thesaurus (ISO 5964-1985, 6). It is further stated and also typographically strongly emphasised that no language should be dominant:

"It is sometimes necessary, on practical grounds, to designate one of the languages as exchange language., i.e. the language which is used as a medium for exchanging indexing data (for example in a multilingual network). Even in these cases, however, it should be possible for indexers and users to use their local languages for indexing and retrieval. All the languages should be regarded as having equal status from the viewpoint of thesaurus construction. The imposition of a dominant language upon other language users is **NOT RECOMMENDED**". (Ibid, 6)

Aitchison et al. (1997) does not consider multilingual thesaurus construction as being much more difficult than monolingual thesaurus construction. The main challenges lie in work organisation and languages belonging to different "language forms", which are:

"Inflectional languages, such as Latin, which use case-endings. These root-suffixes qualify the noun and verbs.

Agglutinative languages, such as Turkish, Finnish and Hungarian where the root-suffixes can, and regularly do, stand as separate words.

Isolating languages, such as Chinese, which make no use of inflection, agglutination or prepositions.

Analytical languages, such as English, which use word order, auxiliary words and some vestiges of inflection to provide the grammatical structure." (Ibid, 135)

Hudón (2001) states, that "[i]n the global information world, the multilingual thesaurus performs three well-defined functions:

- "Allowing individuals to use the language that they feel most at ease with so they can formulate queries as simply and intuitively as possible,

- Providing interpretation support to access information within documents written in a foreign language, and

- Facilitating the integration of information provided in various languages." (Ibid, 68)

She also notes, that for "the multilingual thesaurus to be most useful and efficient in these tasks, its relational structure should reflect the multiple ways of "seeing" the world that its multilingual and multicultural users bring with them when querying an information system. (Ibid, 68)

When listing relevant standards (ISO 5964 – 1985, which should be used in conjunction with ISO 2788) in the context of multilingual thesaurus construction, Aitchison & al. (1997) note, that

"This can be seen to be sound advice when it is considered that the majority of descriptors are nouns (---) and, more importantly, that thesaurus compilation is concerned with the identification of the concepts behind the words. (Ibid, 135)

It is further stated that although basic rules should be discussed in the light of linguistic variations, it is more important to have "a good working knowledge of the languages being handled, including the socio-cultural nuances, particularly present in non scientific subjects" (ibid, 135).

Naturally, the composition and action of working group also partly differ from the monolingual ones. In ISO 5964-1985 standard it is stressed that "[c]lose cooperation, at the international and multilingual level, is essential for the successful construction of a multilingual thesaurus" (ibid, 32). Three kinds of organisational structure are introduced:

> "a) **a centralized structure**, in which all decisions are taken by a central agency. Other agencies contribute terms and suggestions, but do not have a decision-making function. This form of structure allows fast decisions, but it also involves a danger that the views of the other cooperating agencies are not taken sufficiently into account;
>
> b) **a decentralized structure**, in which each of a number of cooperating agencies assumes responsibility for selecting and interrelating the terms which fall within its own language and/or subject areas;
>
> c) **a semi-centralized structure**, in which the work is controlled by a central editorial committee consisting of delegates from the various cooperating agencies. This committee organizes all aspects of cooperation between the agencies, controls the allocation of work, etc., and serves as the final authority in all intellectual and editorial matters." (Ibid, 32)

It is further stated, that a decentralised structure is generally regarded as least effective, and a semi-centralised structure as the most effective. "Whichever type of structure is chosen, its terms of reference should apply not only to the initial stages of thesaurus construction, but should also hold throughout the continuous operations of revision and updating." (Ibid, 32)

The construction of a multilingual thesaurus always necessitates multilingual cooperation. Whichever of the above mentioned working method is chosen, no monolingual part can answer for the terminological work. According to the ISO 5964-1985 standard it is essential "that editorial decisions concerning the terms of a given language should be fully endorsed by a native speaker of that language" (ibid, 33). Concerning the involvement of subject and language specialists is it also stated that:

> "The need for help from language and/or subject specialists may arise at any time during the construction of a multilingual thesaurus. These specialists need not be engaged on a full-time basis, but the information scientists responsible for the work should be able to consult an appropriate specialist in circumstances such as the following:
>
> defining terms and establishing their interrelationships;
>
> selecting the preferred term when synonyms are encountered;
>
> adopting loan terms, or coining new terms." (Ibid, 33)

As already stated in chapter 1.1 *Background* multilingual thesaurus construction faces several serious problems. The conceptional systems are not necessarily successfully translatable into foreign languages representing different cultural systems, and the transfer may lead into meaningless expressions (Hudón 1997). A common way to create a multilingual thesaurus is translating an existing monolingual thesaurus (cf. ISO 5964, and also results in chapter 9.3.1 *Construction practices*). Hudón (1997)

considers a monolingual thesaurus as being culturally biased, and therefore a straight translation might lead to a form of "cultural imperialism". She reminds us that:

> "It is useful to remember at all times that there is more to multilingual thesaurus development than finding equivalents for concepts and terms. There is definite cultural dimension to the process, and in fact it might soon be more appropriate to refer to multicultural thesauri, rather than multilingual thesauri. There is also a political dimension to multilingual thesaurus construction, especially in dealing with languages, which are not, contextually, on the same "standing". Canada, for example, has a good grounding in multilingual thesaurus construction, but it remains struggle to make sure that French (the minority language) and English are given equal treatment in the many thesauri designed and used in the country." (Ibid.)

Hudón (ibid.) also states that designers of multilingual thesauri face many substantial challenges and obstacles, of which some are of an administrative nature, some of a linguistic or semantic nature, and some are related to technology. She further (2001) reminds us, that in order to avoid the traditional problems, "it has been recommended that multilingual tools be built up from the ground up, starting with distinct banks of terms (one for each language represented) and developing distinct structures through semantic relationships, with immediate help from native speakers of each one of the thesaurus languages" (ibid, 69).

Doerr (2001) also states that even though semantic heterogeneity of terminological resources has frequently been referred to as a problem, a systematic analysis of its intellectual basis and structure has not been carried out. According to Doerr, translated thesauri are thesauri, "--- where each concept is optimally interpreted in words of another or multiple languages, to allow speakers of those languages to understand better and use concepts of this thesaurus more effectively". He also stresses that such translations are in general not established indexing terms of the target language. (Doerr 2001)

Nonetheless, controlled vocabularies reduce linguistic problems even in a monolingual searching environment and the benefits are the same – or even greater - in a multilingual environment. According to Milstead and Feldman (1999) metadata attacks three language problems that cause poor precision: polysemy, synonymy and ambiguity. When operating with a foreign language, these problems are even more difficult to solve without any vocabulary or terminological help.

> "The larger the information domain, the more important is to find an effective and efficient way to define narrower domain for searching. One of the major causes of false hits in retrieval is homographs, that is, words that look the same but have different meanings. The advantage of searching within a specific domain is that terms are often ambiguous across several disciplines, but seldom have multiple meanings within a particular discipline or subject domain." (Chan, Lin, & Zeng 2000, 188)

Thesauri usually include scope notes to define or clarify the meaning and to guide the use of ambiguous terms in that particular context, and thus help the indexer and seeker to select a proper equivalent for their purposes. Different virtual libraries, such as the Finnish Virtual Library (*Virtuaalikirjasto*) and the British SOSIG (*Social Science*

Information Gateway) also aim to restrict information retrieval to narrower and more relevant domains.

As earlier stated, a thesaurus can be seen as a cultural product, reflecting its origin. The same also applies on a broader level as well. Hjørland and Pedersen (2005) state:

> "In general, one could say that a specific interest (say that of Scandinavian public libraries) should lead to the design of systems, which are optimal given the interest or purpose and which do not just lead to the acceptance of implicit values inherent in systems that are designed, for example, for commercial purposes. A system designed for a Scandinavian public library should not, for example, tend to identify American commercial websites but should be optimised to identify pages reflecting their cultural and democratic values and purposes." (Ibid, 586)

3.3.3.2 Equivalence in multilingual thesaurus construction

The study concentrates on equivalence as a translation problem between different (sub)languages and (sub)cultures. In monolingual thesaurus work equivalence is generally understood as Townley & Gee (1980, 25) represent it (see more in 3.3.2 *Terms and relationships in thesauri*):

> "--- *Equivalent terms*. These are not necessarily terms literally of equal value but in a given case they may, in the normal habit of the particular thesaurus users, be treated as if they meant the same thing. Where two (or more) terms are used as if they mean the same as each other, normally only one will be used as a descriptor and will be shown in the thesaurus as a 'preferred term' or as *used for* the other ---" (Ibid.)

The most important standards in the construction of a Finnish language version of a multilingual thesaurus are SFS 5471 and ISO 5964. Their most relevant and central ideas for the purpose of this study are discussed next.

In the Finnish SFS 5471 –standard *Guidelines for the establishment and maintenance of Finnish language thesauri* concerning a selected descriptor it is stated that it should represent the concept most aptly, it should be as clear (*selkeä*), short and unambiguous as possible. The meaning of the descriptor should be the same as usually in the thesaurus domain. (SFS 5471-1988, 2) Concerning words of foreign origin it is stated, that if a native word has been established it should be selected in place of a foreign word (e.g. *leirintä* in place of *camping*). If a foreign term does not have an established native equivalent, a foreign word will be selected as a descriptor (e.g. *hakkerit*). In thesauri meant to be used within a certain profession descriptors should primarily represent the expressions from the thesaurus users' jargon. (Ibid, 4)

In the Finnish SFS 5471-standard equivalence is not discussed separately, but concerned in the subchapter 5.2 "*Toisiaan vastaavat termit*" (terms which reciprocate each other). According to the standard (SFS 5471) words, which refer to the same or almost to the same concept can be used as descriptors (*asiasanoina*) in place of each other. The word which is considered to refer to the concept most exactly and which is most commonly known among the users of the thesaurus will be selected as a preferred

term and the others will be marked as non-preferred terms. A descriptor can also represent an opposite concept or a superordinate concept (*yläkäsite*), if needed. (SFS 5471-1988, 5)

This guideline can be understood as a recommendation to select the word that most exactly refers to the concept. If there is more than one exact equivalent then the most well-known word among the thesaurus users should be selected as a descriptor. In special thesauri the well-known aspect among thesaurus users' is emphasised. The aimed equivalence degree is thus not necessarily exact and the usability aspect can (and sometimes should) be prioritised.

The terms that can be used in place of each other are, for example, terms differing from the origin (e.g. *raakaöljy – maaöljy*); common nouns and trade or brand names (e.g. *telekopiointi – Telefax*); the old and new form of the term (e.g. *kotimyynti – suoramyynti*); words of foreign origin and their native equivalents (e.g. *demografia – väestötiede*); abbreviations and the unabridged forms (e.g. *TEL – työeläkelaki*); word pairs representing contrariness (e.g. *työttömyys – työllisyys*); and terms having narrower or broader meanings (e.g. *ennakkosensuuri – sensuuri*). (SFS 5471-1988, 5-6).

ISO 5964- 1985

In the ISO 5964 –standards Documentation- Guidelines for the establishment and development of multilingual thesauri regarding equivalence it is stated:

> "Due to the nature of language itself, terms selected from more than one natural language vary in the extent to which they represent the same concepts. These variations can be regarded as forming a continuum, one end of which is represented by terms which can, for the practical purposes of indexing, be regarded as exact equivalents, further points being marked by various degrees of partial or inexact equivalence, and the final point being represented by those extreme situations in which a term in one language refers to a concept which cannot be expressed by a single, direct and equivalent term in another language. For practical purposes in this International Standard, these graduated distinctions have been organized into five relatively broad categories. These are set down below in order of increasing complexity or difficulty." (ISO 5964-1985, 7-8)

What is especially noteworthy in the above statement is that languages are considered as inevitably varying, which is in correspondence with the ideas represented in modern translation studies (cf. later in chapter 3.4 *Translatability and equivalence*).

In the ISO 5964 –standard equivalence degrees are divided into five groups, and set down in order of increasing complexity or difficulty. The first case is **exact equivalence**:

> "The target language contains a term which is
>
> a)   identical in meaning and scope to the term in the source language;
>
> b)   capable of functioning as a preferred term in the target language." (Ibid, 8)
>
> "Terms from different languages which refer to the same concept should be treated as exact equivalents. Exact equivalents can be morphologically related.

Example:

English   French   German

PHYSICS   =   PHYSIQUE   ....=   PHYSIK

- or they may be morphologically unrelated:

Example:

English   French   German

BLACKBIRDS   =   MERLE   =   AMSEL

- or they may appear to express the same concept from different viewpoints:

Example:

English   French

SOFT DRINKS   =   BOISSON NON ALCOOLISÉE" (Ibid, 11)

Exact equivalence is seen as the analogue of the true synonym, and it is dealt with straightforwardly (Aitchison & al. 1997, 136). It is noteworthy, that in this case is assumed, that an exact equivalent in meaning still might not work as a preferred-term in a target culture.

The second case is the least discussed case in the ISO standard, and like case 1, it is also simply to represent in a thesaurus – but it may not be easy to recognise it to differ from the first degree. According to Aitchison & al. (1997, 136) inexact equivalence is the analogue of the near synonym, and it is also regarded (for indexing purposes) as being equivalent.

The second case is called **inexact equivalence**. In it a term in the target language expresses the same general concepts as the source language term, although the meanings of these terms are not precisely identical. (ISO 5964-1985, 8)

"Examples:

German   French

GEDECK   ....=   MENU" (Ibid, 11)

In the above case solution is that "[t]erms which differ only in connotation should be treated, for indexing purposes, as exact equivalences". (Ibid.)

The third equivalence degree is called **partial equivalence**:

"The term in the source language cannot be matched by an exactly equivalent

term in the target language, but a near translation can be achieved by selecting a term with a slightly broader or narrower meaning". (Ibid, 8)

"This situation covers terms which are generally regarded as referring to the same concept, but one of the terms strictly denotes a slightly broader or narrower concept." (Ibid. 11)

Partial equivalence is similar to inexact equivalence, but in partial equivalence one of the terms denotes a slightly broader or narrower concept. The two terms can be treated as synonyms or as broader-narrower term pair. (Aitchison & al. 1997, 136)

As an example, ISO 5964-1985 (p. 12) gives the German word *Wissenchaft* and the English word *Science*, and provides two solutions. First, the preferred choice, is to treat both terms as exact equivalents for indexing purposes.

Example:

German	English	French
WISSENCHAFT	SCIENCE	SCIENCE (Ibid. 12)

The second choice is to treat the terms as a loan term in the other languages, and organise them hierarchically.

Example:

German	English	French
WISSENCHAFT	=WISSENCHAFT	= WISSENCHAFT
	SN Loan term adopted from German	**NE** Emprunt de l'allemand
	NT SCIENCE	**TS** SCIENCE
SCIENCE	SCIENCE	SCIENCE
D Lehnwort für SCIENCE	**BT** WISSENCHAFT	**TG** WISSENCHAFT (Ibid, 12)

The fourth case is **single-to-multiple term equivalence**, where the term in the source language cannot be matched by an exactly equivalent term in the target language, but the concept to which the source language term refers can be expressed by a combination of two or more existing preferred terms in the target language (ibid, 8). This situation is considered to require a two-stage explanation. Firstly,

"Certain objects or phenomena are named specifically in both the source and the target languages, and these concepts are usually (but not invariably) regarded as members of classes which are also named in both languages. (Ibid, 16)

Secondly,

"In addition, a term in the source language refers to an extra category which has not evolved, for cultural or linguistic reasons, in the target language." (ibid, 16)

ISO 5965-standard provides three kinds of solutions. Generally preferred is to devise a combination in the target language to include the members of the extra category recognised in the source language. Usually a scope note is needed to explain why this combination is necessary. (Ibid. 16) Example,

German	English	French
SCHENKE	=SLUGS + SNAIL	=ESCARGOT +LIMACE
	SN Use this combination as an equivalent to the	**NE** Combination équivalente á l´allemand SCHENKE

German term

SCHNEKE

UB GEHAUHESCHENKE **NT** SNAILS   **TS** ESCARGOT

UB NACKTSCHENKE **NT** SLUGS**TS** LIMACE (Ibid. 17)

It is noted, that this solution is feasible only when the combination consists of a limited number of terms, i.e. four or fewer. If there are more terms, the combination becomes too unwieldy for the purposes of indexing and retrieval. (Ibid, 17)

In the second solution, the thesaurus constructor should assign the status of a non-preferred term to the name of the extra category in the source language, and treat this term as an equivalent to whichever term in the target language is usually regarded as closest in meaning. This can be done by using either a broader or a narrower term. It is noted, that this solution allows the target language to retain all its own terms, but it also involves some loss of retrieval capability in the source language. (Ibid, 17)

Third solution is to adopt the categorical name occurring in the source language as a loan term in the target language, and add a scope note to this term where its meaning is not self-evident. (Ibid, 17)

Aitchison, Gilchrist and Bawden (1997), also following the ISO 5964 –standard, consider single-to-multiple equivalence to be the most complex situation. It covers three kinds of different situations, which demands unique solutions. In the first sub-case a term does not exist in another language but two or more narrower terms do and together these two narrower ones can be used to cover the broader single concept. The second sub-case is similar to the first, but it can include factoring. In the third sub-case two generic terms exist in one language and only one in another, and it is not clear which of these generic terms would be the most appropriate. (Ibid, 136-137)

The fifth case is **non-equivalence**, when target language does not contain a term which corresponds in meaning, either partially or inexactly, to the source language term. Non-equivalence is rarely encountered in indexing. (Ibid, 8, 18-19) [29] Two kinds of examples are given:

"a) a term in the source language expresses an abstract and frequently culture-dependent concept which, at least initially, is unknown to the users of the target language(s).

Example:

German   English

BERUFSVERBOT =   ?

b) newly-developed processes, operations or equipment, notably in the sciences or technology, are named in the language of their inventors, and have not yet acquired vernacular names in the other languages.

---

[29] Finnish equivalents given by Meriläinen (1996, 87-88) for ISO equivalence degrees: Case 1 *täydellinen vastaavuus*, Case 2 *epätäydellinen vastaavuus*, Case 3 *osittainen vastaavuus*, Case 4 *yksi-moneen -vastaavuus*, Case 5 *täydellinen vastaamattomuus*.

Example:
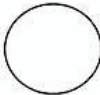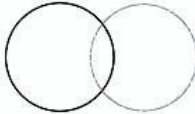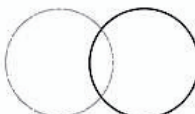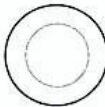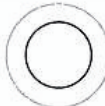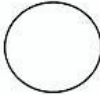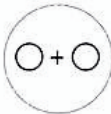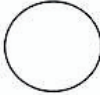
English   French

STEAM CRACKING   =   ?" (Ibid, 18)

The given solutions are not considered a preferred solution and it is stated, that in practice the choice will depend upon a number of factors:

> "in particular on the extent to which the source language term can be "translated", using existing terms in the target language which are also suitable for indexing and retrieval. If the "foreign" concept can be expressed exactly and without difficulty by a combination of existing terms in the target language, this should be considered as an example of single-to-multiple term equivalence (---), not as a case of non-equivalence." (Ibid, 18)

The first solution is to adopt the source language term as a loan term in the target language, defining this loan term in a scope note. The second solution is to coin a term in the target language which expresses the meaning of the source language term. This coined term may amount to a putative "translation" of the original term, or it may be an artificial invention. (Ibid, 19)

To summarise, the degrees of equivalence are illustrated as:

Case	Source language	Target language
1 - Exact equivalence		
2 - Inexact equivalence		
3 - Partial equivalence		
4 - Single-to-multiple equivalence		O + O
5 - Non equivalence		
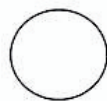
acceptable term exists

acceptable term does not exists

Figure 6: *Degrees of equivalence in ISO 5964-1985 (ISO 5964-1985, 9)*

In the ISO 5964-1985 standard it is also recognised, that - in rare circumstances – for terms involving more than one kind of translation problem it is necessary to proceed in stages when translating the difficult term. These stages might involve more than one of the five categories of problems discussed previously. (Ibid, 19-20.) Other language problems named are choice and number of preferred terms, intra- and inter-language homographs, place names, proper names of institutions and persons, and abbreviations and acronyms. (Ibid, 20-25)

Hudón (2001, 70) has criticised the idea of representing equivalence the way standards and many textbooks do – in them "the identification of cross-lingual equivalents is considered relatively unproblematic, and pragmatic procedures appear sufficient to cope with differences in interpreting relationships in various languages". She further states, that

"---- Such artificialization, realized in part through the manipulation of interpreting relationships within natural language, is deemed acceptable in the context of designing a tool to access an information system: Equivalence relations, whether applied within or across languages, are relations between terms considered to have equal value for information retrieval and for information retrieval only.

The solutions proposed to the most common cases of inexact cross-lingual equivalence may lead to a belief that these problems arise from some gap in the lexicon rather than differences in the deep semantic structure of languages. We can simplify the surface structure of the multilingual thesaurus, but underlying conceptual overlap and conflicts remain. Cross-lingual equivalents that look the same and are good translations of one another, but that do not in fact represent the same concept, are a frequent occurrence in thesauri." (Ibid, 73)

Meriläinen (1997) has studied the concept of equivalence in ISO 5964 –standard and stated, that

"When considering the equivalence concept found in the ISO 5964 in light of the linguistic discussion on equivalence, it seems evident that the descriptor equivalence is an instance of dynamic equivalence. This conclusion is based on the fact that the thesauri include also such equivalencies which could never be tolerated in natural language use. These equivalence relationships depend on the function of thesaurus and are based on non linguistic arguments." (Ibid, 110)

In the LUASPORT-project, which was about "Compatability of the Finnish Liikunnan ja urheilun asiasanasto[30] with the Canadian Sport thesaurus with a special reference to descriptor equivalence", Meriläinen (1997) divided equivalence from three different kinds of context – thesaural, interlingual, and indexing context - into formal equivalence and into two kinds of dynamic equivalence. It was a new method to take all these three perspectives into consideration within one study. (Ibid, 107-109)

In LUASPORT formal equivalence is understood "very narrowly as a correspondence of descriptors on the level of spelling with or without correspondence to the meaning of the descriptors". Descriptor equivalence is understood accordingly as "correspondence on the semantic level with or without correspondence on the formal level" and it is further divided into dictionary equivalence and indexing equivalence. (Ibid, 110) Furthermore,

"Dictionary equivalence between descriptors is a relationship which enables the substitution of one descriptor with the other when translating from one natural language to another – in LUASPORT from Finnish to English – and which can be verified by general-use translation dictionaries. ---" (Ibid. 110)

"Dictionary equivalence has principally only two instances: either it exists or it does not exist. ---" (Ibid, 111)

"Indexing equivalence between descriptors means that they denote the same or almost the same concept and that they may be used as substitutes when translating from one indexing language to the other --- Indexing equivalence is either potential or existing. Equivalence is potential, when it is determined solely by comparing descriptors in the

---

[30] Author note: The controlled vocabulary in the field of sport.

context of the studied thesauri with no reference to their actual use in indexing. Potential indexing equivalence may or may not yield to existing indexing equivalence, when the descriptors have actually been used in indexing the same document and they are determined to name the same or almost the same concept in the document in question. Dictionary equivalence between two descriptors increases the probability of indexing equivalence, but is not its perquisite. Determining the existing indexing equivalence is a process of comparing formal and semantic characteristics of the descriptors starting from possible dictionary equivalence and ending with conceptual analysis." (Ibid. 111)

The tentative results of the LUASPORT-study demonstrated that descriptor equivalence between the source and target thesauri was rather high (70-75 %), when it is measured with dictionary equivalence or with potential indexing equivalence, but it was much lower (30-35 %) when measured with existing indexing equivalence. (Ibid. 111)

What kind of relevance do formal and dictionary equivalence have in practical work? - Should thesaurus constructors aim at dictionary equivalence? One problem is that dictionaries and thesauri differ greatly as text types and in their function, and furthermore, both can be considered as dynamic, but the dynamic nature is not identical. The LUASPORT project provides valuable viewpoints on the concept level. More research about combination of both the terminological and concept level, and also development of more detailed analysis tools would, however, be needed.

One may ask whether there is a literal Finnish thesaurus construction tradition or any basis of such? How is equivalence discussed in Finnish guidebooks? There are not that many guidebooks concerning thesaurus matters in Finnish. The TSK's terminological know-how has been published in The Handbook of Terminology Work (*Sanastotyön käsikirja*, TSK 14; a joint publication of TSK and SFS, The Finnish Standards Association). The handbook concerns practical terminology work and was published in 1989. A more recent guide is the Guide to Terminology (published in 1997) providing a concise introduction to practical terminology work.

In The Handbook of Terminology Work (TSK 1989) equivalence is discussed as correspondence (FI: "*vastaavuus*"): Terms representing different language correspond with each other when referring to exactly the same concept. In multilingual terminology work it is especially important to define the correspondence of the terms. (TSK 1989, 73)

The correspondence is further divided into exact concept-correspondence (*täydellinen käsitevastaavuus*), partial correspondence (*osittainen vastaavuus*) and to narrower and broader concept-correspondence (*laajemman ja suppeamman käsitteen vastaavuus*). In exact concept-correspondence the concepts are identical and terms referring to them refer to the one and only concept. Then harmonising in the concept level is not needed. In partial correspondence the meanings or domains differ. If the differences are irrelevant or minor it is recommended to form only one concept. (TSK 1989, 102)

A concept in one language can correspond in other language with a group of several concepts, or a concept can exist only in one language. If one concept

corresponds with several concepts in other language, it is recommended to consider if the concept should be adopted into the other language. (TSK 1989, 103)

In the guide to terminology (Haarala 1981), correspondence and equivalence are discussed through synonymy, homonymy and polysemy. It is stated that usually in special language (lingo) terminology work the aim is that one concept refers to one term and one term to one concept. In practice, the principle "one term - one concept" is considered to be an ideal, which should be aimed at but only within certain limits. In natural languages – that is also special vocabularies – there exist three common phenomena, which break the rule for one-to-one concept and term, namely synonymy, homonymy and polysemy. (Ibid, 39)

Synonymy is used to refer to the phenomenon when different expressions (*ilmaus*) are conceptually identical, that is, when several expressions refer to one and the same concept. Homonymy is the opposite – one expression refers to more than one different concept. Homonymy is especially disturbing when within the same domain the homonymous concepts are needed in close connection with each other. (Haarala 1981, 39-40)

Polysemy is similar to homonymy in respect that also in polysemy one term is related to two or more concepts. In polysemy, however, the concepts are closely related to each other and usually polysemy occurs because of the broadening of the meaning of an expression. In common language polysemy is a very normal phenomenon and the relation between the concepts is based on figures of speech, and metaphors. (Haarala 1981, 40)

- **Summary and Conclusions**

Thesauri can be seen as representing normative-descriptive vocabularies. They guide language usage, but also mirror the practices. – They are not artificial in a sense, that a thesaurus constructor should not start to create his/her own neologisms nor a new kind of network of vocabulary, but a thesaurus should reflect its users' thinking and practices. When a thesaurus is constructed to be a tool used in indexing and in information seeking, one of the most important criteria is its inter-coherence and predictability. The function and main construction principles are the same for mono- and multilingual thesauri.

In the terminology guides and standards, equivalence and correspondence do not necessarily refer to absolute one-to-one –relationships between terms and concepts. They can vary from the exact equivalence to non-equivalence. Thus in guides and standards it is also admitted to be unrealistic to aim at exact equivalence systematically, both between languages and inside a language due to the characteristics of natural languages.

From the perspective of translation strategies and possibility for different thesaurus types, equivalence is discussed very briefly and sometimes even superficially in the thesaurus construction standards. The rules and norms that guide the construction of

multilingual thesauri remind mainly a domestication strategy. The constructor is supposed to be loyal to customs and the language usage norms of the target language users. In standards, multiculturalism is not directly discussed, but the idea is seen especially in the demand that terms (and translation equivalents) should be accepted by native speakers.

Standards suggest every term has to be translated somehow – it is better to accept imperfect equivalence, than leave a term without any translation. Multilingual thesaurus is considered to have equivalents in one-to-one principle among different languages, and there is neither discussion nor guides for different kinds of solutions, where the contents of the different versions might diverge.

3.4 Translatability and equivalence

"*one can say the same thing in many ways*" (Nida & Taber 1969, 49)

"*The place of practice and of theory, for the translation of any text, is the place of its practice.*" (Meschonnic 2003, 344)

Equivalence both as a topic and as a concept is very problematic. In terminology guides and standards it is often discussed quite superficially, whereas in translation studies it is discussed widely and sometimes with conflicting results. In translation theories equivalence is considered to be one of the most polemic concepts. In general equivalence means correspondence, but as such it does not communicate anything. Saksa (2004, 170) states, that there is also no consensus about *how* to study and define equivalence, but there is an agreement, that it should be done *cross-disciplinarily*.

This chapter discusses what translatability and equivalence are from the perspective of multilingual and –cultural thesaurus construction. Firstly, the art of translation and more general views about equivalence is briefly discussed, secondly how is equivalence understood in communicative translation theories, and thirdly the previously discussed equivalence matters are summarised together with ideas represented in thesaurus construction guidance literature.

3.4.1 The art of translation

The art of translation meant, for Horatius and Cicero, careful interpretation of the source text and composition of source text according to the principle *non verbum de verbo, sed sensum exprime de sensu* (*not word for word, but thought for thought*). This means that a translator is responsible always towards the target audience. (Bassnett 1995, 62)

Translation can be seen from many different perspectives and is already defined and illustrated differently within translation science, as for example:

"Translation may be defined as follows: the replacement of textual material in one language (SL) by equivalent material in another language (TL)." (Catford 1965, 20, cited here Schäffner 1998)

"Translation is a kind of activity which inevitably involves at least two languages and two cultural traditions." (Toury 1978, 200)

"The translation would not simply be a question of linguistics. One should start translating not only words, but also concepts and even contexts." (Carmen 2004, 2)

Not only translations, but also the theories surrounding it, are – or should be - context-bound:

"Any definition of anything is theory-bound, so there is no such thing as a totally objective definition of "translation" that we can take for granted before we start studying it, as there will never be any definition of translation that will be all-inclusive. We start with a preliminary working definition, and refine it as we go along. Different scholars, with different research aims, tend to start (and end up) with different definitions. We should aim to be as aware as possible of why we choose or accept a particular definition and/or conception of translation." (Chesterman & Arrojo 2000, 152)

"--- This phenomenon, whereby a theorist makes global observations on translation in general, but actually means only one, often narrow area of it, still characterizes translation studies today – to the detriment of a general theory of translation." (Snell-Hornby 1988, 14)

"--- The practical and theoretical specificity of translation varies according to the specificity of the language practice to be translated. The place of practice and of theory, for the translation of any text, is the place of its practice." (Meschonnic 2003, 344)

In the next section translation is discussed from the perspective of its function, equivalence and strategy. Thesaurus construction and translation represent the contextual perspective used in this theoretical section.

3.4.1.1 Function of translation

According to Ingo (1984), translation studies have relied very much on linguistic philology when a text has been considered a static object and the translation process itself has not been explained. At various times translation (as a process and action) has been understood in many different ways. Demands and purposes have been passionately polemised as long as there have been translations. The ancient Cicero and the renaissance Dante both considered translation as impossible. If translating is understood as transferring information or a message, i.e. meaning, from a source language into a target language, translating is nowadays seen as generally possible. (Ingo 1984, 1-5)

So-called mirror translations are, by contrast, considered to be impossible: the relation between a symbol, that is a word, and the concept to which it refers to is arbitrary and conventional; in different languages experiences are segmented differently

into linguistic symbols (e.g. in some African languages the whole colour spectrum is covered by three words); the grammatical structures in different languages diverges; language users share different experience worlds and surroundings (concrete objects, religious and moral conceptions). (Ingo 1984, 1-4) In addition to languages, individual language usage also differs, because language users are always creative individuals. Lehtonen (2000) states, that within a culture there are different ways to classify and organise reality and each mirrors the status and perspective of its constructors. Meanings also change over time and may even become contradictory to the original one. (Ibid, 25. See also chapter 3.1 *On culture, language and meaning*.)

According to Oittinen (1995), opinions around *equivalence* are strongly disunited. It is considerably more common to question the concept of equivalence than to explicitly define it. When examining the axiom of translation Oittinen (ibid.) asks: If we consider human being as an interpreting person – and a translator is a human being – all kind of categorising is difficult: if a translation is already an interpretation of the source text, and the translation is not the same as the source text, what is then a modification ("*muunnelma*")? What then does correspondence and faithfulness mean? How do we then define a good translator and a good translation? (Ibid, 18)

Oittinen (1995) states that perhaps the most stubborn axiom associated with translation is the correspondence between the source text and the target text, that is, equivalence as it is traditionally understood. Translating is, therefore, understood as a doomed action, where a translator is aiming at the unattainable sameness. Only during the past two decades have translation researchers questioned the aim of attaining sameness between translation and source text. (Ibid, 31-34)

Oittinen (1995, 146) considers translating as an open and visible manipulation. Traditionally, a good translator is considered to be invisible and humble – a good translation is loyal to the source text, and a translator can be seen as an active manipulator, who aims not at sameness but stretches his/her words to the future target-language readers. Oittinen's idea about translation as manipulation is related to the idea of language and power and how language does not reflect the reality in a passive and neutral way (cf. e.g. Mälkiä & Stenvall 1997; Lehtonen 2000; Koskinen 2000).

Lehtonen (2000) states, that translating can be considered to be a form of re-writing ("*uudelleenkirjoitus*"). Languages are not nomenclatures, but each language produces its own system of differences, and therefore a completely corresponding, exact-equivalent translation is impossible. Texts, the sources of translations, are not same always and for all readers; readers produce texts while reading them. In this sense both translating and reading are re-writing processes. (Ibid, 66)

Although there seems currently to be a general consensus that there is no point in demanding "mirror-translation" translating today is still not seen as an easy act. A translator is commonly seen as "a prisoner" of his/her own culture. Nida and Reyburn (1981) found that a translator usually understands the message in accordance with his /her own cultural-linguistic context. Usually, a translator is aware of that and tries to solve the problem by using foreign terms. Regardless of this, a concept in the source language is not always semantically equivalent with the same concept in the target

language. Generally descriptive phrases are therefore better (more equivalent) than foreign terms in translating a message into another culture. (Nida & Reyburn 1981, 21-25) This can also be seen in thesaurus construction standards and guidelines where loan words are not recommended (in e.g. ISO 5964-1985, ISO 2788-1986, SFS 5471-1988).

3.4.1.2 Idea of equivalence

The early demands made on translations do not conform to the nature of language, and that is also the main trend seen in modern translation theories, terminology guides, and in thesaurus standards. The division of equivalence types in standards and guides and in translation theories the proportion of equivalence to the function of translation is based on this understanding. This study concentrates on the examination of the function of the translation, and of the equivalence type derived from focus. Therefore, outside the scope of this study are the divisions of translation equivalents, which are based on the formal or literal equivalence, such as, e.g. morfem-to-morfem translations.

When studying equivalence it is essential to understand the **limits of equivalence** and to accept these limits as **unavoidable**.

> "Every language is a self-contained system and, in a sense, no words or constructions of one language can have absolute equivalents in another. The idea that there might be some linguistic elements which are universal in the sense of having absolute equivalents in all the languages of the world is of course all the more fanciful." (Wierzbicka 1991, 10)

This kind of uniqueness does not mean a finality, since as Wierzbicka (1991) further states:

> "However, as soon as we abandon the notion of absolute equivalents and absolute universals, we are free to investigate the idea of partial equivalents and partial universals; and if the former notion is sterile and useless, the latter idea is fruitful and necessary." (Ibid.)

More radical critic is also presented, such as:

> "--- the term *equivalence*, apart from being imprecise and ill-defined (even after a heated debate of over twenty years) presents an illusion of symmetry between languages which hardly exists beyond the level of vague approximations and which distorts the basic problems of translation." (Snell-Hornby 1988)

Since no better conceptual tools were found for the study's purposes, the concepts and terms 'equivalence' and 'equivalent' are used in this study. The author, however, wishes to point out, that when speaking of equivalence in general, it is not a question of exact-equivalence, but of similarity or correspondence, and in a more detailed focus it is the **nature** of that similarity (or correspondence). As Zethsen (2004) states, focus "is now on the receiver of the translated text and the skopos of the translation does in most cases mean non-equivalence with certain aspects of the source text. Equivalence is thus not an absolute, but relative concept and functionality, not complete equivalence, is considered the sound aim of the translator." (Ibid, 126-127)

In order to discuss fruitfully partial equivalents, we need conceptual tools and a better understanding of how to do it. Equivalence can be studied from the more structural and linguistic viewpoint (as done in chapter 3.3.3.2 *Equivalence in multilingual thesaurus construction*) and from the more contextual and functionalist viewpoint, as Koller (1989) subsequently does. Koller's model is chosen for further analysis in the empirical section, since its perspectives are considered as practical and relevant also from the thesaurus construction perspective. – Often in translation science the emphasis is on the literature of fiction, and when placed on informative texts, it remains focused on the literature. Dictionary-based translation theories are also considered irrelevant here, since the function of dictionaries differs greatly from thesauri, and since a thesaurus term is always closely related to the whole thesaurus, which is a kind of text that differs from dictionaries.

Koller (1989) lists various categories of equivalence, which makes the definition of the concept 'equivalence' complicated:

> "--- content equivalence (often also: content invariance), stylistic equivalence, formal equivalence, functional equivalence, textual equivalence, communicative equivalence, pragmatic equivalence, equivalence of effect ---" (Ibid, 100)

He has represented an analysis attempting to specify the concept of equivalence more precisely (and at the same time also consider the various categories).

> "(a) The concept of equivalence postulates a relation between SL text (or text element) and TL text (or text element).[31] The concept as such does not say anything about the *kind* of relation: this must be additionally defined. The mere requirement that a translation should be "equivalent" to a given original is vacuous.
>
> (b) The kind of equivalence relation is defined in terms of the frame and the *conditions* to which one refers when using the concept of equivalence. In other words, a *normative* statement is made: there exists equivalence between a given source text and a given target text if the target text fulfils certain requirements with respect to these frame conditions. The relevant conditions are those having to do with such aspects as content, style, function etc. The *requirement* of equivalence thus has the following form: *quality* (*qualities*) *X in the SL text must be preserved*. This means that the SL content, form, style, function, etc. must be preserved, or at least that the translation must seek to preserve them as far as possible." (Ibid, 100)

In other words, when determining the type of equivalence one should express what kind of conditions a translation has to fulfil in order to be equivalent. Koller (1989) separates five factors, which indicate equivalence types and a further five demonstrating translation types. If two words are *true cross-linguistic equivalents*, they agree in all the five levels (factors) (see also Koller 1995; Zethsen 2004, 127). In the specifications for equivalent types the five factors are: extralinguistic content, connotations, text and language norms, receiver and certain formal-aesthetic features of the SL texts. (Koller 1989, 100)

---

[31] Author note: SL=source language, TL=target language

If equivalence is oriented towards *extralinguistic* content it represents ***denotative*** equivalence. ("Terms commonly found in the literature are "invariance in content" or "invariance at the content level"). (Ibid, 100)

> "Translation as the *achievement of denotative equivalence* sets translation theory the task of describing the potential equivalence relations between any two languages, together with the textual factors that determine the choice of a given equivalent in any specific case. Correspondences of different types (one to many, many to one, one to zero, one to part) need to be analysed in order that the translation process can achieve referential identity between SL and TL units. The central area of concern here is the *lexicon* (the words and syntagma of the language), since it is here that languages are (or should be) at their most productive (particularly regarding the use of existing or new methods of word formation), in order to account for ever-changing and expanding communication needs and aims. From the translation point of view, it follows that denotative equivalence is in principle attainable, even though the language may not always be very economically used in attaining it. "In principle" means disregarding the other factors which may play a role in translation (readability and comprehensibility, the receiver, the connotative and formal value of the text, etc.)" (Ibid, 101)

Correspondences of different types listed above are rather similar with the ones represented in thesaurus construction guides and standards (as illustrated in chapter 3.3.3.2 *Equivalence in multilingual thesaurus construction*). When translating thesauri the invariance in content, the extralinguistic, has a key position. To aim at reproducing the semantic invariant in content is important in thesaurus translations, but how far can a translator disregard the stylistic factors, traditional in thesaurus construction, such as favouring short and well-known terms? How much can one favour denotative meaning and ignore the connotative meaning especially within social sciences? - Is it more a problem for translators, indexers or information seekers? In order to consider these questions several types of equivalence should be studied.

Focusing on the connotations transmitted by means of word choice, with respect to level of style, the social and geographical dimension, frequency etc. leads to ***connotative*** equivalence type. (Koller 1989, 100)

> "With respect to translation equivalence, the term connotative indicates that individual expressions in the textual context, and also complex texts themselves, do not only have a denotative meaning; according to the specific means of linguistic expression of the denotatum, additional values are also transmitted --- A single denotative meaning can be expressed in various ("synonymous") ways:
>
> > eat : dine : nosh ---" (Ibid, 101)

Koller (1989) lists several connotative dimensions, relevant for translation:

> (a) connotations of speech level (e.g. elevated, normal)
>
> (b) connotations of socially determined usage (e.g. student language, military usage)
>
> (c) connotations of geographical relation or origin (e.g. non-regional, American English)
>
> (d) connotations of medium (e.g. spoken language, written)
>
> (e) connotations of stylistic effect (e.g. euphemistic, descriptive)

(f) connotations of frequency (common, uncommon)

(g) connotations of register (e.g. normal usage, medical)

(h) connotations of evaluation (e.g. positively evaluative, pejorative)

(i) connotations of emotion (e.g. emotive, neutral). (Ibid, 102)

According to Koller (1989) the achievement of connotative equivalence is one of the hardest problems in translation, and can seldom be absolute (ibid.). Thesaurus constructors and therefore also thesaurus translators aim at neutral, descriptive and precise equivalents (about thesaurus terms see chapter 3.3 *Thesaurus construction* and 3.2.1 *Indexing process, languages and tools*).

Koller calls the third equivalence type **text-normative** equivalence:

> "Legal contracts, instructions for use, business letters, scientific texts and the like all follow lexical and syntactic norms of both selection and usage (i.e. norms of style); to translate in accordance with these norms is to aim at *text-normative equivalence*. ---" (Ibid, 102)

Thesauri designed for indexing and information seeking represent a certain type of text, the construction and usage of which is guided by standards and manuals. However, in practice there is great variation in the representation style and norms.

The fourth equivalence type is ***pragmatic*** equivalence. When aiming at pragmatic equivalence the focus is on the receiver (reader), to whom the translation is directed and to whom the translation is "tuned" in order e.g. to achieve a given effect. (Ibid, 100)

> "The concepts of "usage norm" and "functional form" introduce a pragmatic point of view: in observing the usage norms for particular texts one takes account of the linguistic/textual expectation norm, the expectations that the reader brings to a given type of text. --- the achievement of pragmatic equivalence, then means translating the text for a particular readership --- This may – or even must – result in deviating from the requirements of text-normative, connotative or even denotative equivalence. A translation of a political commentary which sought to persuade the original readers to a particular political action usually has a different function in the target language, and addresses its readers with different presuppositions. ---" (Ibid, 103)

The pragmatic equivalence as represented by Koller (1989) is perhaps the most foreign from a thesaurus construction perspective and not so simply "translatable" into a thesaurus construction context. When translating a manual and/or introduction for the thesaurus users, a translator might think of different readerships and, for example, formulate the motivations to use the thesaurus differently depending on whether (s)he is thinking about Finnish versus British researchers, or students versus experienced researchers. In the social sciences, the pragmatic viewpoint has also another kind of significance – a thesaurus is not meant to irritate its users by including terms that are not commonly accepted by its users or representing terms in a strange context. Therefore, what are the expectations of a thesaurus text and how do the expectations possibly vary? In this study, pragmatic refers primarily to the usability aspect – is the translation version accepted and usable as the original source one. (But is the original one actually accepted and usable?)

The fifth type Koller specifies as **formal** equivalence and admits to it being a heterogeneous concept and adds (in parenthesis), that "the literature also refers in this sense to "artistic-aesthetic equivalence", "expressive equivalence" etc, especially with respect to the translation of poetry". To achieve this in a target language text "is to produce an "analogy of form" in the translation, by exploiting the formal possibilities of TL or even by creating new forms if necessary". (Ibid, 101, 103)

In a thesaurus context this means questions such as - should every thesaurus term have e.g. a hierarchical context? Should the source and the target relationship be identical in translations?

Finally, Koller points out that translation always involves a necessity of a choice and therefore a translator should establish a hierarchy of values to be preserved in the translation and from this (s)he can derive a hierarchy of equivalence requirements for the text or segment in question.

> "This in turn must be preceded by a *translationally relevant text analysis*. It is an urgent task for translation theory – and one on which no more than some preliminary work has so far been done – to develop a methodology and conceptual apparatus for this kind of text analysis, and to bring together and systematize such analyses in terms of translationally relevant typologies of textual features." (Ibid, 104)

It is important to keep in mind that texts and functions do differ. Even in thesaurus construction contexts situations may vary so much, that there is no point in giving strict rules about what to aim at and *how* to do it.

3.4.1.3 Domestication or foreignisation – or internationalisation

When translating something, a novel or a thesaurus, there are also more general than translation strategies to consider and to choose, than those previously presented. In translation studies the two basic strategies are domestication (*kotouttaminen*) and foreignisation (*vieraannuttaminen*), which are in practice exclusive. Lindfors (2001) summarises the idea behind the basic strategies:

> "Translating a text from one culture to another usually requires that a choice is first made between two basic strategies, domestication and foreignisation. Domestication means making the text recognizable and familiar and thus bringing the foreign culture closer to the reader in the target culture, while foreignization means the opposite, taking the reader over to the foreign culture and making him or her feel the cultural and linguistic differences. This choice between domestication and foreignization is linked to questions of ethics, too: should the translator be accountable to the source or target culture, and to what extent? If target-cultural conventions are followed in the translation process, the text will be readily acceptable in the target culture, but it will inevitably lose some of the characteristics that would have given it a foreign or even exotic feeling." (Lindfors 2001, 6)

Translating is not a value-free action and choices are made at all the stages of the process: what to translate, to whom, how etc.

"Translation is often regarded with suspicion because it inevitably domesticates foreign texts, inscribing them with linguistic and cultural values that are intelligible to specific domestic constituencies. This process of inscription operates at every stage in the production, circulation, and reception of the translation. It is initiated by the very choice of a foreign text to translate, always an exclusion of other foreign texts and literatures, which answers to particular domestic interests. It continues most forcefully in the development of a translation strategy that rewrites the foreign text in domestic dialects and discourses, always a choice of certain domestic values to the exclusion of others. And it is further complicated by the diverse forms in which the translation is published, reviewed, read, and taught, producing cultural and political effects that vary with different institutional contexts and social positions." (Venuti 1998, 67)

Venuti (1995, 306) states, that translating should never aim to remove dissimilarities between different cultures entirely. According to him the routineness of fluent domestication has influenced the British and American cultures that are "aggressively monolingual, unreceptive to the foreign, accustomed to fluent translations that invisibly inscribe foreign texts with English-language values and provide readers with the narcissistic experience of recognizing their own culture in cultural other".

When considering multilingual and multicultural thesauri the basic assumption is that the different language versions should work in their own linguistic-cultural surroundings, e.g. the English version of a thesaurus in England and the Finnish version in Finland, and also cross-culturally so, that e.g. a Finnish information seeker could make searches in an international database in Finnish and still retrieve in documents written in English and indexed in English by a British indexer. We can also easily imagine a situation, where an indexer would need a British concept in Finnish, such as when indexing a British-English book into a Finnish-Finnish catalogue. Needs and expectations for multilingual and –cultural thesaurus vary, as well as strategies to construct them. The problem is that the needs and expectations are not necessarily carefully considered nor clearly expressed, and what makes the situation even more challenging is that the cultural aspects are also not directly observable, and that is why we need a multi-dimensional perspective of culture as well as thesauri.

Hall (1982) has made an extension to the Iceberg Theory, in which he represents a third level, which sometimes lies above, and sometimes below the waterline. He termed the theory a 'Triad of Culture'. In the Triad of Culture the three levels of culture are: technical culture, formal culture and informal culture or out-of-awareness control. (See also Hall's triad in figure 4 in chapter 3.1 *On culture, language and meaning*.)

Technical culture is communication "at the level of science, that which can be measured accurately, and has no meaning outside it".

"Let us take the word 'time'. 'Time' has a variety of meanings, depending on context and culture. Technical time, on the other hand, refers only to the technical understanding of the concept and can be broken down into its 'isolates' and analysed. One of its basic isolates is a second, which we all have a feel for. However, very few would be able to define a second. A technical 'second' has no feeling but a clear unambiguous scientific definition (CED):

A second is the basic SI unit of time: the duration of 9 192 631 770 periods of radiation corresponding to the transition between two hyperfine levels of the ground state of caesium-133.2." (Ibid, 44.)

Formal culture is "no longer objective, but is part of an accepted way of doing things. It can, and indeed is, taught. This is the culture of traditions, rules, customs, procedures, and so on." (Ibid, 45) With the third level, informal culture, "Hall means that there are no 'rules' as such. This form of culture is neither taught nor learned, but acquired informally and, even more importantly, 'out-of-awareness'. (Ibid, 46) In social science thesaurus use, the communication occurs primarily at the technical and at the formal levels of culture, but the third level is also represent in the way we react to the words, since, as Ulrych (1992, 254) [32] points out, it is at the level of connotative meaning that we judge and react to words.

According to Katan (2004, 45)

"The idea of English as an international language and the use of a standardized international technical language are attempts at making both language and culture technical. The most extreme examples of this are the artificial or auxiliary languages, such as Esperanto, which are culture-free. The fact that they are culture free may well account for their lack of success in practice." (Ibid.)

In the context of thesaurus construction this means that there are not only two, but in practice **three basic strategies: domestication, foreignisation - and internationalisation**. When writing a fictional novel the author would hardly try to avoid using culture-bound words, but should this apply to thesaurus constructors in multilingual and –cultural thesauri? One example of a thesaurus that strives to be culturally neutral is the ELSST thesaurus.

"The ELSST thesaurus will be created from the current UKDA HASSET. This will involve reducing the present hierarchies so that all cultural and institutional specificity are removed." (Miller & Matthews 2001)

A kind of a modern translation strategy is also "existential equivalence", which Koskinen (2000) found as a typical strategy within a European Union context.

"One of the fundamental tasks of the Translation Service – 'a service with a mission', as the information brochure describes it – is to safeguard the ideal of equality between languages. Within the EU context the symbolic value of translation is high. --- Especially in the case of lesser used languages like Finnish, the communicative function may often be subordinate to a symbolic function. Sometimes the primary function of the translation of a particular official document is simply to be there, to exist. Rather than just conveying a message or providing possibilities for communication, the role of the translation is then to stand as a proof of linguistic equality. For translators, traineed to believe in the communicative function of the profession, this state of affairs can be frustrating: they need to produce a monument, not a text, and have to find new motivation to act as a guardian rather than a

---

[32] "These are "the culturally or socially determined value judgements that are implicit in the semantics of a word". Cited here Katan 1999, 32. Original source= Ulrych, Margherita (1992), *Translating Texts from Theory to Practice*, Rapallo: Cideb editore.

communicator. This could perhaps be called 'existential equivalence', that is, all the language versions need to exist, any other features being rather irrelevant or at least subordinate." (Ibid, 83)

"In spite of its obvious multilingual and multicultural nature, the Commission's Translation Service has not paid much attention to the so-called 'cultural turn' that has taken place in translation studies during the 1990s. In practice, the translation policy aims at *a*cultural communication. This is partly due to the need to draft some documents so that they are applicable in all member states, and it is therefore necessary to avoid culture-specific features. But more significant is the institutional attitude that does not encourage any degree of cultural adaptation, nor perceive translators as experts in intercultural communication. There is a clear, albeit unwritten, preference for surface-level similarity, assumedly guaranteeing that the readers of the various translations all get the same message." (Ibid, 85)

All translators make decisions between foreignisation and domestication translation strategies, although they are not necessarily aware of it (Ruokonen 2004, 63). It would be useful for translators to know how others have acted in similar situations and thus know the predominant translation norms. It would also be useful to know how domestication and foreignisation are considered in translation theories, because theories can give new perspectives to translation problems and offer models of how to solve them. (Ibid, 63) For example, directions for use of different devices are usually translated in accordance with target language norms. It might be useful to discuss, which texts are usually translated according to a foreignisation strategy and which according to a domestication strategy, as well as the reasoning of translating in a certain style and the consequences of the selected strategy. It could also be useful to study translations on the basis of foreignisation and domestication, since it could open new perspectives and help to evaluate predominant practices. (Ibid, 75)

In thesaurus construction literature and practice it could also be useful to discuss these two basic strategies – domestication and foreignisation - and the third type found in the studied thesaurus context, **internationalisation**, and their basic values and implications, which have not been discussed before. This has to do with the ethics of indexing, which is not an established topic within information science as ethics of translation is part of translation science.

3.4.2 Equivalence in communicative translation theories

The basic functional, communicative translation theories are 1) Dynamic Equivalence theory and 2) Skopos-theory. In this study their influence can be seen in the perspective, hypotheses and in the research questions.

Dynamic and functional equivalence theories are on their basis very identical and sometimes used synonymously. Skopos-theory is also very closely connected to them.

3.4.2.1 Dynamic equivalence theory

The theory of dynamic equivalence was developed by Eugene A. Nida in the 1960's and it is the first communicative translation theory. The 'dynamic' refers to an idea, that a translation should cause the same reaction in the target audience as the source text does in the original context. Instead of asking "is the translation correct", we should ask "for whom". (Nida 1964/2000; Vehmas-Lehto 1999)

According to Nida (1964/2000) there are basically two different orientations in translating and thus two fundamentally different types of equivalence: formal and dynamic. If a translator aims at formal equivalence, his/her aim is to try to form the target language equivalent, as much as possible, in coherence with the elements of the original source language equivalent. Nida states, that it is usually more recommended to aim at dynamic equivalence, when the relation between the translation and its receiver is the same as the original text and the receiver of the original has. He remarks, that "--- of course, there are varying degrees of such dynamic-equivalence translations". (Nida 1964/2000, 129-130)

Nida has further defined translation as "reproducing in the receptor language the closest natural equivalent of the message of the source language, first in terms of meaning and second in terms of style." (Vehmas-Lehto 1999, 54)

Taking the case of this study as an example, according to the dynamic equivalence theory "*homemaker*" could be translated into Finnish as e.g. "*kodinhengetär*" (and in Swedish "*husande*" in a novel, to describe a person's character, similar to "a lady / a spirit of the house") or "*kodinhoitaja*" (in Swedish "*hemvårdare*", as an occupational title in a questionnaire, such as housekeeper), but what about in a thesaurus context? The study aims to answer whether the idea of dynamic equivalence is common in the studied multilingual thesaurus context.

Although formal equivalence often means word-for-word translation and the translation units in thesauri translations are thesaurus terms, we can still assume that the fundamental idea behind the dynamic equivalence is closer to the real aims also in thesaurus translations. Translations usually do not aim to express how a foreign thesaurus is constructed, but instead aim to make a functional thesaurus also in the target language and context.

Nida's theory of dynamic equivalence has been further developed and is called the functional equivalence theory. Functional equivalence theory is based on an idea that the function of the source text is the same or similar to the function of the target text. The function of the translation adapts to the source text function. (Vehmas-Lehto 1999, 70) If the function, for example, is to express 'dying' with a euphemistic expression, the English phrase "*kick the bucket*" could be translated into Finnish as e.g. "*heittää veivinsä*" or into Swedish as "*trilla av pinnen*".

Nidas ideas have also been criticised, for example, this was done rather fundamentally by Meschonnic (see Meschonnic 2003, 346-348). Meschonnic questions the idea of 'sourciers' and 'targeters':

"The problem is to react against the conception, as fallacious as it is widespread, that would oppose 'sourciers' to 'targeters', the sourciers squinting toward the source language, trying to copy it closely, the targeters looking straight ahead, as realists, toward the target language, thinking only of preserving the essential, the *meaning*. Whereas the sourciers would care for the *form*. Inessential." (Meschonnic 2003, 339)

He resumes, that

"To translate under the government of the sign includes a schizophrenia of translating. A pseudo-realism demands that the meaning alone be translated – although the meaning is never alone. It demands the illusion of the natural – self-effacing translation. It confines poetry and indeed the entire literary act to a notion of form as residue of what one believes meaning to be, generally taking the word as the unit.

The reply of poetics is that the unit of language is not the word, and can thus not be word meaning. Targeters are looking at the wrong target. Because they only know the sign. But the unit is discourse. The system of discourse.

--- The targeter forgets that a way of thinking [une pensée] does something to language, and what it does is what is to be translated. And there, the opposition between *source* and *target* is no longer pertinent. Only the result counts.

--- Whatever the languages concerned, there is only one *source*, which is what a text does; there is only one *target*, to do in another language what that text does. That is what realism is. What the targeter mistakes for realism is semiotism. A bad sign." (Ibid. 339-340)

The starting point of this study is using particular terms, but they are studied as representing different discourses. The final aim is to identify how certain groups understand different meanings and translation aims in the context of particular discourses, i.e. the system of discourse.

Functional equivalence is not a foreign concept in the LIS field, but the context is different. Within the LIS field, functional equivalence is nevertheless similarly a methodological and theoretical issue. Hjørland (1997, 51) states (and further suggests) that:

"At one point in history, astronomy and astrology were regarded as related subjects. In modern science they are obviously not. Astronomy and astrology both study planets and stars, and some of the same words may therefore occur in texts from astronomy and astrology. But this kind of resemblance or similarity does not make the two fields related. Such a view of relatedness is superficial and naive. Real subject-relatedness does not depend on perceptions of similarity, but on theoretical analysis!

No advanced indexing, subject analysis, categorization, or classification can therefore be based on common properties or similarities; rather, subjects should be interpreted in the light of the theoretical context. Similarity is therefore an unfruitful concept. What should be grouped together for the purpose of the subject analysis are documents with identical or related functions. This is one of the reasons for the success of citation indexing, where related documents can be identified without any form of linguistic similarity. The concept of similarity should therefore be replaced by the concept of functional equivalence (or isomorphism). This insight also has important implications with regard to algorithmic retrieval, since a match between query and text representation can only be based on some kind of similarity." (Ibid.)

In this research, the emphasis is, according to Nida, on ideas concerning dynamic equivalence, and on meanings and the 'targeter'. The meanings are not seen as something self-governing, independent, or constant. The context plays a crucial role when discussing equivalence matters. The starting point is – and will be throughout - that the translation units represent the discourse of thesauri. There is little point in discussing terms and meanings without a reference to the discourse where the studied phenomenon takes place.

Context is also considered when discussing translating as action, since

> "Each cultural domain, each culture-language, has its historicity, without (total) contemporaneity with the others. The Russians do not translate French in the same way as the French translate Russian." (Meschonnic 2003, 342)

When considering functional equivalence issues the aim is, accordingly to Hjørland's (1997) ideas about functional equivalence, to also take into account what is reasonable to link and compare. However, similar to translations in translation research there are no absolute true pairs to compare, since each corpus concept and discourse differ in terms of their basis.

3.4.2.2 Skopos-theory

Skopos-theory was developed by Reiss and Vermeer in the 1980's. Skopos means the aimed purpose of the translation. In Skopos-theory it is more important to fulfil the function of the translation than to translate in a certain style. However, the function of the target text is not necessarily the same as that of the source text in its original context. (Reiss & Vermeer 1986, 54-59; Koskinen 2001, 380)

In Skopos theory the translation action is stressed – translating is doing something and for a certain purpose. Vermeer (1989) describes a translation action as a particular sort of behaviour:

> "for an act of behaviour to be called an action, the person performing it must (potentially) be able to explain why he acts as he does although he could have acted otherwise. Furthermore, genuine reasons for actions can always be formulated in terms of aims or statements of goals (as an action "with a good reason"---). (Ibid, 176)

In the translation process the phase of analysis means considering the function of the texts.

> "In order to place a functionally equivalent TL text beside an SL text the translator should clarify the functions of the SL text. This may be done in a three-stage-process, which may, in principle, be carried out either by starting from the smallest textual unit and ending with the text as a whole, or by beginning with the text as a whole and ending with the analysis of the smallest textual unit." (Reiss 1981/2000, 162)

First the translator should establish the *text type* and therefore should ask – and answer – questions such as: which basic communicative form is realised in the concrete text with the help of written texts? – Is it

"a. The communication of content – information type

b. The communication of artistically organized concept – expressive type

c. The communication of content with a persuasive character – operative type." (Ibid, 163)

Thesauri are informative, but to which degree do they also represent expressive text type? (In standards and guides, style receives considerable emphasis.)

The second stage is to establish the *text variety*, "i.e. the classification of a given text according to specifically structured socio-cultural patterns of communication belonging to specific language communities". (Ibid, 165)

As regards thesauri, therefore, are they alike in different languages and cultures e.g. in style and form?

The third stage of the analysis means the analysis of style.

"Now the *text individual* is placed in the foreground. This analysis is of supreme importance, because the translators "decisive battle" is fought on the level of the text individual, where strategy and tactics are directed by type and variety." (Ibid, 166)

In thesaurus construction this means the term selection stage. After analysing the function of the text, text type and text variety, comes the phase of reverbalisation. The theory behind this is:

"The text type determines the general method for translating;

The text variety demands consideration for language and text structure conventions." (Ibid, 166)

If functional equivalence is sought during the process the mode for translating is "*according to the sense and meaning* in order to maintain the invariability in content". (Ibid, 167)

Accordingly to the Skopos-theory the Swedish wordplay "Knut satt vid en knut och knöt en knut. När Knut knutit knuten var knuten knuten."[33] could have as an English equivalent "The sixth sick sheik's sixth sheep's sick". The explanation to choose these equivalents is to conform to the function of the source text unit – pass on a "tongue twister" into another language and confirm its common language usage.

Skopos-theory has been criticised for giving too free a hand to the translators – there may be no loyality to the source text and to the writer of the source text (Koskinen 2001, 381). Therefore, what happens in the thesaurus context? – How "freely" can (or should) one translate and why is this possible?

---

[33] The literal translation would be: Knut sat in a corner and tied a knot. When Knut had tied the knot, the knot was tied.

3.4.2.3 Cultural bias

After discussing different translation strategies the problem of selecting a proper translation strategy is reflected in the perspective of cultural words and cultural implications, since when translating cultural words the translation strategy clearly plays a greater role than when translating universals. The reflection also summarises previous discussions about translation strategies that are target versus source language biased.

Newmark (1998) states:

> "I define culture as the way of life and its manifestations that are peculiar to a community that uses a particular language as its means of expression. More specifically, I distinguish 'cultural' from 'universal' and 'personal' language. 'Die', 'live', 'star', 'swim' and even almost virtually ubiquitous artefacts like 'mirror' and 'table' are universals – usually there is no translation problem there. 'Monsoon', 'steppe', 'dacha', 'tagliatelle' are cultural words – there will be a translation problem unless there is cultural overlap between the source and the target language (and its readership). Universal words such as 'breakfast', 'embrace', 'pile' often cover the universal function, but not the cultural description of the referent. And if I express myself in a personal way – 'you're *weaving* (creating conversation) as 'he's a *monologger*' (never finishes the sentence) – I use personal, not immediately social, language, what is often called idiolect, and there is normally a translation problem." (Ibid, 94)

Newmark (ibid.) also points out, that "[a]ll these are broad and fuzzy distinctions. You can have several cultures (and sub-cultures) within one language".

In a thesaurus context, idiolect is not a potential problem in the way it is when translating e.g. a poem or a novel, whereas thesauri, in all likelihood, also include cultural words. As previously stated (see chapter 3.3.3 *Multilingual thesauri*) thesauri are cultural products and therefore we can assume they to include many cultural words (and not only universals).

James (2002) has discussed cultural implications for translation and Nida's theory of dynamic equivalence in connection with Newmark's (1988) and Hervey's and Higgins' (1992) ideas. Hervey and Higgins have visualised source and target language and culture bias on a scale, as follows:

Exoticism	Cultural	Calque	Communicative	Cultural
	Borrowing		Translation	Transplantation

Figure 7: *Source and target language and culture bias (Hervey and Higgins, 1992:28, cited here James 2002)*

The model is used in the teaching of different translation strategies and to illustrate the source language (SL) and the target language (TL) bias. Exoticism is an extreme form of source language bias and cultural transplantation an extreme form of target language bias.

For example, the term "matador" could be translated into Finnish as, i.e. "matador", as "*matadoori*", as "*härkätaistelija*" (as a calque), as "*poroajaja*" (reindeer-racer) or as "*eukonkantaja*" (wife-carrier[34]). The citation loan word "matador" would be most SL biased and "*eukonkantaja*" most TL biased – the first one would reflect only the source language and culture and the last one would reflect only the target language and culture. Could e.g. the term "matador" be of use in a Finnish general thesaurus context? Although "eukonkantaja" would be the most target language and culture biased, it might be relevant in a Finnish thesaurus context e.g. in the context of machismo or popular entertainment.

Newmark (1988, 45) has summarised various translation methods and places them in the form of a flattened V diagram as follows:

SL emphasis	TL emphasis
Word-for-word translation	Adaptation
Literal translation	Free translation
Faithful translation	Idiomatic translation
Semantic translation	Communicative translation

_____
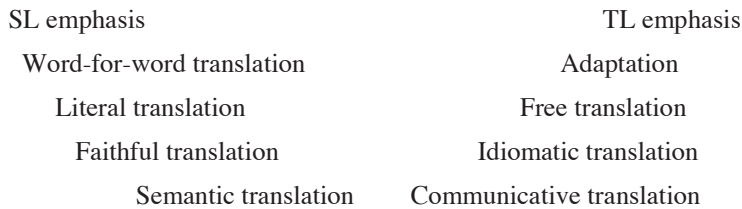
(Newmark 1988, 45)

According to Newmark (Ibid, 47) "only semantic and communicative translation fulfil the two main aims of translations, which are first, accuracy, and second, economy". When considering the nature and function of thesauri it is natural to assume fulfilling the previously mentioned main aims to be also relevant in a thesaurus context. Nevertheless, in practice, thesaurus translations operate on a word level, and if the construction method of a multilingual thesaurus is to translate an existing thesaurus **and** if the translation corpus is static, the result will inevitably be a source language bias and also call for those kinds of strategies. When considering these strategies they are thus seen in this study in more generally terms, such as an attitude or a perspective followed in the translation process. Newmark (ibid.) also says, that cultural components tend to be transferred and explained with culturally neutral terms in informative texts and replaced by cultural equivalents in vocative texts.

We can assume a thesaurus to represent both an informative text (how concepts and terms are understood and represented in a specific field and context), and in a way also a vocative text (the presented thesaurus terms should encourage the reader/user to interact, i.e. to use the terms in indexing and/or information seeking).

According to Newmark (1998)

"Communicative translation attempts to render the exact contextual meaning of the original in such a way that both content and language are readily acceptable and comprehensible to the readership." (Ibid, 47)

---

[34] Concerning wife-carrying see Sonkajärvi, <URL: http://www.sonkajarvi.fi/?deptid=15136> (last visited September 29th 2009.)

Communicative translation is also the fundamental idea behind the design of the empirical part of the study and therefore emphasis is given to different aspects of functionality, i.e. relevance and usability, and consideration is given to the opinions of thesaurus constructors, indexers and potential thesaurus users (social scientists).

- **Summary and Conclusions**

In this study, equivalence is understood in a similar way to how dynamic, functional equivalence is commonly understood in translation studies. Translating is seen as a decision-making process, where a translator often has different kinds of possibilities to choose in order to fulfil the function of the translation. Accordingly, and as a starting point for the construction of the empirical part, the function of the source text is considered to be the same or similar to the function of the target text, that is, a functional thesaurus both in source and target context. The aim of the study is to examine the idea of dynamic equivalence in thesaurus construction. The style of the equivalents is discussed, but prior to this functionality and reasoning are considered.

To conclude and summarise the main points previously discussed from the perspective of multilingual thesaurus equivalence matters, indexing term equivalence in thesaurus construction is reflected here from the point of view of translation science. Koller's division (1989) of equivalence types are reflected to the aims represented in thesaurus construction guides and standards for good indexing terms.

In the definition of equivalence type the essential characteristics are extralinguistic content conveyed by a text, connotation expressed by word selection and word order, text and language norms of the certain text type, the presumed receiver of the translation, and certain formal-aesthetic characteristics of the source text. The adequate equivalence types are denotative, connotative, text-normative, pragmatic and formal equivalence. (Koller 1989, 100-101)

In multilingual thesauri the aim is to express the extralinguistic message, such as the content of a searched database. Furthermore, in terminology work one operates usually at a word and syntagma[35] level, which is also the case in thesaurus work. It is still taken as an axiom, that aiming at word-for-word equivalence (i.e. one Finnish indexing term for one English indexing term and vice versa) is unrealistic, and that a thesaurus represents much more than its words may apparently suggest.

In multilingual thesauri connotative equivalence is also an aim. For example, in information retrieval vocabulary "client" could have as an equivalent stylistically similar equivalent in Finnish, "*asiakas*", rather than perhaps a more expressive equivalent in the

---

[35] Syntagma consists of the group and certain combination of symbols, selected for the whole. For example, a painter can produce a still-life by painting a drinking class, an apple and a skull, and these three symbols together create its meaning. If one symbol is removed, e.g.the skull, the meaning of the picture changes. In written and spoken language it is the sentence that forms the syntagma. If a word is removed or word order is changed, the whole meaning of the sentence may change. (Fiske 1992, Seppänen 2001a.)

terms of transparency, "*tiedonhaun teettäjä*". As in the case of equivalence in monolingual thesaurus work, in multilingual thesauri equivalent terms are (cf. Townley & Gee 1980, 25) "not necessarily terms literally of equal value but in a given case they may, in the normal habit of the particular thesaurus users, be treated as if they meant the same thing".

Idealistically, the terms in a thesaurus should conform to the language usage norms of its domain. In other words, the terms should be in accordance with practice and not just represent documentary language. Finnish terms should be "good" Finnish, English terms "good" English. They should conform to phonetics and morphology of their language, the language planning rules, and terminological principles. In other words, a thesaurus constructor also aims at text-normative equivalence. This view is also represented in the ISO-5964 standard.

A thesaurus represents informative text. It does not aim at the analogy of form (such as rhythm) between Finnish and English terms. When selecting a preferred term the focus is not on whether it represents the conventions of its origin in this sense. Formal equivalence in these matters is random. On the other hand, in ISO standard 5964-1985 equivalence is divided into different degrees, and in the division a meaningful distinction is made if the source and target term consist of different amount of terms.

It is very difficult, if not even impossible, to exhaustively prioritise all these aspects. Situations and expectations for thesaurus translations vary. The following is thus only indicative. The first level consists of two types of equivalence. In multilingual thesaurus equivalence is aimed for at a factual, subject matter level (denotation) of its targeted users (pragmatics). Denotative and pragmatic aims are reciprocal.

Connotative and text-normative equivalence are also seen as similar in their importance. The preferred term should conform to text-normative demands, and in the context of thesaurus construction it means that the connotative aspects should also be similarly neutral in the target and source language. If there is a paradoxical situation, the notes and examples given in the ISO-5964 standard suggest a preferred denotative equivalence.

In this study, Koller's division of equivalence types is used as a tool for analysis. With the help of this tool it has been specified what kind of equivalence is generally aimed at in thesaurus translation and it will also be used as an analysis tool in the empirical part.

When compared to the main translation strategies previously discussed – domestication, foreignisation and internationalisation - the ideology behind ISO 5964-1985 standards is closest to the idea of internationalisation. In practice this means that the content of the thesauri should be designed to make it possible to aim at equivalence between different language versions. The standards do not recommend designing versions differentiating in their content, except with regard to the amount of scope notes needed, and the non-preferred terms. As shown earlier with the term *nationalism*, equivalence at the term level would lead to a general content.

4 Linguistic and social context

The research aims to study translatability in a particular context and from a certain perspective – a current multilingual thesaurus construction, focusing on family role terminology and translatability from British-English into Finnish. The aim of this chapter is to illustrate some differences related to family roles from a social perspective and to some of the characteristics of the Finnish language when compared to the English language, as well as European foreign language usage. It is noteworthy, that the study is not multicultural, and the perspective is clearly Finnish and consequently the Finnish practices are more emphasised.

After the introduction the language policy attitudes within the European Union (EU) context is first briefly discussed, and secondly, Finnish and English as languages are illustrated. After that a part of the sociological background of the study is briefly presented focusing on gendered labour division and paternity leave models in Finland and in the UK.

As stated previously, in order to have, a heterogeneous sample from the viewpoint of equivalence problems and translatability, and a homogeneous sample from the viewpoint of the sociological context, the selection criterion in the study is a theme. Within this group are presumed to be found several different translation problems, which occur on different levels – on the concept, term and/or on the indexing term level. The theme selected is *family roles*. According to the ideas discussed earlier in chapter *3.1 On culture, language and meaning* we can state that 'family roles' are socially constructed and thus not only bound to the language but also to the surrounding culture.

Five **terms** were selected for further examination and to be used as a starting point when examining the Finnish and British indexing practices and several thesauri: family roles, breadwinners, heads of household, homemakers and housewives. 'Family roles' is the principal term and therefore used in the social context as a framework. It is understood here to refer to the tasks and responsibilities within a family. Economic and caregiver roles are emphasised in accordance to the case (i.e. translation corpus), and in order to focus two examples were selected (one for each perspective). (Concerning the theoretical sampling and the case study see chapter 6.1.1 *Using a case as a tool and multiple strategies*.)

Families, family roles and gender in educational, working and family life represents not only an interesting translation problem, but is, as a theme, topical since today an increasing amount of attention is being paid to equality between men and women, which can be seen, e.g., in legislation and the media. It is also a clear research theme in Finland and there are many studies about the topic.[36]

---

[36] E.g. Työelämän sukupuolistavat käytännöt / The gendered practices of employment (Kinnunen & Korvajärvi 1996), Kamppailu suomalaisesta äitiydestä. Maternalismi, väestöpolitiikka ja naisten kertomukset / Struggle for Finnish motherhood. Maternalism, population policy and women's stories (Ritva

The material used is mainly statistics and studies of the actors of policy making, both in Finland and in the UK.

4.1 On linguistic background

This chapter provides background information regarding the linguistic context of the study. The focus is limited to some aspects considered relevant and which highlight from the perspective of the study, the general level of the translatability issues. The chapter illustrates briefly the European language policy and language use attitudes, and discusses the different natures of the studied language-pair - Finnish as the Finnish language and English as an international language.

4.1.1 Language policy attitudes and Europeans as language users

The reason for the variety of languages in the world is that different speakers aim to mark their territories with language. Language barriers protect and even create the identity of the language society. The main problem is to find a balance between the two main tasks of the language. The main tasks are communication, which places value on intelligibility, and self-expression, which places value on the speaker's own identity. (Steiner 1975, cited here Chesterman 2005, 115-116)

In the context of European Union, Johansson and Wiberg (2005) state, that tomorrows Europe can be compared with other areas in world, where English is a *common lingua franca* between numerous local languages, as is already the case in South-Africa and India. The motto of the European Union is "United in diversity" (in Swedish "*Förenade i mångfalden*", in Finnish "*Moninaisuudessaan yhtenäinen*"). Johansson and Wiberg ask whether sometime in the (near) future the struggle between the different languages, the monolingual trend and the integration development, will abbreviate the motto to only one single word - united? (Johansson & Wiberg 2005, 45)

Johansson and Wiberg (2005) have summarised arguments concerning conflicts and problem situations, where the institutional multilingualism of European Union arises. Negative arguments about multilingualism suggest that multilingualism is an obstacle for efficient communication and for practical attitudes towards languages.

Nätkin 1997), Äidit, isät ja ammattilaiset. Sukupuoli, toisto ja muunnelmat asiantuntijoiden kirjoituksissa / Mothers, fathers and professionals. Gender, repetition and modifications in the writings of professionals (Vuori 2001), Isänä olemisen uudet suunnat / The new directions of being a father (Huttunen 2001), Family Policy, Work Incentives and Employment of Mothers (Forssen & Hakovirta 2001), Työn ja perheen väliset ristiriidat ja niiden taustatekijät naisilla ja miehillä / The conflicts between work and family and their background factors for women and men (Lahelma, Winter, Martikainen, Rahkonen 2005), Perheen vastaisku / The backlash of the family (Jallinoja 2006), and Diverging paths? The dual earner - dual carer model in Finland and Sweden in the 1990s (Haataja & Nyberg 2006).

Multilingualism slows down business discussions, delays decision-making and the enforcement of the decisions made. Multilingualism is an obstacle for European integration. The reasons and justifications for monolingualism are that everybody knows at least some English, which could thus work as a lingua franca. Discussions about language switch situations are considered a political problem which should be avoided. Multilingual communication is also considered problematic and limited. Monolingualism would deepen European integration and would increase the mobility of labour. It would be necessary for European federalisation, because a uniform European policy is considered to need one common language. (Ibid, 41)

Wiberg and Johansson remind us that the ideal of unilingualism is an old belief (see also Eco 1995). In monolingual communication the cultural and symbolic dimensions are ignored and communication is then considered as just conveying the message. Using only English gives excessive advantage for those, who speak English as their mother tongue. (Ibid, 41-42, see also Moring 2005)

The reasons for retaining multililingualism are that the costs are rather small when compared to the whole budget of EU. Multilingualism must be appreciated, because each citizen has a right to use his or her own language, and this quarantees democracy. Multilingualism is a European richness. (Ibid, 42)[37]

In the study Finnish and British-English speakers are considered as Europeans, and the multilingual thesaurus context is also Europe.

Crystal (1997, 61) estimates that a total of up to 1 800 million people, that is 30 per cent of the world population, are already "reasonably" competent in English and of these, 670 million have a native or native-like command of English. According to a 2001 Eurobarometer survey English is the mother tongue of 16 per cent of the European Union population. The foreign language most frequently known by Europeans is English (41 %), followed by French (19 %), German (10 %) and Spanish (7%). English is also the first foreign language most used by Europeans (average 33 %, in Finland 47 %). The most useful language in addition to the mother tongue was considered to be English (by 75 % of respondents). 74 % of Europeans do not know a second foreign language. 47 % of Europeans claim to know only their mother tongue. In the United Kingdom, French is the most used foreign language (17 %). (INRA 2001)

Europeans were also asked about their attitudes to foreign languages. The results showed, that 42 % Europeans believe that knowledge of foreign languages is or would be useful to them personally, while 38 % believe it would be fairly useful, 10 % not very useful, 12 % not at all useful, and 6 % do not know. Europeans were also asked which two languages they *thought were most useful to know* apart from their mother language. English (75 %) was considered as the most useful language to know. English was followed by French (40 %), German (23 %) and Spanish (18 %). (Ibid.)

---

37 Although Europe is considered a multilingual, only 3.5 percent of living languages are spoken in Europe. - There are an estimated 6,900 living languages in the world, out of which 239 are spoken in Europe. (Lewis 2009) As a continent Europe is linguistically one of the most unified (Moring 2005, 76).

An analysis made by socio-demographic variables showed that the tendency to know other languages in addition to the mother tongue diminishes with age. 66 % of the 15-24 year-olds claim to speak English, compared with 53 % of the 25-39 year-olds, 38 % of the 40-54 year-olds and 18 % of the over-55s. The same pattern applies regardless of the language. Furthermore, the higher the level of education, the greater the likelihood of knowing a foreign language. Knowledge of another language is proportionally highest among students (78%), managers (67 %) and white-collar workers (59 %), and lowest among pensioners (17 %) and house persons (27 %). In addition, the younger age-groups are more likely to regard knowledge of foreign languages as very useful or fairly useful. Presented in numbers, 87% of 15-24 year-olds share this view, compared with 78% of 25-39 year-olds and 76% 40-54 year-olds. The over-55s are much less likely to share this view (57%, compared with an EU-15 average of 72%.) (Ibid.)

In content management these issues are worth serious consideration. If almost half of the European citizens know only their mother tongue, how can we expect them to seek information more globally, without proper tools such as multilingual thesauri? How can we expect them to find documents published in their mother tongue but indexed in another culture and another language than their own? – In Finland, for example, a considerable amount of information is published in English, which, without multilingual information retrieval tools may not be accessed by non-English speakers.

4.1.2 On Finnish and English

Before going into more detail concerning family roles and family policy as the sociological context of this study **some general information** about Finnish and British-English is given. The language use in Finland and in the United Kingdom is also discussed in the context of Europe.

On the World Wide Web, English is a dominant language and Finnish a very minor language (see Grefenstette & Nioche 2000, Chesterman 2005). A common and growing trend in Finland's academic life is to write and publish in English. It has been estimated that in Finland in 2001 the ratio between Finnish and English in scientific publication was 70:30 English being thus clearly superior (Wilson 2002, cited here Chesterman 2005, 118). There has been a wide and lively debate on whether to use Finnish or English. On the one hand it is important to have a larger audience, but on the other hand, it is essential to publish research results in our native language and thus also ensure that the special terminologies will also be up-to-date in Finnish (see e.g. Leiwo 2000; Oittinen & Väyrynen 2001). Documentary languages are facing the same problem and therefore there is a growing need to develop tools for multilingual indexing and cross-lingual information seeking (cf. chapters 8 and 9).

The nature of Finnish and English is very different when considering multilingual communication. Finnish belongs to Finno-Uralic languages, whereas English is international and belongs to the Germanic languages, and the differentiation has also

influenced the research design. The fact that English is the most common language on the Internet (see e.g. Grefenstette & Nioche, 2000; Chesterman 2005) and also the common working language between international social science documentation providers, such as the *International Society for Knowledge Organization* (ISKO) and the *members of the Council of European Social Science Data Archives* (CESSDA) adds to its importance, also in the international environment of social sciences and information seeking, even in Finland.

Finland's population is about five million. In Finland there are two official languages: Finnish and Swedish. Their status is also defined in Finland's legislation. Finnish is spoken by approximately 92 percent of the population as a first native language. About 300 000 are Swedish speakers (mainly on the southwestern and southern coasts and on the Åland Islands in the Baltic Sea), constituting almost six percent of the population. Of Finland's 2,500 Lapps about 2000 speak the Lappish language as their first native language. (Katzner 2002, 347, Karlsson 2000, 73, Hiidenmaa 2000, 56) However, Finland is not a homogeneous country if we consider the number of spoken languages: in Finland 150 languages are spoken (Hiidenmaa 2000, 56).[38] As a language Finnish received its official status in 1863 (Allardt & Starck 1981, 185).

In free-text searching the richness of noun case inflections in Finnish is often considered a problem: Finnish has fifteen cases. In Finnish the inflection of words is done by adding grammatical affixes instead of using prepositions as in English and other Germanic languages. An illustrative (and typical) example within the field of information studies is English *search* versus Finnish *haku*:

> - *in searches – hauissa* (means also *in pikes*, the basic, nominative form being *hauki/pike*)

> - *searches – haut* (nominative) / *hakuja* (partitive). (Järvelin 1995)

In controlled vocabularies, the noun inflections are not a problem, but another problem (also influencing free-text search) are compounds. In Finnish, it is often not clear whether to write compound words together or not. (See also Pirkola & al. 2001; Airio 2009)

The other language discussed is **British-English**. The population of the United Kingdom is approximately twelve times greater than Finland's: 60 million. More importantly, English is spoken universally. About 600 000 people in Wales speak Welsh, 100 000 in Northern Ireland speak Irish, and about 75 000 in Scotland speak Scottish Gaelic, but all these people also speak English. (Katzner 2002, 366)

English is stated to be the most used language in world wide communication, although it is the mother tongue of "only" 350 million people.[39] English is the most

---

38 Satistics do not give a perfect image of the reality and do not indicate the exact truth about the language skills of a population. In language registration there is a principle, according to which one can only have one language as mother tongue, so bilingual people have to choose and mark only one specific language as their mother tongue. (Johansson & Pyykkö 2005)

[39] The most common mother tongue, used by over a milliard people, is Chinese. (Häkkinen 1994, 38)

common official language (50 countries), followed by French (28 countries), Arabic (23) and Spanish (21). (Karlsson 2000, 73)

Why is English so superior? An English-speaking country has been in a leading position in three central historical phases. First, in the 18th and 19th centuries England, due to its navy, was the leading country of colonialism. English was then taken together with imperialism to North-America, Australia, New Zealand, Africa and areas of the Pacific Ocean. Second, the industrial revolution also began in England and started to grow especially strongly in the United States in the 19th century. A significant part of the new technology spread out first via the English language. The third reason for the domination of the English language has been the economic and military superiority of the United States of America in the 20th century and in the beginning of the 21st century. (Chesterman 2005, 116-117, see also Johansson & Wiberg 2005, 131)

In multilingual communities a *lingua franca* is often needed. According to Karlsson (2000) there are many kinds of lingua francas. In extremes cases it is so called pidgin language. A more pure form of lingua franca is that one individual language develops to become the dominating language because of economical and practical reasons. (Karlsson 2000, 75-76)

From a global perspective *English* has become a lingua franca. A lingua franca is needed especially in business life, in the political-administrative sector and in the international academic community. The increase in turism has also created pressures. (Karlsson 2000, 75-76, Chesterman 2005, 115-120) English is the most used (85%) language of all international organisations (Crystal 2003, 87).

4.2 The social background

Finland is considered to be one of the most egalitarian of countries. Finnish women were the first in Europe to be granted the right to vote and the first in the world to obtain the right to become candidates at an election, which occurred in 1906. (The Council for Equality 1999, 2) However, according to the United Nation's Human Development Report 98 comparing the position and quality of life of women, Finland was not the pioneer:

> "Finland is the fifth most equal country in the world. Canada ranks first, followed by Norway, Sweden and Iceland. In a comparison of the quality of life, Finland ranks sixth." (Council for Equality 1999, 15)

The cornerstones of gender equality as well as financial independence are parental leave, home care leave and safe day care for children (Council for Equality 1999, 8). The Finnish Equal Opportunity Barometer Study revealed that the clear majority of both men and women feel that the social position of women is inferior to that of men, and

---

When comparing the numbers of mother tongue language users - the most used language in Europe is German, followed by French, Spanish and Italian. Spanish and Polish could be categorised as middle-size languages. (Johansson & Wiberg 2005, 33)

that one out of two men and women believed that equality will increase over the next ten years. Finns stated that women are entitled to work irrespective of their family circumstances and that men ought to participate more in child care and rearing. (Ibid.) The two clear aspects in gender equality as well as in family roles (see also in thesaurus context in appendix 8) are economic and care-giving roles (men and women in parenthood and in occupational life) which are discussed in the next section.

4.2.1 Participation in labour life and paternity leave modes

In Finland, women have been gainfully employed for over a century now. Today, there are almost as many women as there are men in the labour force[40]. Both genders are engaged in full-time work. In addition, the mothers of small children work outside the home. (Council for Equality 1999, 2, 5).

According to the Council for Equality (1999), at the turn of the new millennium the Finnish fathers were taking responsibility for the care of their children and the home. Other influencing factors mentioned that aid women to be able to go out to work were e.g. that children get a warm meal at school and the majority of parents can also have a meal at their workplace, as well as well run public transport and geriatric care. Since 1973, the municipalities have been responsible for arranging children's day care. The entitlement of children to a municipal day-care place has since become a subjective right, and in 1996, this right has applied to all children under the age of seven (until the age of compulsory education) regardless of whether the child's parents are at work or not. When children are under school age, the parents also have the option of choosing shorter working hours. If the children are under the age of three, employees on such leave are also entitled to partial care allowance. The payments of public day-care system are scaled according to the parents' income. (Ibid. 1999, 8-9) Finland is thus in principle often considered as rather gender-equal, but in practice there are differences in e.g. labour force participation.

In the table below the labour force participation rate in Finland during 1970 – 2001 is illustrated (population aged 15 – 74):

---

[40] The labour force consists of employed and unemployed.
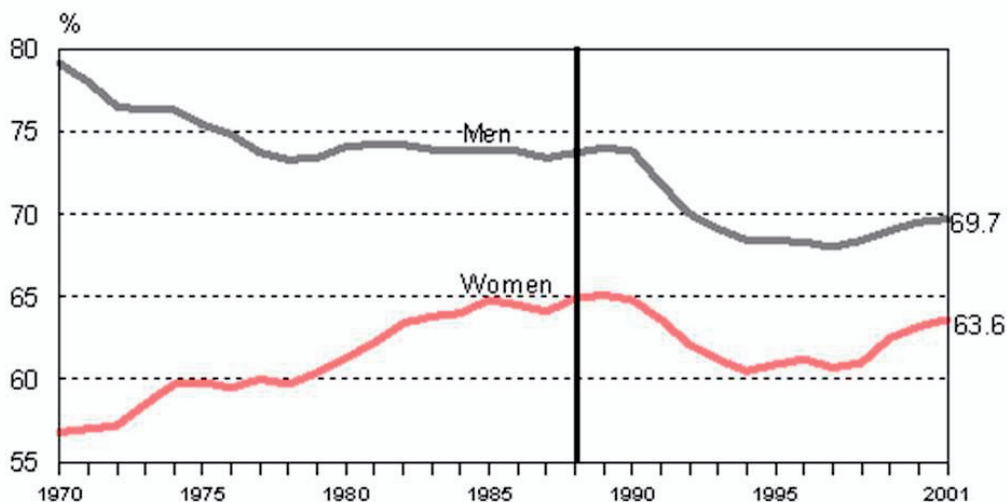
Table 1: *Labour force aged 15 – 74, 1970 – 2001, Finland. (Source: STAT 2002b)*

When comparing only the numbers of people in the labour force in Finland we can state that women are rather equal to men. As will be discussed later, the equality becomes more uneven, when other facts are taken into consideration such as differences in wages and participation in part-time employment.

In 1996, among couples of working age, in approximately 25 % of the families the spouses' personal net income was about the same. In about 60 % of the couples the man had a larger income. (Council for Equality 1999, 5.) In only 15 per cent of the families was the working women's income more than their spouses, which is therefore rather rare when compared with the men's sixty per cent. The salary gap between women and men with the same educational background working within the same sector in the mid-1990s was approximately fifteen per cent. The differences in women's and men's pay in the private sector was greater than those prevailing in the public sector and women directors earn on average 35 per cent less than their male counterparts. (Ibid 6-7.)

Although statistics show that women are not equal with men in occupational life when considering e.g. salaries, there is also inequality the other way around. In society nowadays there is a lively debate about family roles from the men's perspective, which can also be seen in the titles of the documents – such as *Flexible Work and Flexible Fathers[41], Caring as Social Right: Cash for Child Care and Daddy Leave[42]* How should we index these documents in conformity to controlled vocabularies and current

41 Brandth, Berit & Elin, Kavnde 2001: Flexible Work and Flexible Fathers. Work, Employment & Society, Vol. 15 (2): 251-267.

42 Leira, Arnlaug 1998: Caring as Social Right: Cash for Child Care and Daddy Leave. Social Politics, Vol. 5 (3): 362-378.

practices? How should we express an approach which is different to the main stream and controlled vocabularies?

The terminological case *family roles* also cover economical roles within a family, and in the studied ELSST sample they are stressed. According to the statistics (STAT 2001a[43]) there are some differences in Finland versus in the United Kingdom (UK) regarding gender division of labour force at the turn of the millennium.

First, the number of people in the labour force was rather even with regard to Finland (50.4 %) and the UK (49.2 %) at the beginning of the new millennium. A slightly clearer difference was seen regarding the female labour force (see table below):

	1 000 persons	% of labour force	employed persons (1 000 persons)	% of population
Finland	1,238	44.5	2,356	44.6
UK	13,074	47.4	27,793	46.4

Table 2: *Female labour force - Finland versus UK 2000-2001 (see STAT 2001a)*

The greatest difference concerns participation in part-time employment, which was clearly more common in the UK than in Finland and was also more female-gendered:

	part-time employment %	–of whom females %
Finland	10.4	63.8
UK	23.0	79.9

Table 3: *Female part-time labour force - Finland versus UK 2000-2001 (see STAT 2001a)*

The numbers and estimations as well statistical classifications vary, but there is a clear trend in moving towards two full-time employed parents also in the UK (see e.g. O´Brien & Shemilt 2003).

In 1995 in Finland 79 per cent of the fathers and 52 per cent of mothers were full-time employed. Seven per cent of fathers and 25 per cent of mothers were economically inactive. In the United Kingdom, 84 per cent of the fathers and 20 per cent of the mothers were full time employed. Six per cent of the fathers and 36 per cent of the mothers were economically inactive. (Deven et al. 1998, cited here O´Brien & Shemilt 2003, 9)

Around 2000, that is a large majority of part-time working mothers in the UK gave as the reason to prefer part-time work over full-time work, the need to meet domestic commitments. In the UK, the common mode for families with dependant children is 'long hours for men' and 'short hours' for women, but 40 per cent of the families with dependant children contained two full-time working adults. In these families especially the fathers worked long hours and overtime. (Cousins & Tang 2003, 9-14)

---

[43] Stat 2001a summarises the results of and refers to various sources of United Nations (UN) as Economic Commission of Europe (ECE), International Labour Organization (ILO), and Organisation for Economic Cooperation and Development (OECD) and World Bank.

There exist great differences among different practices on paternity and parental leave in Europe, in which the UK and Finland are not totally opposites, but still have clear differences. The paternity and paternal leave system in 2002/2003 was different in Finland versus in the UK:

Finland	Paternity leave: 1 week, to below 70 % of earnings. Maternity leave: 9.5-12.5 weeks. Parental leave available as a family entitlement until child is 26 weeks (with an additional Care Leave up to 36 months). Paid part flat-rate and part earnings-related.
UK	Paternity leave: 2 weeks, universal flat rate. Maternity leave: up to 52 weeks, part earnings-related, part flat-rate and part unpaid. Parental leave available as an individual entitlement until child is 6 months (no payment).

Table 4: *The paternity and parental leave system in 2002 Finland / 2003 UK (O´Brien & Shemilt 2003, 37-38)*

A recent trend in Nordic countries has been to mark a proportion of paid parental leave to be devoted exclusively to fathers. It was first implemented in Norway in 1993, Sweden in 1995 and in Denmark in 1999 (O´Brien & Shemilt 2003, 38). In 2009, fathers in Finland can take between 1 and 18 days of paternity leave after the birth of their child, in order to look after their child at home together with the mother. They can take paternity leave at any time after the birth of their child while Maternity or Parental Allowance is being paid. The paternity leave can be taken in up to 4 separate periods. (Kela 2009a) In addition, fathers are today entitled to additional leave days called "daddy-month". (Kela 2009b) The two week's paternal leave was introduced in the UK as late as 2003. The system is in flux, and in the UK is today a gender-equal plan is being discussed with the possibility that both parents could take six months off work during the first year of their baby's life (see e.g. BIS 2009)

In Finland, about one half (46 per cent) of all children under school age make use of municipal day care services. For the children aged three and more it is most common for them to be in day care centres. (Ministry of Social Affairs and Health 2002, 11)

- **Summary and Conclusions**

Finnish as a language belongs to a different language group than English. More importantly, Finnish is culturally very Finnish, while English is globally used as a common lingua franca, and commonly used also in the context of Finnish academic discourse.

In Finland women are more engaged in full-time employment than in the United Kingdom (UK). In Finland the parental leave system enables fathers to concentrate more on their new-born child than in the UK. Using these as examples it is obvious that the gendered male breadwinner model and female homemaker model are stronger in the UK than in Finland, but it is noteworthy, that as a phenomenon they are not strange for Finnish practices.

5 The study

As already stated in the introduction (chapter 1) the focus of this study is twofold:

On the empirical level it focuses on the translatability of British-English social science indexing terms into Finnish language and culture on a concept, term and indexing term level, and; on theoretical level its main interest is on the focus of translation and on the concept of equivalence.

The interest is thus on the human dimension and accordingly to the ideas represented within functional, communicative translation theories (see chapter 3.4 *Translatability and equivalence*) translating is seen as a decision-making process. The study represents a unique field of research and therefore the research questions are not only empirical, but also methodological and theoretical.

In chapter 5.1, the research questions are first represented at a general (as in chapter 1 Introduction) and then at a more detailed level and after that, in chapter 5.2, the research questions, methods and material used are aggregated and the working order illustrated. (The methods and material used will be discussed more detailed in the next chapter 6 *Methods and material*.)

5.1 Research questions

The overall goal of the study is to explore translation equivalence in multilingual and multicultural thesaurus construction. The general research questions for the study are:

1. Can differences be found between Finnish and British discourses regarding family roles as thesaurus terms, and if so, what and which kind of implications do these have for multilingual thesaurus construction?

2. What is the pragmatic indexing term equivalence?

These overall questions will be explored through a term which may be interpreted differently depening on the cultural context, namely "family roles".

The first question studies how the same topic (*family roles*) is represented in different contexts and by different users, and furthermore, focuses on how the possible differences are handled in multicultural thesaurus construction. The second question is based on the findings of the other two, and the answers to the final question of what kind of factors should be considered when defining translation equivalence in multicultural thesaurus construction.

The practical questions for the study are

- How is equivalence understood in thesaurus construction guidelines and standards as opposed to modern translation studies versus in practice?

- How is the studied case, *family roles*, understood in different types of material and discourses?

- What is the goal of thesaurus term translations?

The more methodological questions for the research are

- How to study equivalence in a context of multilingual and multicultural thesaurus construction?

- What kind of theoretical framework would be suitable for a study which considers translation problems from an information science perspective?

- How to seek equivalents?

- How to evaluate them?

- How to operationalise the results?

Also a background question for the empirical case, *family roles*, has been asked (see chapter 4):

- What is the socio-cultural context of the studied phenomenon (family roles)? - How do Finnish and British societies differ?

To illustrate this, mainly Finnish and British statistical data and official reports have been used focusing on two aspects – the gendered division of labour and the paternal leave model.

As illustrated in chapters 5 and 6, the studied material is heterogeneous and multiple methods are used. The main material consists of focused interviews. The themes and questions given to the interviewees are partly the same and partly different, depending on their role in the study as a social scientist, indexer or as a thesaurus constructor (see also material in chapter 6 and interview guides in appendices 1-6) and whether they are Finnish or British participants. The main interview themes are:

 educational background and working history

- simulated indexing task

- tools and guidelines commonly followed in indexing and in thesaurus

construction

- word associations for the studied terms

- definitions for the studied terms

- translations for the studied terms

- different discourses

- relevance in multilingual and multicultural thesaurus context

- equivalence in multilingual and multicultural thesaurus context

5.2 Research design

The research questions, the methods used and the material are summarised in the table below:

Research Questions	Data Collection and Analysis Methods	Material
What is the socio-cultural context of the studied phenomenon (family roles)? - How do Finnish and British societies differ? (Examples)	Background analysis	Mainly Finnish and British statistical data and official reports, Eurobarometers etc
How is the studied theme (family roles) represented in commonly used thesauri?	Context analysis Discourse analysis	6 mono- and 3 multilingual thesauri
How are the studied terms understood and defined/translated in different cultures and subcultures?	Focused interview Discourse analysis	Finnish and British indexers, thesaurus constructors and specialists as interviewees, in all 27 interviews (29 informants), 7 online dictionaries
How do the given equivalents differ from each other?	Component analysis	Equivalents and definitions given to the studied terms
How are the studied terms used in Finnish and British indexing?	Co-word analysis ------------------ Simulated indexing situation/Focused interview	Datasets retrieved from Finnish and British databases (tot.3), as a starting point 5 keywords, i.e. ~15 cases ------------------------ 3 Finnish and 3 British indexers using 5 documents
Similarities and differences found in the Finnish and British discourses?	Classification Discourse analysis	Results of the interviews, component analysis and co-word analysis
What is pragmatic equivalence in multicultural thesaurus construction?	Conclusions	Theoretical discussion + empirical studies

Table 5: *Summary of the research questions, methods and material*

6. Methods and material

In this chapter, the nature of the research, methods and material used in this study and the reasons and motivations to use them are illustrated. First, the nature of the research is discussed, then the methods and material are presented. (The research design has been summarised in chapter 5 *The Study*.)

Multilingual thesauri are little studied, and as there is no research tradition this study, therefore, uses several types of material and also mixes quantitative and qualitative methods. Concerning thesaurus construction Lykke Nielsen (2004) states:

> "--- The thesaurus should be developed according to the nature and needs of the information environment in which it is going to be used. This implies a thorough knowledge of the information environment and the discourses and languages in play. This knowledge may be gathered by different methodologies. Recently, there has been some discussion about what approach to take in order to get the needed insight. However, it is generally accepted to take a holistic approach and study the information environment and its users as an interacting whole." (Ibid. 74)

6.1. Quantitative and qualitative methods

The study focuses on equivalence and translation strategies in multilingual thesaurus construction. The emphasis is on human-effort and translation is seen as a decision-making process (see chapter 3.4 *Translatability and equivalence*). This study uses **quantitative and qualitative methods**. Two different methodological approaches are needed in order to answer the research questions of this study. Quantitative and qualitative methods are thus used to supplement each other and to give answers partly to the same and partly to different questions. A quantitative method is used for data collection purposes, but due to the terminological nature and perspective of the study and the small data sets, the interpretation of the quantitative material is in its nature qualitative. In its methodological nature the whole study is a **qualitative study**.

The starting point in a qualitative study is that the researcher performing it openly admits their subjectivity and that the researcher is the central research tool. (Eskola & Suoranta 1998, 209, 211, 216). The role of the researcher is emphasised, when the data collection method is through interviews.

Qualitativeness is also emphasised in developing the analysis method for this specific research material. Traditional terminology sources have been supplemented by the theories and concepts of translation studies, linguistics and communication theories, in order to achieve a broader, deeper and more pragmatic perspective (see especially chapter 3.4 *Translatability and equivalence*).

During the 1990's, there has been a lively debate about quantitative versus qualitative approaches and it has become more acceptable and reasonable for researcher

to use the both approaches within a specific research work. The methods are not seen as opposed to each other anymore, but rather complementary and giving answers partly to the same questions. (Hirsjärvi & Hurme 2000, 21-33. See also Töttö 2000, 68-101.) Brannen (1992) states, that

> "--- there are grounds for arguing that both qualitative and quantitative approaches need to be applied in combination, especially where investigations are carried out on social groups whose material situations and perspectives have been under- or mis-represented in social research. While the qualitative approach may overcome some of the problems of giving a voice and language to such groups, through which they may better express their experiences, the quantitative approach would serve to indicate the extent and patterns of their inequality at particular historical junctures." (Ibid, 22)

According to Brannen (1992, 24-25) qualitative methods may play a subsidiary role in a project for several reasons:

> "First, they may act as a source of hunches or hypotheses which quantitative work may then go on to test. A second usage is in the development and piloting of research instruments – questioners, scales and indices. A third and often covert usage is in the interpretation and clarification of quantitative data." (Ibid.)

Accordingly to the third function in this research, quantitative methods are used to illustrate how certain terms are used in Finnish and in British indexing and qualitative methods are used to examine the discourses and the nature of the problems occurring in the research material. The use of certain approaches and methods depends both on the problems examined and on the stage of the research process. Since the problems are partly overlapping and complex ones, several methods are also used simultaneously.

This kind of research can be called a multistrategy approach. According to Layder (1993) the multistrategy approach has two principal aims:

> "First, it is 'open' to as many strategies and analytical cuts of the data as possible, in order to produce robust and firmly grounded theory. Secondly, this 'openness' of the approach is tempered by an overarching vision of the relations between macro and micro elements." (Ibid, 109)

As Layder (ibid.) notes, these aims may appear on the surface to be contradictory. He further clarifies, that

> "In this sense, the most appropriate and productive research strategies are those which incorporate this vision. Therefore, the multistrategy approach should not be misinterpreted as anarchic or eclectic. Rather, it should be understood as a form of 'disciplined flexibility'". (Ibid, 109)

Layder (1993) illustrates the quantitative and qualitative analysis in multistrategy research as follows:

		Qualitative	Quantitative
HISTORY	CONTEXT	Theoretical/interpretive characterizations	Demographic characteristics Traditional quantitative data
	SETTING		
	SITUATED ACTIVITY	Importance of qualitative data from observations and interviews	Simple forms of counting
	SELF		

Figure 8: *"Qualitative and quantitative analysis in multistrategy research" (Source: Layder 1993, 114, figure 6.2)*

There is a shift in emphasis in the kinds of analysis and data that are pertinent to different layers of social organisation. Self and situated behaviour is best studied by the use of classical forms of qualitative research, such as observation and/or semi-structured interviewing, "because these topics demand some ethnographic, interpretive account of the meaningful world of the individuals involved". (Ibid, 114-115)

The analysis of the study occurs in stages. According to Alasuutari (1994) qualitative analysis consists of two phases, namely reducing the observations and solving the mystery. The two parts can be separated in order to simplify the observations. First, the material is always studied from a certain theoretic-methodological perspective. Therefore, when analysing the material the focus is only on those matters, which are relevant from the point of view of a certain theoretical framework and of the viewpoint of the research questions, although the material could be studied from several perspectives also within a specific piece of research work. The simplification on the second phase happens by combining the observations. Unique observations are combined into one observation or at least to a more limited group of observations. The starting point is the idea that in the material examples or evidence exists about the same phenomenon. In a qualitative research the second part, solving the mystery, means that the interpretation of the studied phenomenon is made on the basis of the leads and clues produced. (Ibid, 30-31, 35)

In the next subchapter, the multiple strategy approach of this study is discussed in more detailed.

6.1.1 Using a case as a tool and multiple strategy

A case is selected as a practical approach and a concrete tool for the research problems. It is done in order to gain insight into the different discourses and practices, because only they are considered to be able to provide proper answers to the research questions of the study.

The study uses multiple cases aiming at theoretical completion and replication. The study is thus a qualitative study and in its nature descriptive, generating hypotheses and illustrating new concepts such as a "pragmatic indexing (term) equivalence". (Pragmatics is the study of language use in context. Here it refers to the indexing and information seeking environment, which can be considered as a unique discourse type.)

The study is formed of several samplings. According to Seale (1998), sampling as a method means the selection of units of analysis (for example, people or institutions) for study. It can involve attempts to statistically represent a population, in which case a variety of random methods are available, and alternatively, sampling can be opportunistic, or formed by emerging theoretical concerns of the researcher. (Ibid. 329)

The samples studied are theoretical (not random) and thus cases assumed to give answers to the questions of this study. Strauss and Corbin (1990, 176-179) consider theoretical sampling as sampling on the basis of concepts that have proven theoretical relevance to the evolving theory.

> ---"The term **proven theoretical relevance** indicates that certain concepts are deemed significant because (1) they are **repeatedly present** or **notably absent** when comparing incident after incident, and (2) through the coding procedures they earn the status of categories. --- The aim of theoretical sampling is to sample events, incidents, and so forth, that are indicative of categories, their properties, and dimensions, so that you can develop and conceptually relate them." (Ibid, 177)

In its nature, theoretical sampling is cumulative and, it increases the depth of focus, ensures the noting of variation, process, and, density, and it is planned, and also flexible. Theoretical sampling is primarily used with qualitative data and in Grounded Theory it refers to sampling until the researcher reaches a point of data saturation. It is similar to the triangulation method. (Ibid.)

Denzin (1978) separates four ways to triangulate in research.

> "1. *Data triangulation*: the use of variety of data sources in a study;
>
> 2. *Investigator triangulation*: the use of several different researchers or evaluators;
>
> 3. *Theory triangulation*: the use of multiple perspectives to interpret a single set of data;
>
> 4. *Method triangulation*: the use of multiple methods to study a single research problem." (cited here Janesick 2000, 391)

In this study the triangulation of data, theory and methods is used, but the different types are mainly used to supplement each other – no single type alone could provide the proper answers, but together they can lead to satisfying results.

- **Summary and Conclusions**

Several data collection and analysis methods will be used in order to answer the research questions. The use of multiple methods and data thus aims to complement different approaches and the triangulation of evidence. The research process starts from the hypotheses (statements) and ends through observations and generalisations of modelling translation strategies and equivalence in different types of multilingual thesauri.

The data collection methods are both quantitative and qualitative, whereas data-analysis methods are by nature mainly qualitative. The methods are used iteratively and simultaneously. This kind of method aims at increasing the validity and "density of empirical coverage" (see Layder 1993).

6.2 Methods

The study examines Finnish equivalents for British-English indexing terms. The key terms of the research corpus will be studied:

- by **interviewing** people representing different cultures (Finnish versus British) and subcultures (indexers, specialists, multilingual thesaurus constructors);

- by **co-word analysis** (the use and **indexing practice** of the key terms studied in Finnish, British and international databases);

- by comparing several **thesauri** and;

- by **component analysis** (clarification the meaning of the studied terms and recognition of the equivalence type).

The foremost method used is **discourse analysis**. In the following, the methods used in this study are illustrated in more detail, that is, the reasons and motivations for using them in this study.

6.2.1 Discourse Analysis

Discourse analysis is both context-bound and a human activity, showing which aspects are needed in development of information products such as thesauri. Olsson (2004, 2) states, that

> "- - - It is only through an understanding of people not simply as isolated intellects or rational decision-makers, but as social beings intimately connected to their life-world,

that we can hope to develop truly effective, guinely user-centred, information products and services." (Ibid.)

Fairclough has presented his approach to discourse analysis and its use in investigating social changes, in his work *Discourse and Social Change* (1992). He pointed out some conditions that needed to be fulfilled in a useful study:

"Firstly, it would need to be a method for multidimensional analysis. --- Secondly, it would need to be a method for multifunctional analysis. --- Thirdly, it would need to be method for historical analysis. --- Fourthly, it would need to be a critical method. Relationships between discursive, social and cultural change are typically not transparent for the people involved. ---" (Fairclough 1992, 8-9)

Fairclough (1992, 72) has also illustrated the concept of discourse as three-dimensional and thus attempted to bring together three analytical analyses, each of which, according to him, are indispensable for discourse analysis.

"These are the tradition of close textual and linguistic analysis within linguistics, the macrosociological tradition of analysing social practice in relation to social structures, and the interpretivist or microsociological tradition of seeing social practice as something which people actively produce and make sense of on the basis of shared commonsense procedures." (Fairclough 1992, 72)
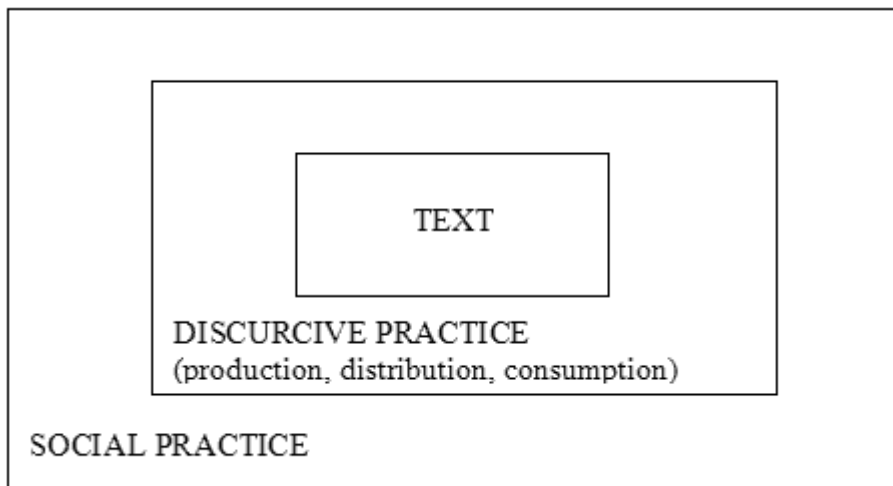
Figure 9: *"Three-dimensional conception of discourse" (Source: Fairclough 1992, 73.)*

As stated earlier (in chapter 2), at a general level discourse can be defined to refer to language use in a social context (Pälli 2003, 22). It is also noteworthy to realise, that although discourse is a substantive, it also means doing something and is an active process (see Lehtonen 1994; Potter & Wetherell 1990; Pälli 2003).

In 1980 Busha and Harter (ibid, 171) stated, that many library science studies are focused on the users of various media of communication and less frequently are inquiries concerned with the actual content of the media. The tendency is still the same. Discourse analytical methods have been scarce in the field of librarianship and information science research, although the interest has been growing (Hedemark,

Hedman & Sundin 2005, 17). The starting points of discourse analysis have been discussed in several articles, but it has seldom been used as a method for empirical studies (Talja 1998, 18). However, discourse analysis is a natural choice, when we think of the key element in our field – information. - Talja (1997, 70) describes information as being concerned with what people do with language and what language does to people.

Talja (1999, 460) says, that

"Discourse analysis studies practices of producing knowledge and meanings in concrete contexts and institutions – be they in library organizations, information studies, information society strategies, database interfaces, or the Dewey Decimal Classification. Discourse analysis systemizes different ways of talking in order to make visible the starting points on the basis of which knowledge and meanings are produced in a particular historical moment. It pays attention to the way in which discourses produce and transform social reality, and makes it possible to evaluate the practical consequences of different ways of approaching a particular phenomenon." (Ibid.)

It is not reasonable to characterise discourse analysis as a distinct research method, but more as a loose theoretical framework (Potter & Wetherell 1990; Suoninen 1992, 125) that allows different emphases in focus and methodological applications (Jokinen & al. 1993, 17-18).

The idea of dynamic equivalence and translation as an active decision-making process (see 3.4.2 *Equivalence in communicative translation theories*) corresponds well with discourse analysis. Potter (1997) states:

"DA [discourse analysis] has an analytical commitment to studying discourse as *text and talk in social practices*. That is, the focus is not on language as an abstract entity such as a lexicon and set of grammatical rules (in linguistics), a system of differences (in structuralism), a set of rules for transforming statements (in Foucauldian genealogies). Instead, it is the medium for interaction; analysis of discourse becomes, then, analysis of what people do. ---" (Ibid, 146)

Content analysis and **semiotics** can be seen as types of discourse analysis, the emphasis in this research being on semiotics. Slater (1998) describes the roots and use of the methods:

"**Content analysis** and **semiotics** represent two important attempts to introduce consistent methods to the interpretation of culture. Both are forms of textual analysis, aiming to provide convincing readings of cultural texts, and to draw various conclusions from them, by looking at texts themselves rather than at the ways in which people actually consume these texts. - - - Content analysis, as an old and rather positivist-inclined method, characteristic of mid-century American sociology, tends to fairly mechanistic readings and conclusions; semiotics, on the other hand, seeks to draw out the full complexity of textual meaning, as well as the act of reading texts, but with little rigour in a conventional sense." (Slater 1998, 234)

"At bottom, content analysis simply measures frequency, and typical research questions might be: 'how prevalent soap opera are sexist images of women?' or 'how often are women depicted in soap operas as mothers, as opposed to sex objects, workers or mainstains of the community?' --- " (Ibid, 234-235)

"A content analysis is rather like a social survey of a sample of images, rather than of people, using a tightly structured and closed questionnaire. The stages are much the same. Content analysis firstly involves suitably defining a population and drawing a sample from it. - - - The second major component of content analysis involves deciding on categories for *coding* the data and carrying out the coding. --- " (Ibid. 235-236)

The discourse analytic approach offers a possibility to study differences and similarities both on the individualistic and community level. It also enables the use of several layers. An individual can be seen, for example, as a Finnish, as a social scientist and/or as an information seeker depending on the perspective and question. Pälli (2003, 18) points out that what is important is that the discourse analysis does not limit to only the linguistic structure, but takes into consideration also the social context of the language usage.

Slater (ibid.) further points out, that much of the apparent rigour of content analysis rests on the structure of categories used. The categories must follow the common guidelines of categorisation – they must be exhaustive and they must be mutually exclusive. "The development of coding frameworks like these involves hard conceptual work and usually a great deal of piloting or trial and error." (Ibid. 236) To Slater semiotics

"- - - represents the exact opposite to content analysis along every dimension. It is closer to interpretive methodologies than to quantitative and survey methods and is utterly open-ended rather than closed in its questions and investigations. It is strong on rich interpretations of single texts or codes but offers almost no basis for rigorous generalization outwards to a population. It argues that elements of a text derive from their meaning from their interrelation within a code rather than looking at them as a discrete entities to be counted. Where content analysis is all method and no theory, hoping that theory will emerge from observation, semiotics is all theory and very few practical guidelines for rigorously employing it. Above all, semiotics is essentially preoccupied with precisely that cultural feature which content analysis treats as a barrier to objectivity and seeks to avoid: the process of interpretation." (Ibid. 237-238)

6.2.2 Component analysis

The interest of this study is in concept systems for terminological analysis, which, according to Nuopponen (1994) is a broad area and can be studied from different viewpoints and with varying tools. Concept systems are of interest not only to terminology science but also to philosophy, the general theory of science, psychology, artificial intelligence, lexicology, semantics, pedagogy, information science, classification theory, database development, etc. (Ibid, 236) Concept systems are part of our everyday life, and often needed to evaluate and to develop:

"Concept systems are fundamental to human existence. They are part of man´s mental activity, of his ability to structure experience and knowledge. The organization of concepts and the linguistic coding of experiences, observations, etc. are especially important to the specialist, since more exacting demands are made on concept

formation and classification in special subject fields than in everyday life. The concepts required in everyday communication are vague and more or less unconscious in comparison with those in special fields. The specialist needs a sophisticated conceptual apparatus and a system of terms to cope with his professional duties. He is also frequently faced with the task of developing, revising, defining, standardizing, structuring, classifying and naming concepts." (Nuopponen 1994, 236)

Nuopponen (ibid.) states, that researchers in the field of terminology science are concerned with the conscious structuring of concept systems as carried out in different special fields, while, for instance, psychologists also investigate unconscious structures. Concept analysis is also done in everyday life without thinking about it (Nuopponen 2003, 1). In this study emphasis is solely on conscious structuring, and that is thus also the case in the word associations.

Component analysis concerns concepts, i.e. the basic starting point and basic for the terminological analysis is a concept, not a term. The starting point is often questioned – would the term be a natural base and is it already visible in the name of the field – terminology. However, in theories and principles of terminology and terminological work the primary nature of concepts is traditionally emphasised, and the founder of terminology Eugen Wüster[44] has stated, that all terminological work begins with concepts, its aim is to make precise distinction between different concepts. (Nuopponen 2003, 2)

In a definition of linguistic or formal equivalence a semantic component analysis is used. It is commonly used by translators and commonly known as denotation analysis[45]. In component analysis the denotative – and sometimes also connotative - meaning of a word is divided into smaller parts, i.e. into semantic components. Component analysis is considered useful especially in polysemous and related cases and in near synonyms. (Häkkinen 1994, 173-175; Vehmas-Lehto 1999, 67; Nida 1975, 32-67; Engelberg 1998; Ingo 1982, 71; Ingo 1990, 143-166)[46]

It is not common to relate in terminological guidebooks and standards how to recognise polysemous, homonymous and synonymous expressions, and although terminological worker should have considerable linguistic competens, it is not an axiom to recognise these expressions. For example Ingo (1982, 77) points out the importance of recognising polysemous expressions. According to him a translator must be especially careful, when one word has several meanings. In linguistics polysemia and

---

[44] Eugen Wüster 1979, 1: "Jede Terminologiearbeit geht von den Begriffen aus. Sie zielt auf scharfe Abgrenzung zwischen den Begiffen." Cited here Nuopponen 2003, 2. Org. title: Wüster, Eugen 1985/1979, Einführung in die Allgemeine Terminologielehre und Terminologische Lexikographie. 2. Auflage herausg. Vom Fachsprachlichen Zentrum, Handelshochschule Kopenhagen.

[45] Vehmas-Lehto (1999, 67) also uses the term *piirreanalyysi* (characteristic analysis). In this study, form component analysis is used, because it may also refer to connotative aspects and as discussed before, it is not always possible and/or reasonable to separate connotative and denotative aspects of a word.

[46] The originators of component analysis were Floyd G. Lounsbury and Ward H. Goodenough, its most important developers were Jerry J. Katz and Jerry A. Foder, from whom its main ideas were transmitted via Noam Chomsky into generative grammar. (Ingo 1982, 144)

homonymia are usually separated. The criterian is whether it is (language historically) a question of several meanings of one word (polysemia) or about accidental, phonological or graphical, similarity of two or more (homonymy), although it is not easy in every case to make a distinction between these two. (Ibid.)[47]

Ingo (1982) states, that for the translator the nature of the ambiguous (polysemous vs. homonymous) is not so significant, but in the analysis of denotation it is more important to study what the word means in a certain context, that is, what sememe of a lexem[48] is it about. The translation equivalents for the ambiguous terms should be sought for at the level of sememes, when the context of a word is crucial in the meaning of the kept characteristics. (Ibid, 77, 81)

Nida (1975) gives an example, where four different words refer to similar objects: *chair*, *stool*, *bench* and *hassock*. These objects have in common that they are made by people, they are articles of furniture and meant for sitting. What is then distinctive? A chair has legs, a back and is for one person. A stool has legs, no back and is for one person. A bench has legs, is with or without back and is for two or more people. A hassock does not have legs or a back and is for one person. Distinctive characteristics are the absence or non-absence of legs and backs and concern how many people they are meant to be used by. (Nida 1975, 69-72, also in Vehmas-Lehto 1999, 67-68) Similarly, for example, the word stay-at-home mother (homemother, *kotiäiti*) can be divided into semantic components as 'woman', 'mother', 'stays at home', 'not in gainful employment', and 'takes care of her own children'.

In component analysis one can use, for instance, a semantic characteristics matrix, where words are defined by semantic characteristics. With a matrix we can recognise the common characteristics of different words and the distinctive components.

In the material used in this study (dictionary definitions, see chapter 2 *The central concepts*) semantic component analysis also has its limitations. According to Häkkinen (1994) some parts of the vocabularies are easily describable with semantic characteristics. Words referring to human beings are often used as examples. For example the words *man, woman, boy* and *girl* can be separated from each other with the help of opposite pairs such as "adult- non-adult" and "masculine-feminine" or with the help of minus(-)- and plus(+) signs like "+/- adult, +/- masculine". (Häkkinen 1994, 174)

---

[47] As homonyms are often considered homograps (graphical similarty) and/or homophones (phonological similarity). According to Häkkinen (1989) it is of no benefit to make a distinction between homograps and homonyms in Finnish, since a homograph is usually also a homophone. However, in English the distinction is needed, since a homonym may be a homograph without being a homophone and vice versa. Cf. for example homograps lead [li:d] '*johtaa*' - lead [led] '*lyijy*' and homophones sight [sait] '*näky*' - site [sait] '*sijainti*'. (Häkkinen 1989, 86)

[48] "A lexeme is the minimal unit of language which has a semantic interpretation and embodies a distinct cultural concept. It is made up of one or more form-meaning composites called lexical units." (Glossary of linguistic terms 2004/2009) "A Sememe is a proposed unit of transmitted or intended meaning; it is atomic or indivisible. A sememe can be the meaning expressed by a morpheme, such as the English pluralizing morpheme -s, which carries the sememic feature [+ plural]." (The Language of Linguistics 2004/2009) (See also Ingo 1982, 77)

The inadequancy of the semantic component analysis is, however, that the numbers of potential characteristics are considerable and partly indefinable. In addition, some parts of the vocabularies are difficult to define with the help of semantic characteristics. For example, different mushrooms, grasses and metals are impossible cases. It is not enough to find distinctive components; the characters should also be as universally applicable as possible. In practice most often synonyms, parallel expressions, euphemisms or sometimes even foreign language translations are used to describe the meaning. In these cases the entities of the objects and the description of the lexical meaning are extremely difficult to keep separated. According to some theories there is no reason to aim at this kind of distinction. (Ibid 1994, 174)

In this study, component analysis will be used in studying the meaning of a word represented, especially in dictionary definitions and also in defining (the degree of ISO standard 5964) equivalence. As case terms referring to human beings and human roles are used, they are partly far more complex than the samples given by Häkkinen (in 1994, 174). The connotative meanings of the Finnish and British-English terms could also be studied with similar methods and tools, but in this study component analysis concentrates mostly on the denotative meanings. This is supplemented with the other methods used in this study, most importantly, with the explanations given by those people interviewed and with discourse analysis.

Below, an example of the way component analysis is used in the study is presented. In the table, the relevant semantic characteristics used in the studied English online dictionary definitions of the keyword *homemaker* are illustrated. Symbol ++ refers to a static, invariable characteristic, + to a dynamic characteristic, which is in definition expressed with words such as "often", "especially":

HOMEMAKER characteristics:	Merrian-Webster	OED Online(a)	OED Online(b)	Cambridge International Dictionary of English	Dictio-nary.com	WordNet	NHD
woman	+	++				++	
mother	+						
wife	+	++				++	
housewife	+	++				++	
family member	+	++	+	++		++	++
manages a household	++	+	++	++	++	++	++
non-paid	+	++		+		++	++
takes care of children	+			+			++

Table 6: *Homemaker's characteristics in dictionaries*

The Cambridge International Dictionary of English (2002), for example, defines homemaker as "a person who manages a home and often raises children instead of earning money from employment". From the definition semantic characteristics are

selected such as 'family member', 'manages a household', 'non-paid', and 'takes care of children'. The two first mentioned are considered as static and thus invariable characteristics, whereas the two last mentioned as dynamic characteristics, which is expressed in the definition with the word "often".

The characteristic common to all of the studied definitions was 'manages a household'. The second most common characteristic was family member' – managing a household is often done within one's own family. The third most common character was 'being unpaid': taking care of household is usually not a paid duty or work. The fourth most common characteristic was 'gender': four out of the six expressed in some way that the homemaker was a married woman. Only one definition was totally neutral in the gender issue while the other ones referred to a female in its example of use. (More analysis in chapter 7.1.1.4 *Homemakers and housewives dictionaries*.)

6.2.3 Co-Word Analysis

Co-word analysis (in Finnish *yhteissana-analyysi*, in Swedish *cowordanalys*) is a quantitative bibliometric method. In brief, bibliometrics (FI: *bibliometriikka*, SW: *bibliometri*) is to study document collections with the help of quantitative methods (Persson 1991, 6). In this study, co-word analysis is used to study indexing frequency and the lexical network of the key terms of this study in several databases, and also in interview material. Attention is also paid to Finnish and British indexing practices.

According to Callon & al. (1991) the methodological foundation of co-word analysis is the idea that the co-occurrence of keywords describes the contents of the documents in a file. Methodologically, it is therefore a question of using one (or more) index (or indices) to measure the relative intensity of these co-occurrences and to achieve simplified representation of the networks to which they give rise. (Ibid. 161) (About the history, origins and scope of bibliometrics see e.g. Forsman 2005; Kärki & Kortelainen 1998; Persson 1991)

Co-word analysis is very similar to co-citation analysis. Co-word analysis deals with co-occurrences of terms in documents, while co-citation analysis deals with shared citations. (Persson 1991, 51; Horton et al. 1998) It is thus about the relatedness of terms rather than documents. (About the methods origin, see e.g. Schneider 2004, 135-136.)

Co-word analysis is used in different kinds of studies. Von Ungern-Sternberg has compared the documentary languages in four international databases in the subject field of biotechnology using co-word analysis and bibliographic coupling (see von Ungern-Sternberg 1998 and 1994). With co-word analysis she studied the co-occurrence of index terms in different articles. Although the method and tools appeared to be very useful, there were also some problems:

"1. The weakness of the databases ---

2. The citing practice ---

> There were many problems in building up the terminology for a documentation language. Different databases use different depth and specificity for their indexing. ---
>
> The two methods for mapping the knowledge structure have certain weaknesses. The results of the two methods differ somewhat. This can partly be explained by the roughness of both methods. The indexing language is built in a hierarchical structure which is not noted in the coword analysis. --- The methods used are quantitative and neglect to some extent qualitative aspects of research. As a result the methods show the knowledge structure according to the number of articles on a topic, but tell us little about the depth or value of the research." (von Ungern-Sternberg 1994, 317-318)

Co-word analysis has also been recently used to examine how to organise keywords in a web search environment. Ding, Chowdhury and Foo (Ding et al. 2000) have used the co-word analysis method to identify the relationships among words and to create keyword maps that may be useful for information retrieval purposes in the domain of information retrieval. The results of the co-word analysis were compared with traditional thesauri to identify the difference. The research shows, that

> --- "the associations of words identified by co-word analysis were different from those obtained from traditional thesauri. --- co-word analysis can catch the changes of its domain area to provide better and timely information guide for users. --- co-word analysis can be used for organizing knowledge through keyword maps and they can be quite useful to compliment the traditional vocabulary tools." (Ding et al. 2000)

Schneider (2004) has studied the methodological and theoretical aspects of bibliometric methods in semi-automatic[49] thesaurus construction. The bibliometric methods investigated were document co-citation analysis, citation context analysis, co-word analysis, and bibliometric ageing methods. When studying the creation of a conceptual network of noun phrases within a concept group by use of co-word-analysis (in a certain sub-field of dentistry), he found, that

> "Frequently occurring noun phrases in a concept group are agreed upon, contextual, and most likely semantically related to each other. In addition, the indirect approach ensures that the noun phrases assigned to a concept symbol appear together because they share the same textual context surrounding the specific reference marker in citing papers. All noun phrases thereby become related to each other, either directly by occurring together in the same context, or indirectly through their common co-occurrence with the cited reference representing the concept symbol. Together all concept symbols and their assigned noun phrases refer to the common concept of the group. Consequently, first and second order co-occurrence analyses can be performed by use of co-word analyses in order to investigate its ability to disclose equivalence, hierarchical, and associative relationships. In addition, network analysis is used as a visual aid for the interpretation of the relational network structures." (Ibid, 321)

Schneider (ibid) concludes, that the "applied bibliometric methods are very suitable for selection of candidate thesaurus terms in the specialty area of periodontology", and further, that "co-word analysis is able to identify thesaural relationships, albeit with a certain error rate", and in addition that, "in all evitability the co-word analysis applied in

---

[49] When automatic approaches are used as a tool for thesaurus constructers, and not as a means in itself, then we speak of semi-automatic thesaurus construction (Soergel 1974). (Cited here Schneider 2004, 14)

the present methodology will detect valid as well as defective definitional relationships." The method is still quantitative, and needs to be supplemented by human intellectual interpretation, which is considered time consuming. – "Please keep in mind it is time consuming to manually identify and verify candidate thesaurus terms, as well as investigate their potential relations." "--- automatic methods can assist in, but not replace, the intellectual effort needed for the construction of an indexing language or a thesaurus". (Ibid, 323-329)

There are two opposite perspectives about using titles versus keywords in co-word analysis. Those who prefer titles want to get rid of the "indexer-effect": indexing is a subjective action and standardised vocabularies used in indexing are sometimes outdated for the present studies. Therefore, words from titles are considered to give more subtle and current conception of ongoing research. From the authors' viewpoint the main idea of the titles is perhaps to create as high a level of precision as possible so that the most unique feature will be distinguished. The indexers task is thus the opposite – to find the balance between precision and recall so that seekers will not miss any relevant information and on the other hand will not be too loaded with irrelevant information. Thus indexers discover those documents, whose titles are insufficiently formulated. (Persson 1991, 52)

Persson (1991) specifies several problems in the use of title words, and points out that they are often undervalued. Sometimes authors form provocative and rhetorical titles in order to get more attention. The titles are also not standardised, which means that one concept can be described by several terms, and on the other hand one term can have several meanings. Furthermore, titles also include a lot of trivial words which separated from the context have no special meaning, and in the analysis there is a subjective stage if the researcher sometimes has to modify material (for example select between singular and plural forms of the words). (Ibid, 52-53)

Whittaker (1989) prefers the use of keywords in co-word analysis and states, that

> "As long as titles are available, co-word analysis can be performed and may be expected to yield coherent results. For document sets which are tightly focused on a relatively narrow research area, title words are probably entirely adequate, but they may not be capable of revealing the rich detail of more heterogeneous fields or document sets. For such material, properly chosen keywords are clearly superior. --- On the negative side, title words suffer from some serious drawbacks. The main one is that the set of themes addressed by a long and complex article is less likely to be fully expressed in the title than in a set of keywords, so that the aim of the co-word analysis is to expose the network of concepts (and the like) to the fullest comprehensible extent, title words would almost inevitably yield inferior results. Lesser, but still important, problems arise from the non-standardization of title words and the subdivision of concepts which may occur. The present study has shown that sensible results can be obtained despite these difficulties, but in the end the keyword analysis provided substantially more information." (Whittaker 1989)

In this study the aim is to discover the similarities and differences of the use of certain Finnish and English indexing terms and keywords with the help of a co-word analysis. The final aim is to see how these practices and needs diverge and how they can be served in a multilingual thesaurus. This study focuses on metadata (present indexing

practices) and thus keywords (indexing terms, descriptors) are used. It is still good to be aware of the indexer effect and time delay factors when discussing indexing discourse. The indexers may have out-dated tools and lexicology when compared with social scientists, who use up-to-date vocabulary and can apply the entire range.

For the study's purposes it is essential that the words used in searches appear in certain fields (titles, indexing terms, author keywords etc). Iannella (1998) reminds us that metadata is *structured* data about data and states, that "this structure is the crucial element that gives metadata the edge over full-text indexing. The benefit is that the structure alludes to the semantics of the metadata".

The tool used in co-word analysis is Bibexcel (Bibexcel 2002), developed by Olle Persson. It is also used to visualise the results. With these co-word maps, produced with Bibexcel, it is possible to make observable what has been occluded or invisible. One still has to keep in mind, that "[it] is not the "object" that presents the figure itself; rather, figures stand out relative to interest, attention, cultural and macroperceptual features" (Chen 2003, 25).

With Bibexcel the terms that occur the most (e.g. descriptors) are listed first, and then how often they co-occur together is counted. With the help of MDS-technique (multi-dimensional-scaling) the co-occurrences can then be represented as a map, where the distance between terms is vice versa in relation to the number of terms that occurred simultaneously. – The more the terms co-occur together the closer they are placed in the map. Because the frequency of a single word strongly influences co-occurrences, the effect of word frequencies has been eliminated via a normalising distance unit (the observed frequency divided by the expected frequency). (Persson 1991, 56)

The maps are used to create a picture of the studied words and their relationships. However, it is important to remember that a picture can also be experienced in different ways.

The number of aspects of visual thinking can be stated - vision is a unique source for thinking – insight, foresight, hindsight, and oversight. Before we focus in words, we examine the overall picture. (McKim 1980, cited here Chen 2003, 12) One cannot represent the whole picture of a phenomenon, since the maps are results of a complex selection process. With maps one can shed light on the phenomenon. In maps all the nuances are usually not included, and some information may be lost. If the analysed material consists of dozens or even hundreds of descriptors, it is clear that one picture cannot include them all in a sense-making way. In this study the terms represented in maps are of many kinds. In the case of descriptors (indexing terms) the number of terms shown in maps is sometimes limited to descriptors that co-occur most often (within every map is reported the minimum number of co-occurrences to be included in the map and how the terms represented are selected). Because of the richness in variations of natural language expressions, in the case of word associations a need is expected to occur for some modification and standardisation. – When a term occurs in the dataset both in a plural and in a singular form, the most common version is selected. If they occur evenly, the plural form is selected. If the word associations are more similar to stories than terms, they are keyworded. In addition, Finnish terms will be translated into

English, since the publication language is English and the reader of the study is not expected to know Finnish. Word associations are thus expected to demand a lot of subjective decisions (cf. subjective stage in Persson 1991, 52-53). It is noteworthy, that the word association maps are used rather to provide an interpretation and illustration of the material than to represent it as authentic. Semantic lexicological maps of the thesauri studied are supposedly most authentic and neutral in the respect to their origin, since only modifications have been performed of possible different versions in the use of the plural versus the singular form.

The maps have another significant advantage; the thematic maps improve recall for textual information and inferences (Rittschof et al., 1994). (See also Chen 2003, 6.)

Co-word analysis is thus used according to the purposes of this study: the aim is at finding which and what kind of concepts and terms co-occur with the certain indexing terms and what is their frequency and networks. It is thus used to gain contextual information 1) about the content of the studied concept and 2) about the context of their use, i.e. about the documents, which are indexed with certain descriptors (*family roles, breadwinners, heads of households, homemakers, housewives*).

6.2.4 Focused Interview

Interviewing is a typical method of collecting qualitative data, but it can be used also to gain quantitative data. The aim of the interview is to find out what someone has in mind. It requires discussion, but it is initiated and lead by the researcher. (Eskola & Suoranta 1998, 86)

The aim of the project is to find out the translatability of British indexing terms into Finnish. To achieve a broader perspective and a better understanding of the meaning (connotation and denotation) and use of the key terms studied, people representing different cultures and subcultures will be **interviewed**. The aim is thus to find out the tacit assumptions and practices, which are linked with the key terms. Terminologies and thesauri usually operate at the denotation level of the words, but human communication – both formal and informal – uses also connotative meanings. (About connotative and denotative meanings see chapter 3.1 *On culture, language and meaning*.)

The interview types can be divided and defined in many ways. Eskola and Suoranta (1998) have used as criteria for their division, on one hand, the level of solidity in question formulation and, on the other hand, to what degree the interviewer structures the interviewing situation. This leads to four types of interviews:

1. In a **structured interview** or form interview (FI: *strukturoitu haastattelu, lomakehaastattelu*) the formulation of questions and the representation order is the same for each interview. This procedure is based on the idea that the meaning of the questions does not vary according to the interviewee. The alternative answers are also given. The situation therefore is similar to a guided filling in of a questionnaire.

2. A **half-structured interview** (FI: *puolistrukturoitu haastattelu*) is different from a structured interview in the respect that although the questions are the same for each interviewee there is no ready-made alternative answers and the interviewee thus answer with their own words.

3. In a **focused interview** (FI: *teemahaastattelu*) the subject matters, themes, have been defined beforehand. There are still no exactly formulated questions and asking order as in the structured interview. The interviewer has determined the interviewed themes on beforehand and uses a kind of checking-list, but the order and the specificity of the questions varies in the interviews.

4. The **open interview** is the one that is very similar to a common discussion. The interviewer and interviewee discuss a certain topic, but e.g. all the subject matters are not touched on with every interviewee. (Eskola & Suoranta 1998, 87)

Because of the aim of the study the method used was a **focused interview** (in Finnish *teemahaastattelu*, see Hirsjärvi & Hurme 2000, 9, 47). It was originally developed to meet certain problems arising from communications research and propaganda analysis (Merton & al. 1990, 5).

The focused interview has its advantages in this specific research context. According to Merton, Fiske and Kendall (1990, 3) the focused interview differs in several respects from other types of research interviews, which might appear similar at first glance. In broad outline, the distinguishing characteristics of the focused interview are:

> "First of all, the persons interviewed are known to have been involved in *a particular situation* --- Secondly, the hypothetically significant elements, patterns, processes and total structure of this situation have been provisionally analysed by the social scientist. Through this *content or situational analysis*, he has arrived at a set of hypotheses concerning the consequences of determinate aspects of the situation for those involved in it. On the basis of the analysis, he takes the third step of developing an *interview guide*, setting forth the major areas of inquiry and the hypotheses which provide criteria of relevance for the data to be obtained in the interview. Fourth and finally, the interview is focused on the subjective experiences of persons exposed to the pre-analyzed situation in an effort to ascertain *their definitions of the situation*. The array of reported responses to the situation helps test hypotheses and, to the extent that it includes unanticipated responses, gives rise to the fresh hypotheses for more systematic and rigorous investigation." (Merton & al. 1990, 3-4)

The distinctive prerequisite of the focused interview is thus a prior analysis of the situation in which subjects have been involved (Merton & al. 1990, 4). The criteria of the effective focused interview are briefly:

> "1. *Range*. The interview should enable interviewees to maximize the reported range of evocative elements and patterns in the stimulus situation as well as the range of responses.
>
> 2. *Specificity*. The interview should elicit highly specific reports of the aspects of the stimulus situation to which interviewees have responded.
>
> 3. *Depth*. The interview should help interviewees to describe the affective, cognitive and evaluative meanings of the situation and the degree of their involvement in it.

4. *Personal context*. The interview should bring out the attributes and prior experience of interviewees which endow the situation with these distinctive meanings." (Merton & al. 1990, 11-12)

In this study, the participants interviewed were selected accordingly to the first characteristics: they were known to have experience concerning the studied terminological matters. – One interviewed group has experience as indexers, the other as thesaurus constructors and the third as researchers/specialists, which are known to be familiar with the studied concepts and terms. Secondly, before constructing the interview guide the author studied the context of the phenomenon – what is characteristic for the Finnish and British indexing practices and how are the terms studied used and understood in Finnish and British indexing in general. This was done mainly by comparing several thesauri and studying the use of certain terms in Finnish and British indexing (co-word analysis). The sociological context and dictionary definitions were illustrated and analysed before the actual interview process took place.

Hirsjärvi and Hurme (2000) have discussed the views about human beings and their relation to reality in connection with focused interview. They have concluded eleven ontological and epistemological aspects from the methodological viewpoint:

1. **The ability to symbolise the world is the most important characteristic when describing a human being.** Due to their consciousness, human beings are able to create symbols, change their meanings and pass on the knowledge to the next generations. The most advanced symbol system is language. In its nature language is social and contractual: it is based on the need to communicate meanings, and it necessitates both the ability to create symbols and understand symbols created by others. This point has seldom been taken into consideration when discussing research interviews.

2. **The starting point for human research should be organismic models instead of mechanistic models.** Humans can set down new goals and also change because of their experience and interaction with their surroundings. Humans are not only reactive but also active. Therefore, because surroundings can differ also humans and their change also vary. The historicality of human beings should also be taken into consideration in research either through a longitudinal study or by explaining the implications of history.

3. **Human behaviour and human actions are multiform and reflect an open system, of which humans form a part.** When thinking of the causality in human research we have to take into consideration the intentionality as well as the factors systematically connected with each other.

4. **Reality is socially constructed**. There is no absolute reality, common to all humans. Social constructions are also related to science and scientific concepts. Even if we wanted to, we still cannot separate ourselves from language: it is inbuilt in the world surrounding us.

5. **We cannot find the absolute truth**. In accepting point 4, social construction, and knowing that the same phenomena are understood differently at different times and in different cultures, we also accept the research results are more of an approximation, thus not representing the absolute truth. What researchers and the scientific community consider as truth changes over time.

6. **Research concerning humans beings involves meanings; interpretation and understanding are the central methods.** The interpretation and understanding are problematic in research, and the researcher must pay serious attention to their position in the research unity.

7. **Presumed inter-subjectivity**. As every human being creates his own idea regarding reality, one can never be absolutely sure about another person's ideas. Generally, in order to be able to act, we must presume that we can understand things similarly to other people, that is, we must postulate the inter-subjectivity.

8. **The researcher is a part of social reality and thus also a part of the research process**. The researcher influences the research during the different steps: the choice and interpretation of the concepts, data collection and analysis, and the reporting. The researcher can thus be considered one of the research tools. For instance, in sociological and behavioural sciences the researcher is also a layman, and the concepts learned as a layman also affects the research. Although it must be admitted that the subjective ways of thinking will affect the research process, the research must aim at reflecting the researched world. The interpretations should be supported in the world under research, and the researcher should be able to recognise his/her place.

9. **Power relations affect science, which is thus value-bound**.

10. **Research is more contextual rather than universal**. In social and behavioural sciences it is not realistic to aim at universal principles because of the different ideas of reality constructed by different individuals and cultures. The research should consider these contextual factors as similar to history and culture.

11. **Research should aim at ecological validity**. The research must be done in the field, where people live their everyday life and act. It must also consider the individual contexts because it is only from them that we can understand individual meanings. (Hirsjärvi & Hurme 2000, 16-19)

According to the ideas represented by Merton, Fiske and Kendall (1990), and Hirsjärvi and Hurme (2000) the context plays a key role in this study.

As a part of the interview method, the free **word association method** will also be used. Respondents will be asked to give freely and spontaneously approximately five response words to certain stimuli words. The aim is to obtain information about the studied terms semantic network and to work as a starting point for further discussion about the terms meaning and usage.

This method has also been used earlier in the context of thesaurus construction. According to Lykke Nielsen (2002) word association method may result in a usable and workable thesaurus, performing as well as a thesaurus designed by traditional thesaurus construction methods. The advantage of the word association method is that it may enrich thesaurus construction by providing current, relevant, and domain-specific information. (Ibid.)

When summing up previous word association research (relevant to thesaurus construction) Nielsen (2002) states, that the "method provides to the thesaurus compiler:

> "A varied and complex set of related terms to the stimulus, reflecting the perspective of the knowledge domain to which the respondent belongs
>
> A large set of variant forms for the stimulus word, reflecting the language use of individual user and the knowledge domain, respectively." (Ibid, 71)

She also points out, that

> "--- the word association method is a semi-automatic and economic method to collect terms for thesaurus construction and to identify relationships among terms, but the later semantic analysis demands considerable cognitive, intellectual resources." (Ibid. 73)

In accordance with the scope of the study the Finnish and British practices will be emphasised in the interviews. Interviews will be made preferably at the interviewees' place of work.

As discussed in chapter 3.4 *Translatability and equivalence* in this study different types of equivalence are used as conceptual tools to analyse especially the interview material. The aim is to develop a division of different equivalence *aspects*, which would be of use in practice – both in construction and in analysis of thesauri.

As previously mentioned the attention in this study is on human effort and therefore special attention is paid to different **choices** and **actions**. Koller's division (1989, 1995, see also Zethsen 2004) of different types are kind of layers and perspectives for the analysis of empirical material in this study and lead to **further questions from five different perspectives**:

> 1. Denotative
>
> - Have the two words (or expressions) exactly the same truth value? Do they refer to the same concept at the denotative level?
>
> 2. Connotative
>
> - What impact does the connotative level have? Are the two words equally positive/neutral? Can the connotative level be superior to the denotative level (in choice-making of equivalents)?
>
> 3. Text-normative
>
> - How does the text-type affect? Do the equivalent words represent documentary languages?
>
> 4. Pragmatic
>
> - Is the translation accepted and usable in the target context (in the way it was in source context)? Is the effect same? - Does the translation produce similar results e.g. in information seeking as in the source version?
>
> 5. Formal
>
> - Are the hierarchies the same in the source and target languages? - How is the hierarchical thesaurus system considered? (If, for example, the hierarchy in the target language (Fi) does not conform to the hierarchy in the source language (Eng) how has the problem been handled?)

In the interview analysis attention is also paid to possible prioritising between these different aspects.

6.2.4.1 Simulated indexing task

As a part of the interview a **simulated indexing task method** is used. Four studies are of particular interest for this part of the study – those made by Saarti (1999), Iivonen (1989, 1995) and Borlund (2000). The models and experiences of these studies are briefly illustrated in the following.

Iivonen (1989) has studied how indexing is dependent on the indexing environment. She (1989, 8) used ten test participants from different libraries: they were asked to index certain documents and give the reasoning behind each indexing choice. The test individuals indexed ten documents according to four different tasks:

a) The documents were required to be indexed accordingly to the test person's normal working situation *without* any controlled vocabulary.

b) The documents were requested to be indexed accordingly to the test person's normal working situation *using* a controlled vocabulary.

c) The documents were requested to be indexed *without* any controlled vocabulary the way they should be found in information seeking situation in their own library.

d) The documents were requested to be indexed *using* a controlled vocabulary the way they should be found in information seeking situation in their own library. (Ibid.)

After every indexing situation the test persons were asked to give their motivations for each indexing term they selected. The indexing situations a and b were conducted on the same day. There was a four weeks break before the indexing situations c and d were carried out, since it was hoped that the test persons would forget in the meantime the indexing terms they selected in indexing situations a and b. (Ibid, 9)

Iivonen (1989) showed how the intraindexer consistency (consistency of a same person) is greater than interindexer consistency (consistency between different persons). The consistency between different indexers was significant on the concept level and about the major theme of the document. Consistency is thus likely to be greater in the conceptual analysis step and less in the term selection step.

Iivonen (1995) has also studied consistency in the formulation of query statements in online bibliographic retrieval. Similar to indexing, here "consistency means the degree of agreement in performing the same search task on various occasions, and is considered in the study as intersearcher and intrasearcher consistency" (Ibid, 281).

The empirical data consisted of test situations, where searchers with different backgrounds (24 experienced searchers, each from different work place, and 8 students of information studies of University of Tampere) analysed 12 search requests and formulated query statements from them for a search in the Finnish database KOTI. After

two months the same test persons analysed eight of the same search requests again and formulated query statements from them for a search in the KOTI database. As search terms they could use both descriptors taken from YSA and free-text terms. After test situations the searchers had to explain their query statements. The study showed that there is inconsistency in the formulation of query statements. However, the intrasearcher consistency was considerably higher than intersearcher consistency. Here also the concept consistency in the formulation of query statements was higher than term-consistency. Iivonen (1995) states, that

> "It is important that in future studies, too, consistency is considered as a practice. The analysis of the nature of inconsistency is more useful than astonishment at its existence. It requires, however, the ability to accept and understand dissimilarity. Seeing the variety and alternatives in information retrieval offers, however, elements for a better understanding of information retrieval as a whole." (Ibid, 294)

As stated before, when discussing the consistency of indexing it is also important to keep in mind that the aim of the documentary languages, e.g. thesauri, is to increase **predictability**, not consistency. Even bad indexing may be highly consistent. (See e.g. Iivonen 1989, 12, Fugmann 1994, 420-421)

Saarti (1999) has used a similar method in a study on indexing and abstracting of fiction in the public library environment. 30 test participants from five different libraries were studied and five fictional novels were used for the experiment. Half of the test persons were working in the library and half were library users, clients. The sample was constructed accordingly to the qualitative research model, thus its scientific criterion was rather a qualitative representativeness than statistical scope. (Ibid, 99-101)

The test participants were asked to index and to abstract the novels. The abstracting and indexing took place in their own respective libraries and the researcher was present at the test situation. Books were at first abstracted and immediately after that indexed using the Kaunokki-thesaurus (Swedish version Bella). In each library the process lasted about three hours. (Ibid.)

Saarti experienced his methods and the triangulation of evidence to be fruitful. He thought that the touchstone for the research results is not statistical generalisations, but how well the conclusions are justified and how well, in this sense, they can be generalised. (Ibid.)

Borlund (2000) has studied the evaluation of interactive information retrieval systems (IIR systems) and used as a method a *simulated work task situation*. According to her (ibid, 24) "it serves two functions: 1) triggers and develops a simulated information need by allowing for user interpretations of the situation, leading to cognitively individual information need interpretations as in real life; and it 2) is the platform against which the information need is updated. Further, by being the same for all test persons experimental control is provided".

In the study five topically different simulated situations were used. One information need was real and personal (developed by a test person) and four were simulated information needs. The test participants were volunters: altogether 31 university students from various academic fields and educational levels. Four of them

participated in the pilot experiment. 24 participated in the main experiment. (Ibid, 104-111)

The collected and analysed data of the main experiment were 24 sets of responses to the questionnaires as to demographics and searching skills of the test persons, 120 protocol of shorter verbal statements, and 120 corresponding search transaction logs. (Ibid, 112)

A similar holistic research design is not considered realistic in this study section, since e.g. in Borlund's study (ibid, 113, cf. also Saarti 1999, Iivonen 1989 and 1995) the participation took three hours on average per test person, and in this study the indexing task is just a part of a larger entity. Some of its (Borlund 2000, see also Ingwersen 1994) empirically based recommendations (with respect to the employment of simulated situations/simulated work tasks) still work as a methodological reflection tool. It was recommended, for example, to employ both the simulated situation/simulated work task situation *and* real information needs within the same test, to permute the order of search jobs between the test persons. in order to avoid the possible bias of the relevance assessments due to human relevance assessment behaviour, and to pilot test prior to actual testing. (Ibid, 167-169)

In the interviews the terminological case is conducted both without any reference to a certain material, and with a reference to specific kinds of material. Some of the references were known to be familiar to the informants from their working tasks. The working order is kept the same with all the respondents. – The material to be indexed was delivered beforehand and with identical instructions. In addition, in the interviews the case was discussed in a specific order the indexing task being conducted after more material-free discussions. (The research design was also discussed with several actual respondent-like test-persons before the actual research interviews.)

For the purpose of this study a way to study indexing practices and the use of certain concepts in monocultural versus multicultural indexing and information seeking environment was to ask the indexers to index the documents first as their "ordinary task" and then ask them to index the same documents for European social scientists. After the indexing tasks, they were then asked the reasoning for the selected concepts and indexing reference terms. It is essential that the interviewed indexers could use in simulated indexing situation their normal tools and methods. The time distance between delivering the indexing task and making the interview was rather short, from few days up to two weeks.

- **Summary and Conclusions**

In this study ***discourse analysis*** is understood not only as a method, but as a theoretical framework, that allows different emphasis in the focus and methodological applications. Discourses are here understood to refer to language use in a social context, and although discourse is a substantive, it means also doing something and is an active process.

In this study the *focused interview* is used to gain information about:

Indexing

- How are the studied terms understood by the indexers? How do they use them and why?

- How do they consider the meaning and usage of the terms in multicultural (European) indexing environment?

- What tools are used? What are the reasons and motivations mentioned?

- How would the Finnish indexers translate the British terms into Finnish and why?

- What are the indexers' assumptions about equivalence?

- Academic language usage versus common language usage and indexing practice

- How are the studied terms understood by the specialists? ('Specialists' refers to the researchers, teachers, professors etc in academic institutions.) How do they use them and why? Are they aware of other discourses in the use of the studied terms in content management and how do they consider them?

- How do they consider the meaning and usage of the terms in the multicultural publishing environment?

- How would the Finnish specialists translate the British terms into Finnish and why would they use these terms?

- What are the specialists' assumptions about equivalence?

Thesaurus construction

- How are the studied terms understood by the thesaurus constructors? How do they use them and why? What would be the thesaurus context (subject field, BT, NT etc.)?

- How do they consider the meaning and usage of the terms in multicultural indexing environment and other discourse types?

- What tools and methods are used? What are the reasons and motivations mentioned?

- How would the Finnish thesaurus constructors translate the British terms into Finnish and why?

- What are the thesaurus constructors' assumptions about equivalence?

The **word association method** is also used as a part of the interview method in order to obtain information about the studied terms semantic network and as a starting point for further discussion about the terms meaning and usage.

Simulated indexing task situation is used to gain information about

- How Finnish and British indexers index documents related to caregiving roles? – How they use terms related to roles and gender?

- Are there any differences when indexing for different environments?

In the recognition of the formal, linguistic meaning and further equivalence type a **semantic component analysis** is used. It is commonly used by translators and known also as denotation analysis. In component analysis the denotative meaning of the word is divided into smaller components, semantic characteristics. A component analysis is commonly considered helpful especially in polysemy cases and in co-ordinated and related cases.

Bibliometric co-word method is used to gain information about:

1. What are the documents about indexed with certain terms? What are the certain terms about according to their use in indexing? What are the similarities and differences in the use of certain terms in British and Finnish databases? - The emphasis is both on indexing practices (depth, used vocabularies and indexing guidelines) and on the contents of the retrieved documents (aboutness).

2. What is the conceptual context of the studied terms, according to several commonly used thesauri, and to Finnish and British informants?

6.3 Material

The material is collected according to relevance criteria and thus theoretical sampling has been used. The material is qualitative and its collection principles have been derived from the theoretical framework. The material should, therefore, allow multiple strategy approach and represent different types of discourses, which are relevant to consider when constructing a multilingual and multicultural thesaurus.

This study concentrates on the translatability of British-English social science index terms into the Finnish language. The focus is on terms representing abstract concepts. Accordingly to the starting point for the study the studied terms were selected from (the working version of) the multilingual thesaurus, ELSST (European Language Social Science Thesaurus), which was being translated from English into Spanish, German and French at the time when this research was planned. The Finnish Social Science Data Archive (FSD) was then interested in also translating it into Finnish.

The terms for this study are selected according to **a theme - family roles.**

The selected terms for family roles will be studied in five ways:

1. Three databases are selected in order to obtain pools of documents about the topics in the above-mentioned theme: British COPAC and Finnish ARTO and LINDA.

2. Nine thesauri are used to examine and compare the indexing terms used in the received documents in the database searches: Thesaurus of Inter-University Consortium for Political and Social Research (ICPSR),

Sociological Abstracts Thesaurus (CSA), The UNESCO Thesaurus, Eurovoc, European Language Social Science Thesaurus (ELSST), ERIC, General Social Science Thesaurus (SOSIG), Humanities and Social Science Electronic Thesaurus (HASSET) and General Finnish Thesaurus (YSA).

3.   Three Finnish and Three British indexers will be asked to index the same documents, that is, 5 articles retrieved from Finnish and British databases having the studied terms as one of the central ones (e.g. in title and/or as a descriptor and thus likely to be used in indexing).

4.   Finnish and British indexers, thesaurus constructors and social scientist (in all 29 participants) are asked to give word associations, definitions and equivalents for the studied terms.

5.   Six online dictionaries are studied to find out how the theme studied is represented and generally understood.

In the next section the selection criteria and motivations are discussed in more detail together with the selected terms, databases, thesauri and documents.

6.3.1 Family roles terms

By nature the study is holistic, aiming at a wide-ranging perspective within one specific problem area. As already stated, at the general level the study is about translatability and equivalence in the context of multilingual and multicultural thesauri. When studying translatability, there has to be a starting point – a translation unit with which to start.

In order to have, on the one hand, a heterogeneous sample from the viewpoint of equivalence problems and translatability, on the other hand, a homogeneous sample from the viewpoint of the sociological context, the selection criterion is a theme. Within this group are presumed to be found several different translation problems, which occur on different levels – on the concept, term and/or on the indexing term level. As a theme family roles have been selected, which seem to represent *gender related terms* and more precisely, terms related to families and parenthood and also to gender in working life.

Accordingly to the idea of theoretical sampling five indexing terms were selected for further examination: **family roles, breadwinners, heads of household, homemakers,** and **housewives.**

The above terms represent several different translation problems from the viewpoint of a multilingual thesaurus constructor. They are used as a starting point when examining the Finnish and British indexing practices and several thesauri. Translation problems are expected to occur because the division into different roles seems foreign and gendered from the Finnish aspect. The case is considered to represent a typical translation problem within abstract social science discourse.

In chapter 4.2 *The social background* is discussed a part of the sociological background of the above mentioned terms from the point of view of gendered

participation in labour life and on paternity leave modes. In order to understand the differences in indexing practices it is essential to be aware of the sociological context (Finnish and British practices) of the studied terms. In accordance to the research questions the emphasis is on Finnish practices and therefore it is these that are stressed.

6.3.2 Thesauri

As a context and in order to achieve a broader perspective a total of nine thesauri have been selected. The aim is to find out the similarities and differences in them - using as a sample the theme described in the previous sub-chapter 6.3.1.

When selecting thesauri the aim was to find both general and specific as well as commonly used indexing and information seeking tools, which are available online. In accordance with the aim of the study – to study translatability and equivalence into Finnish - the selected thesauri are mainly focused on social sciences and English. All are available online, via the Internet. Most are designed for the use of a specific database and collection, some for a much wider audience. All are well known in the Finnish university library context.

The nine thesauri used to examine and compare the representation of the studied terms (case family roles) are:

- multilingual UNESCO Thesaurus, Eurovoc and ELSST

- monolingual HASSET (Br), SOSIG (Br), CSA Thesaurus of Sociological Indexing Terms (En-Am/In), ICPSR Subject Thesaurus (En-In), ERIC (En-Am) and YSA (Fi).

Accordingly to the research design more emphasis is given to YSA. The thesauri studied will be illustrated, in the following section, in more detailed, such as, for example, the content, the purpose, the constructor.

The empirical terminological case is supposed to represent a problematic case for the informants. In its evaluation, tools adopted from translation science are used. Their application calls for special care, since thesauri as a text type are very unique and different from that which is traditionally considered a translation object when discussing equivalence matters within translation science. Although the typologies used are not directly transferable to the new area of application, multilingual thesaurus construction, they are considered to provide the necessary analytical tools and perspectives for the analysis, which cannot be found in traditional LIS literature. Finally, it is important to keep in mind that translation units do not only represent different languages and cultures, but also different discourses and sub-cultures (see chapter 3.4 *Translatability and equivalence*).

YSA

The YSA Thesaurus (*Yleinen suomalainen asiasanasto*) is a general thesaurus in Finnish and it covers all fields of research. The Thesaurus is maintained by the National

Bibliographic Services. YSA has been used for indexing Finnish publications since 1987 in public and scientific libraries and data archives. (Helsingin yliopiston kirjasto 1999)

YSA includes approximately 14 000 preferred terms and 3 000 non-preferred terms. It has been created following the SFS 5471 standard concerning construction and maintenance of Finnish language thesaurus (*Suomenkielisen tesauruksen laatimis- ja ylläpito-ohjeet*). (Ibid.)

YSA is meant for the indexing and information retrieval of books, articles, electronic material and other material types. The vocabulary is stated to help information recorders and seekers to use a shared language. Its purpose is also to be a general source vocabulary when developing special vocabularies. (Ibid.)

YSA is included in VESA – *Verkkosanasto* (VESA-Webbthesaurus), which also includes special thesaurus for music and Swedish translation (*Allärs thesaurus*). (Ibid.) In the study YSA is used online.

HASSET

The UK Data Archive for use with its retrieval system BIRON has developed a Humanities And Social Science Electronic Thesaurus, HASSET. (BIRON, Bibliographic Information Retrieval Online, is a WWW interface providing access to a complete source of information about studies in the UK Data Archive's collection.)

The purpose of HASSET (see UKDA 2002) is in accordance to general aims of thesauri (cf. chapter 3.3).

HASSET includes over 4,000 preferred terms, 2,500 non-preferred terms, 260 standalone terms and 330 hierarchies, and it is based on the UNESCO thesaurus (ibid.). It is also broadly used outside the UKDA – e.g. material used in the SOSIG's General Social Science thesaurus[50] is derived from HASSET.

ELSST

Language independent metadata browsing (LIMBER) of several European social science data-archives use the multi-lingual ELSST (European Language Social Science Thesaurus), derived and translated from the current UKDA HASSET. The aim is to reduce the present HASSET hierarchies and remove all cultural and institutional specificities. In addition, new areas such as methodology will be added. (Miller & Matthews 2001)

> "--- The resulting broad-based social science thesaurus will be suitable for use by any resource in the social science domain. Due to time limitations, a target of 1500 preferred terms from a minimum of 20 hierarchies has been set. The thesaurus will also include all synonyms to these terms and all top terms of hierarchies in the existing

---

[50] See URL: <http://sosig.ac.uk/help/thesaurus.html>

HASSET that either map to existing thesauri or which, although not in the major 20 hierarchies, would have been present if resources were available. Each hierarchy will be sent to the CESSDA archives for evaluation of coverage and usefulness.

As each hierarchy is reduced it will be translated --- Although it is hoped that, at this broad level, one-to-one equivalence will be possible for the vast majority of terms, the format will allow for non-equivalence and different structures in each language. Extensive use of scope notes will resolve ambiguities, translation assumptions and subject coverage of hierarchies. The translated hierarchies will be sent to the appropriate archives of CESSDA for evaluation and addition of language specific synonyms." (Ibid.)

ELSST is divided into thematic parts, which (in the working version) are: Addiction, Age groups, Attitudes, Disadvantaged groups, Discrimination, Ethnic groups, Equipments, Families, Family environment, Housing, Offences, Economics, Labour and employment, Political institutions, Political systems, Politics, Social problems, Social structure, Social welfare, Sociology, Analysis, Conflict, Data, Demography, Development, Emotional states, Environmental changes, Environmental sciences, Human behaviour, Identity, Life histories, Methodology, Nationality, Quality, Businesses, Consumption, Education, Educational environment, Health, Human behaviour, Human settlement, industries, membership, Population migration, Products, Property, Ownership and tenure, Religion and Resources.

In September 2001, the working version of ELSST included approximately 1,500 preferred terms, 860 non-preferred terms and 270 standalone terms divided into ten major hierarchies and thirteen additional hierarchies.

In September 2009, ELSST is being used in the Madiera portal, which provides unified access to European data resources. About the thesaurus is informed:

"The ELSST is a multilingual social science thesaurus. It is available in German, Danish, Greek, English, Spanish, Finnish, French, Norwegian and Swedish. It includes more than 3000 terms.

There are two versions of ELSST in the portal. The ELSST version matches on keywords only. The ELSST Free Text version matches on a few key text fields e.g. title, abstract, keywords, variables and subject." (MADIERA 2009)

SOSIG General Social Science Thesaurus

SOSIG is funded by the Electronic Libraries Programme and by the Economic and Social Research Council. It is based in the Institute for Learning and Research Technology at the University of Bristol, and has been used as a model for the creation of several UK based gateways in other subject areas. (Worsfold 1999)

The Social Sciences Information Gateway (SOSIG) aims at locating "high quality sites on the Internet, which are relevant to social science education and research." It provides three different thesauri to aid in searching. In this study, the General social science Thesaurus is used. This was developed to "provide alternative terms that will generate hits in the SOSIG Internet Catalogue." The SOSIG Thesaurus is derived from

the HASSET Thesaurus, and developed by the UK Data Archive (SOSIG 2001) in co-operation with several institutions:

> "SOSIG will be working with the UK Data Archive, IBSS [13] , the Centre for Economic Performance at LSE [14] and Qualidata [15] to establish a social science Thesaurus based on HASSET. Keywords used by the contributing services which are not currently held in HASSET will be submitted as candidate terms which can then be included in updates of the Thesaurus, thereby increasing the value of the Thesaurus to all users of the various services and avoiding duplication of effort." (Hooper 1997)

It is constructed according to the generally accepted principles.

Eurovoc

The Eurovoc thesaurus is published in the official languages of the European Community and it thus includes English and Finnish. All the languages have equal status. The Eurovoc Thesaurus covers the fields in which the European Communities are active, i.e. politics, international relations, European Communities, law, economics, trade, finance, social questions, education and communications, science, business and competition, employment and working conditions, transport, environment, agriculture, forestry and fisheries, agri-foodstuffs, production, technology and research, energy, industry, geography, international organisations. (EC 2001)

Eurovoc is divided into 21 fields and 127 microthesauri including altogether 6,075 descriptors (of which 508 are top terms). In English it also includes 516 scope-notes and 5,672 non-descriptors. In Finnish it includes 628 notes and 4,817 non-descriptors. (EC 2001). It is continuously being updated, and in 2009 (V4.3, last visited 29.1.2009) Eurovoc comprised 6,645 descriptors (of which 519 are top terms). The English version had 759 scope notes and 6,769 non-descriptors, while Finnish version had 859 scope notes and 5,445 non-descriptors. (Ibid.)

In addition to equal status between different languages, as regards equivalence it is stated, that "there is no equivalence between the non-descriptors in the various languages, as the richness of the vocabulary in each language varies from field to field". The equivalence relationship between descriptors and non-descriptors is shown by the commonly used abreviations: "UF" (Used For and "USE"). It is also stated that the equivalence relationship covers relationships of several types, such as "genuine synonymity, or identical meanings; near-synonymity, or similar meanings; antonymy, or opposite meanings; inclusion, when a descriptor embraces one or more specific concepts which are given the status of non-descriptors; because they are not often used". Thus the numbers of non-descriptors and scope notes vary from language to language. (Ibid.)

The Eurovoc thesaurus has been compiled in accordance with the standards of the International Standards Organization, namely ISO 2788-1986 - *Guidelines for the establishment and development of monolingual thesauri*; and ISO 5964-1985 - *Guidelines for the establishment and development of multilingual thesauri*. What is noteworthy, and contary to common practice, is the preference for the singular form -

"You can look for a descriptor or a non-descriptor using an expression, a term or part of a term. Enter the term(s) in the singular, then click on "search". (Ibid.)

UNESCO Thesaurus

The UNESCO Thesaurus is developed by the United Nations Educational, Scientific and Cultural Organisation. It covers the subject fields of education, science, culture, social and human sciences, information and communication, and politics. It is constructed to facilitate subject indexing in libraries, archives and similar institutions. It was first published in 1977 and the second edition was issued in 1995. (UNESCO 2001)

UNESCO Thesaurus is widely known and used, and has also been used as a basis for other thesauri, e.g. for HASSET.

ERIC

ERIC, the Education Resources Information Center, is an online digital library of education research and information. It is sponsored by the Institute of Education Sciences (IES) of the U.S. Department of Education. "ERIC provides ready access to education literature to support the use of educational research and information to improve practice in learning, teaching, educational decision-making, and research." (ERIC 1999, 2009)

ERIC is constructed according to the general thesaurus construction rules and actively updated. ERIC is used world-wide.

> "The Thesaurus of ERIC Descriptors (Thesaurus) is a controlled vocabulary - a carefully selected list of education-related words and phrases assigned to ERIC records to organize them by subject and make them easier to retrieve through a search. Searching by Descriptors involves selecting relevant terms from this controlled vocabulary to locate information on your topic."

> "ERIC has an ongoing commitment to maintain the Thesaurus of ERIC Descriptors (Thesaurus). In addition to adding new terms, ERIC may modify the status of an existing Descriptor if it has been rarely used in indexing, overlaps with other terms, or becomes obsolete. ERIC may also reinstate terms if necessary. These changes, generally based on literary warrant, are considered routine maintenance, and are in accordance with standard practices of thesaurus development and maintenance as outlined and defined in ANSI/NISO Standard Z39.19-2005, Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies. Other Thesaurus maintenance activities include updating hierarchical relationships and adding or changing scope notes." (ERIC 2009)

CSA Sociological Abstracts Thesaurus

The **Sociological Abstracts Thesaurus** contains an alphabetical listing of Main Term descriptors used for indexing and searching the SA database and printed index,

beginning with the April 1986 issue. (CSA 2001) It is produced by Cambridge Scientific Abstracts, CSA, which is a privately owned information company located in the United States with offices in Hong Kong, France and the UK.

The thesaurus is constructed according to the thesaurus construction standards. When compared to thesauri in general it is exceptional that the Sociological Abstracts Thesaurus also includes detailed historic notes, which "provide the range of years in which a term was in use, its former Descriptor Code, and the word form if it has changed. Often they provide search instructions. History Notes appear for both Main Terms and discontinued terms" (CSA 2001).

ICPSR

Inter-University Consortium for Political and Social Research, ICPSR, was established in 1962. ICPSR is the world's largest archive of digital social science data and an active partner in social science research and instruction throughout the world. It acquires, preserves, and distributes original research data, and also provides training in its analysis. It also offers access to publications based on its data holdings. Physically it is a unit within the Institute for Social Research at the University of Michigan, but it is a multinational organisation. - ICPSR is a membership-based organization, with over 640 member colleges and universities (including Finnish Universities via Finnish National Membership, Finnish Social Science Data Archive, FSD), as well as around the world. (ICPSR 2009a)

ICPSR provides a thesaurus, which is composed of three separate lists: Subject Thesaurus, Personal Names Authority List, and Geographic Names Thesaurus. In this study only Subject Thesaurus is used.

> "Subject Thesaurus is an alphabetical listing of social science subject terms. The scope of this thesaurus is multidisciplinary and is intended to reflect the subject range of the ICPSR archive. Social science disciplines represented include: political science, sociology, history, economics, education, criminal justice, gerontology, demography, public health, law, and international relations." (ICPSR 2009b)

ICPSR has compiled a bibliography of reference documents and thesauri that were used to prepare the ICPSR controlled vocabulary system, and links are provided to these (PDF 12K) on their website. Development of the ICPSR Thesaurus was supported by the National Science Foundation (SES-9977984). The structure and format conventions used to construct Subject Thesaurus follow the recommendations outlined in the Guidelines for the Construction, Format, and Management of Monolingual Thesauri, Z39.19-1993 (NISO 1993). (Ibid.)

The sources consulted (reported in Thesaurus refs. Sullivan Feb. 7, 2002) were various, 25 were listed, and they include several classics in the field of thesaurus construction, such as Aitchison, Gilchrist and Bawden (2000): *Thesaurus Construction and use: A Practical Manual*; Cleveland and Cleveland (2001): *Introduction to Indexing and Abstracting*; Hjorland (1997); *Information Seeking and Subject Representation: an Activity-Theoretical Approach to Information Science*; and Lancaster (1991) *Indexing*

and Abstracting in Theory and Practice. The thesauri used were also various, and totalled 26, out of which 24 are online. They represent commonly known thesauri and are mostly in the field of social sciences.

6.3.3 Dictionaries

The six English dictionaries selected were gathered from the link pages of Finnish translation science departments (most from The School of Modern Languages and Translation Studies in University of Tampere) and university libraries. – The purpose is to use common language dictionaries, which are considered useful also in academic contexts and freely available in university networks. – Links in university web-pages are thus seen as a guarantee of (sufficient) high quality and usability. In addition, one Finnish-English dictionary is studied, commonly used in the Finnish universities and provided by university libraries. The emphasis is not on dictionaries, and they are neither evaluated nor compared, but used to provide information about how the studied terms are understood in dictionaries *in general*. The dictionaries were studied in 2002-2003 and checked for possible changes at the end of the study, which is 6/2009.

The aim was to include British, American and international English dictionaries. In practice many of the dictionaries linked in the Finnish university web-pages are American-based, but aim at international content, which is also seen in this study. There was a great variety in information provided to the user in their web-pages. All reported the name of the publisher, but most did not inform the users of how many keywords they include, or how the dictionary is updated, and what corpus was used etc.

The studied online dictionaries, most freely available on the Internet and all commonly known and widely used, are:

1 Newbury House Dictionary of American English

<URL: http://nhd.heinle.com/>

Heinle's Newbury House Dictionary of American English contains over 40,000 entries. It represents short definitions and sample sentences and idioms. (Newbury 2003)

2 WordNet

<URL: http://wordnet.princeton.edu/>

WordNet is a large lexical database for the English language provided by Princeton University, Cognitive Science Laboratory. The most recent version is 3.0, published in 2006, in which the total of all unique noun, verb, adjective, and adverb strings is 147278. (WordNet 2003 & 2009)

3 OED Online

<http://dictionary.oed.com/>

OED Online is the Oxford English Dictionary (online version), published by Oxford University Press. It claims to represent international English. It has been available from March 2000, and provides authoritative definitions of over 500,000 words, traces the usage of words from their first recorded occurrence to the modern period through 2.5 million quotations from a wide range of international language sources, and it offers etymological analysis and detailed listings of variant spellings and uses the International Phonetic Alphabet to show pronunciation. (OED 2004, 2009)

OED OnLine is not freely available, but needs a private subscription or e.g. a connection provided by a university network.

4 Merriam-Webster OnLine

<URL: http://m-w.com/>

Merriam-Webster Online Dictionary is produced by Merriam-Webster Online.

> "The Merriam-Webster Online Dictionary is based on the print version of Merriam-Webster's Collegiate® Dictionary, Eleventh Edition. The online dictionary includes the main A-Z listing of the Collegiate Dictionary, as well as the Abbreviations, Foreign Words and Phrases, Biographical Names, and Geographical Names sections of that book. It also includes 1,000 illustrations and 25 tables. Selected sections of the print Collegiate Dictionary, notably the Signs and Symbols section, are omitted from the online Collegiate Dictionary because they include special characters and symbols that cannot readily be reproduced in HTML." (M-W OnLine 2004, 2009)

5 Cambridge Advanced Learner's Dictionary

<URL: http://dictionary.cambridge.org/>

The Cambridge Advanced Learner's Dictionary is produced by Cambridge University Press, "dedicated to the development of resources for a changing educational environment", and stating to be "a leading educational publisher in the UK, publishing high-quality, curriculum-based books and software for secondary school in the UK and for international schools". (Cambridge 2004, 2009)

6 Dictionary.com

<URL: http://dictionary.reference.com/>

Dictionary.com is produced by Lexico Publishing Group, LLC. It is a multi-source dictionary search service, thus using several other dictionaries. (Dictionary.com 2003, 2009)

> "The dictionaries that appear on Dictionary.com include:

* Dictionary.com Unabridged (v 1.1), Based on the Random House Unabridged Dictionary, © Random House, Inc. 2006.

* Webster's New Millennium Dictionary of English, Preview Edition (v 0.9.6) Copyright 2003-2006 Dictionary.com, LLC

* Dictionary.com Word of the Day

* Dictionary.com Crossword Solver

* The American Heritage® Dictionary of the English Language, Fourth Edition Copyright © 2000 by Houghton Mifflin Company. Published by Houghton Mifflin Company. All rights reserved.

* The American Heritage® Dictionary of Idioms by Christine Ammer. Copyright © 1997 by The Christine Ammer 1992 Trust. Published by Houghton Mifflin Company.

* The American Heritage® Stedman's Medical Dictionary Copyright © 2002, 2001, 1995 by Houghton Mifflin Company. Published by Houghton Mifflin Company.

* WordNet ® 2.0, © 2003 Princeton University

* Investopedia.com. Copyright © 1999-2005 - All rights reserved. Owned and Operated by Investopedia Inc.

* Wall Street Words: An A to Z Guide to Investment Terms for Today's Investor by David L. Scott. Copyright © 2003 by Houghton Mifflin Company. Published by Houghton Mifflin Company.

* Merriam-Webster's Dictionary of Law,© 1996 Merriam-Webster, Inc.

* Merriam-Webster's Medical Dictionary, © 2002 Merriam-Webster, Inc.

* Webster's Revised Unabridged Dictionary, © 1996, 1998 MICRA, Inc.

* The Free On-line Dictionary of Computing, © 1993-2005 Denis Howe

* Jargon File 4.2.0

* U.S. Gazetteer, U.S. Census Bureau" (Ibid, 2009)

In the analysis of Dictionary.com only results that are not overlapping with the other studied dictionaries are included.

7 MOT

<URL: http://www.kielikone.fi/en>

MOT is a Finnish-English dictionary, constructed by specialists on the English language. It is mainly a general language dictionary, but also includes special vocabularies on various fields such as medicine, law, economics and ecology. It makes distinction between American and British English. The first used version in 2002 is 4.0, and the most current version of MOT used is 4.7 from 2007. Version 4.0 included about 110,000 Finnish and English keywords and version 4.7 includes about 115,000 Finnish keywords and 110,000 English keywords. (Kielikone Ltd. 2002, 2009)

6.3.4 Databases

Three databases were selected in order to obtain pools of documents. The databases used are British COPAC and Finnish ARTO and LINDA. The studied databases are not identical – the contents, depth and ways of indexing differ a considerably.

1 COPAC

COPAC is a British union online catalogue of members of the Consortium of University Research Libraries. Today 26 institutions contribute records to COPAC (University of Manchester 2004). The indexers do not conform to any established or commonly accepted indexing guidelines, or any controlled vocabularies.

2 ARTO

ARTO is a reference database of Finnish articles (an index of articles from approx. 1000 Finnish journals). Altogether about 40 institutions index in the ARTO database. (Helsinki University Library 2005) A common indexing tool used is YSA, but also special vocabularies are used when needed.

3 LINDA

LINDA is the union catalogue of Finnish University Libraries and also includes the National Repository Library, the Library of Parliament, the Library of Statistics and Lahti Science Library. It contains references on monographs, serials, cartographic materials, audiovisual materials, electronic resources, multimedia and archives. (Helsinki University Library 2005). As in the case of ARTO, the numbers of institutions conducting indexing for the LINDA database are extremely many, and a common indexing tool (YSA) is used and in addition also special vocabularies may be used.

6.3.5 Informants

The people interviewed were different kinds of professionals, representing social science researchers, indexers and information specialists with thesaurus construction experience, who were both Finnish and British. Within this study this meant that there were six kinds of sub-groups:

> 1 a) Finnish **social scientists** known to be familiar with family role matters.

> 1 b) British social scientists known to be familiar with family role matters.

> 2 a) Finnish **indexers** doing social science indexing

2 b) British indexers doing social science indexing

3 a) Finnish thesaurus constructors (i.e. information specialists with experience on thesaurus construction) mainly representing social sciences

3 b) British thesaurus constructors (i.e. information specialists with experience on thesaurus construction) mainly representing social sciences

The participants, therefore, represented six different views – that of social science experts and of (social science) information specialists representing Finnish versus British cultures. The respondents needed to represent several different institutions in order to get a rich (both in terms of depth and breath) view of the phenomenon. They were also required to be experienced in their subject area.

Thus, some of the informants represented thesaurus specialists having experience in indexing and/or in thesaurus construction and the other informants represented family role research specialists from a variety of social science perspectives. Half of the respondents were Finnish, half were British. The aim was thus to interview specialists representing different cultures and sub-cultures in order to study the different aspects of multilingual and multicultural social science thesaurus construction and use. Accordingly to the perspective of this study, the emphasis here also rests on the Finnish context and usability (pragmatics).

The primary way of contacting informants was sending an e-mail and if necessarily to call them. In the e-mail, the aim of the research was briefly described, but not the precise terms and procedures in a detailed way (see appendix 9: Example letter of the interview request).

Before the actual research interviews, the ideas and questions were tested in Finland by interviewing two social scientists and an indexer with experience in thesaurus construction. At the final stage of the interview process a Finnish statistician representing the statistics office and a British technical thesaurus constructor were also interviewed. All these five interviews were important for the final design of the interview questions and analysis. However, they are not part of the actual research material. The actual research material consists of 27 focused interviews and a total of 29 participants. (See more about the participants in chapters 7-9, which report the results.)

6.3.5.1 Indexed documents

Originally the idea was that the selected documents should represent the studied topic (family roles) and include the studied terms as central and key terms. 'Key terms' refers here to terms, which are also supposed necessary in the indexing of the documents. Translation problems may occur due to the linguistic, cultural or institutional differences and the common characteristic for them is that the aboutness is somehow difficult to determine in multicultural environment homogeneously.

After studying some of the documents the topic 'family roles' was narrowed to a caregiver role for two reasons: firstly, the corpus term *homemaker* was considered most interesting and concentrating on this would allow deeper insights, and secondly, the amount of work (both for the indexers as informants and for the author) was considered to be more tolerable when focusing more on a single concept.

The documents were selected after the preliminary study of some Finnish and British indexing practices. The selected documents were all articles that could be easily posted and enabled the reading of the whole original document. The **five articles** to be **indexed** and **studied** are:

> 1. Arnlaug Leira 2002: Updating the "gender contract"? Childcare reforms in the Nordic countries in the 1990s' NORA no. 2 2002, Volume 10, pp. 81-89.

> 2. Bridges Judith S. & Etaugh Claire & Barnes-Farrell Janet 2002: *Trait judgments of stay-at-home and employed parents: A function of social role and/or shifting standards?* Psychology of Women Quarterly, 26:2 (2002), pp. 140-150.

> 3. Nock Steven L. 2001: *The Marriages of Equally Dependent Spouses*. Journal of Family Issues, Vol. 22 No.6, September 2001, pp. 755-775.

> 4. Jalovaara Marika 2002: *Socioeconomic Differentials in Divorce Risk by Duration of Marriage*. Demographic Research, Volume 7, Article 16, Published 29 November 2002, URL: <www.demographic-research.org>, pp. 537-564.

> 5. Smith Calvin D. 1998: Men Don't Do This Sort of Thing. A Case Study of the Social Isolation of Househusbands. Men and Masculinities, Vol. 1 No. 2, October 1998, pp. 138-172.

Due to the starting point and the nature of the research the selected articles are all published in English and selected because the authors were both Finnish- and English-speaking. English as a language area and social sciences (in it closest different approaches to family sociology) as a field of study are also understood here in broad terms. After a content analysis the articles were considered also as to their capacity to represent the terminological theme studied in this section – caregiving role – in its broadest sense – although, here, naturally not in an exhaustive way. The number of articles was decided because of the term case and time limits – on the one hand several articles were needed in order to acquire different aspects for the studied terminology, and on the other hand it is not realistic to expect indexers to be able to use too much time for this kind of extra work.

The original abstracts of the articles are provided in appendix 5.

6.3.5.2 Process

Due to the qualitative nature of the study, the research plan is under process during the empirical studies and several sub-questions arose, such as the question of genderism in the context of thesaurus construction.

In the British guides for non-discriminatory language use for academic staff, *homemaker* is considered a modern and gender-neutral equivalent for the term *housewife* (see e.g. UCL 2000, Brunel University 1999, School of Psychology 2002, University of Southampton 1999). In dictionaries (which represent a common language usage and "consensus") *homemaker* is represented as a woman and usually as a synonym for *housewife*.

Examples:

"Gender has traditionally been associated with the words for particular roles, for example 'foreman' and 'housewife'. Wherever possible it is best to use gender-neutral terms such as 'supervisor', 'carer', 'homemaker' or another suitable alternative. " (Brunel University, 1999)

Homemaker=one who manages a household **especially as** a wife and mother (Merriam-Webster 2002)

Homemaker=A housewife, **esp**. one in charge of the domestic arrangements (as opp. to a paid housekeeper)**; also**, one who manages a household. So home-making n. and a. (OED Online 2002)

In this context some questions, which rose during this study, were: Should the thesaurus constructors of today also consider other language usage norms than the thesaurus construction standards? Do thesauri represent an isolated language discourse? If social scientists are advised to use in other academic contexts gender-neutral term forms, should they when seeking information and keywording for their own documents use terms conforming thesauri practices? – In thesaurus construction guidelines and standards genderism is not discussed, but the main criteria for good descriptors are terms expressing most precisely the concept and preferred equivalents most commonly known by the users.

A more detailed interview guide was constructed after the examination of the use of studied terms in Finnish and British indexing (see appendices 1-6). The original plan was to have a minimum of 3-4 Finnish and 3-4 British interviewees from the above-mentioned groups.

Before the actual interviews, the interview guides were discussed in research seminars held at the home institution (several participants had broad experience of library and thesaurus matters), and also two social scientists and one indexer who had experience of thesaurus matters were met with and test-interviewed. – In the meeting not only were the interview guides tested, but the research design was also discussed.

Although the emphasis is on the Finnish language and culture, the British aspect (as a source language and culture and representing several discourses) was equally

needed, and therefore there is no significant difference in the amount of Finnish versus British informants.

- **Summary and Conclusions**

The terms for this study are selected according to the criterion of theoretical sampling, and they represent the theme *family roles*. The selected terms are *family roles, breadwinners*, *heads of household*, *homemakers* and *housewives*. They were studied in several ways:

1) Three databases were selected in order to obtain pools of documents about the topics in the above-mentioned theme: British COPAC and Finnish ARTO and LINDA.

2) Nine commonly used thesauri (mono- and multilingual, mono- and multicultural) were used to examine and compare the indexing terms used in the received documents in the database searches: Thesaurus of Inter-University Consortium for Political and Social Research (ICPSR), Sociological Abstracts Thesaurus (CSA), The UNESCO Thesaurus, Eurovoc, European Language Social Science Thesaurus (ELSST), ERIC, General Social Science Thesaurus (SOSIG), Humanities and Social Science Electronic Thesaurus (HASSET) and General Finnish Thesaurus (YSA).

3) The Finnish and British indexers were asked to index the same documents, that is, articles retrieved from Finnish and British databases having the studied terms as one of the central themes (e.g. in the title and/or as a descriptor and thus likely to be used in indexing).

4) The Finnish and British indexers, thesaurus constructors and social scientist (tot. 29 informants) were asked to give word associations, definitions and equivalents for the studied terms.

5) Seven online dictionaries were studied to find out how the theme studied is represented and generally understood.

7 Equivalence and family roles in common and academic language usage

In order to gain knowledge of how the terminological case, family roles, is understood and used in common and social science language, several commonly used online dictionaries (cf. chapter 6.3.3) were studied. Additionally, British and Finnish social scientists representing different social science disciplines were interviewed (cf. chapter 6.3.5) in order to find out how the studied terms are understood and defined/translated in different cultures and subcultures, and in order to study how the given equivalents differ from each other. For the purposes of gaining knowledge on what is pragmatic equivalence in multicultural thesaurus construction equivalence and the aims of translations were also studied.

As stated previously, especially in chapters 2, 3 and 6, the study approaches certain terminology and language use from several different perspectives. In semantic analysis one is interested in what the words conventionally mean (see Yule 1996, 114; Hudson 1998, 176), whereas pragmatic analysis focuses on what is actually meant in specific circumstances (Yule 1996, 127). The literal material, i.e. dictionaries, gives the closest answers to the questions of semantics, and the oral material, i.e. interviews, the closest to the questions of pragmatics. In addition, equivalence issues are studied from several perspectives, which can also be seen to represent different degrees of decision-making in translation processes as well as in analysis.

The six British social scientists interviewed represented four universities, located in three cities. The six Finnish social scientists interviewed represented three different universities locate in three cities. Despite the differing departmental names the Finnish and British social scientists were considered to be equivalent as groups both in terms of scope and depth. It is important to realise that the department divisions do not reflect their perspective or focus of study as such. For example, in the department of sociology the area of the informant's study could also be considered a gender study or to belong to economic geography. The respondents were selected based on their research areas, which were considered relevant from the study's perspective and on their experience. In the Finnish group the gender division was more even in the planning stage, but men tended more often to neglect the interview inquiry. The demographics of the respondents representing social science researchers is summarised in the table below:

	Finnish (6 respondents)	British (6 respondents)
Cities	3	3
Universities	3	4
Genders	5 women / 1 male	4 women / 2 male
Educational levels	4 doctors / 2 licentiates	6 doctors
Departments	4	5

Table 7 : *Demographics of the Finnish and British social science respondents*

In the examples from the interview material, the British social scientists are referred to with acronym BR SS and the Finnish in similar manner with FI SS. – The style of abbreviations is the same with the other studied groups – instead of SS indexers are represented as IN and thesaurus specialists as TS.

It is noteworthy, that due to the qualitative material and the small numbers there is no statistical significance in the representation and analysis of the bibliometric co-word maps.

7.1 Dictionary meanings and equivalents

> *"The question is," said Alice, "whether you CAN make words mean so many different things." "The question is," said Humpty Dumpty, "which is to be master -- that's all."*
> - Lewis Carroll, Through the Looking-Glass

The focus of this chapter is to see how several online dictionaries represent the studied terminological case family roles and what kind of idea of equivalence is behind the studied term representations.

7.1.1 Background

As stated before, the six English – British, American, and international - dictionaries selected were gathered from the link pages of Finnish university libraries and translation science departments (see chapter 6.3.3 *Dictionaries*). In addition one Finnish-English dictionary, commonly used in the Finnish universities and provided by university libraries, is studied. In the study, the emphasis is not on the dictionaries, as they are only used to provide information about how the studied terms are understood in dictionaries in general. The dictionaries were studied in 2002-2004 and checked for possible changes at the end of the study, 6/2009.

The collected samples were further explored with the help of component and discourse analysis (about the methods see chapters 6.2.1 *Discourse analysis* and 6.2.2 *Component analysis*). The aim of the analysis was to find out the extension of the

studied concepts and the essential characteristics, especially from a common language perspective. In this section the main emphasis is on how the terms are understood in dictionaries, but additionally other language use examples, illustrative and relevant from the social science (thesaurus) perspectives, are also given.

The dictionaries studied varied in depth. Some provided several explanations or definitions for the keyword, some only one, or in some cases listed only synonymous terms. However, the idea behind all the studied dictionaries was similar: to provide a linguistic description of the concept and outline the extension of the concept and its objects (cf. TSK 1999). (The function is actually rather similar with thesauri – to provide information about how the keyword is understood and used in a certain context, and sometimes even with a historical background.) It is noteworthy, that the length of the samples did not necessarily correlate with depth and specificity.

7.1.2 Family roles in dictionaries

In the studied dictionaries, the keyword *family role* was not found and thus did not have any definitions or equivalents, whereas as separate keywords both "family" and "role" were found. With the separate words the results were similar, Merriam-Webster Online Dictionary providing the most informative results.

In Merriam-Webster Online Dictionary the keyword "**role**" was clarified to mean:

"1 a (1) : a character assigned or assumed (2) : a socially expected behavior pattern usually determined by an individual's status in a particular society b : a part played by an actor or singer

2 : a function or part performed especially in a particular operation or process <played a major role in the negotiations>

3 : an identifier attached to an index term to show functional relationships between terms"

"role."

- (Merriam-Webster Online Dictionary 2004)

In the case of roles within a family, the role refers rather unambiguously to "a socially expected behavior pattern usually determined by an individual's status in a particular society". However, in the dictionaries the concept of family is much more complicated.

In this case, the Merriam-Webster Online Dictionary again provided the most informative sample, and gave 23 entries for the keyword "**family**". As a noun it was explained to refer to:

"1 : a group of individuals living under one roof and usually under one head : HOUSEHOLD

2 a : a group of persons of common ancestry : CLAN b : a people or group of peoples regarded as deriving from a common stock : RACE

3 a : a group of people united by certain convictions or a common affiliation : FELLOWSHIP b : the staff of a high official (as the President)

4 : a group of things related by common characteristics: as a : a closely related series of elements or chemical compounds b : a group of soils that have similar profiles and include one or more series c : a group of related languages descended from a single ancestral language

5 a : the basic unit in society traditionally consisting of two parents rearing their own or adopted children; also : any of various social units differing from but regarded as equivalent to the traditional family <a single-parent family> b : spouse and children <want to spend more time with my family>

6 a : a group of related plants or animals forming a category ranking above a genus and below an order and usually comprising several to many genera b in livestock breeding (1) : the descendants or line of a particular individual especially of some outstanding female (2) : an identifiable strain within a breed

7 : a set of curves or surfaces whose equations differ only in parameters

8 : a unit of a crime syndicate (as the Mafia) operating within a geographical area"

- (Merriam-Webster Online Dictionary 2004)

The first and fifth samples, household and traditional family, represent the way family is commonly seen in the other discourses studied. As a compound, family + roles is on a general level considered to be equivalent to the other discourses studied and thus to refer to roles or task within a family or household.

7.1.3 Breadwinners and Heads of household in dictionaries

According to Merriam-Webster the keyword *breadwinner* dates back to 1771. In two dictionaries it is represented as having several meanings, one dictionary representation of which is similar with the thesauri studied (cf. 9.1.2 *Breadwinners and Heads of Household in thesauri*), and the other meaning "a means (as a tool or craft) of livelihood (Merriam-Webster and OED Online).

In the table below, the semantic characteristics, the definitions used in the six dictionaries are illustrated, wherever the meaning is relevant to the study. (Symbol ++ refers to a static, invariable characteristic, + to a dynamic characteristic, which is in definition expressed with words like "often", "especially".)

BREADWINNER characteristics:	Merriam-Webster	OED Online	Cambridge International Dictionary of English	Dictionary.com (a)	Dictionary.com (b)	Word-Net	NHD
in working life / have income or earnings	++	++	++	++	++	++	++
supports others	++	++	++	+	++	++	++
Have children/dependants	++	++	++	+	++	++	+
male		+	+				
female							+
family member	++		++				

Table 8: *Breadwinner's characteristics in dictionaries*

According to the studied dictionaries *breadwinner* refers to a person, who has an income and earnings, and who supports others. The dictionary definitions do not often directly express her/him to be a family member, but it happens via dependants - in practice a breadwinner supports her/his family. In principle breadwinner is represented as a gender-neutral word (see example 1 below), and in the examples given it occurred once as a feminine (see example 2 below) and twice as a male word, out of which one was not value-free (see example 3 below).

Example 1:

"S: (n) breadwinner (one whose earnings are the primary source of support for their dependents)"

(WordNet 2002)

Example 2:

"a person who works to support others: She is a breadwinner who supports two children."

(NHD = Newbury House Dictionary of American English 2002)

Example 3:

"the member of a family who earns the money that the family needs:

Men are often expected to be the breadwinner in a family."

(Cambridge Advanced Learner's Dictionary 2002)

MOT (4.0 and 4.7) offers for the keyword *breadwinner* the Finnish equivalents *"perheen elättäjä"* (supporter, in more of an economic role) and *"perheen huoltaja"* (provider, more in a caretaking role). The shorter form *elättäjä* also includes *breadwinner* among its equivalents, whereas *huoltaja* does not:

MOT: *elättäjä* = supporter, provider, meal ticket, breadwinner *Hän on tämän perheen ainoa elättäjä* He is the sole supporter of this family

MOT: *huoltaja* = huoltaja s 1 supporter, provider 2 guardian, care giver, carer, caretaker (am)

■ s 3 oik person having care and control, person having custody

According to the studied online dictionaries *breadwinner* is part of common language usage and refers to an economic provider role within a family. No changes during the researched period (2002-2009) were found.

The keyword **head of household** resulted in remarkably less material than *breadwinner* did in the studied online dictionaries. There were no results in four dictionaries (Newbury House Dictionary of American English, OED Online, Merriam-Webster OnLine, Cambridge Advanced Learner's Dictionary), and in one case was provided extremely brief information (WordNet: "head of household (the head of a household or f amily or tribe)"), and therefore no component analysis table was produced. Only one of the six dictionaries studied provided an informative explanation about the kewords content. Dictionary.com resulted in two entries, one from Investopedia.com (example 1 below) and the other from Merriam-Webster's Dictionary of Law (example 2 below), in both of which the meaning was understood similarly:

Example 1:

"Someone who is unmarried and resides with a dependent.

Investopedia Commentary

Head of household status can result in a lower tax rate in certain jurisdictions. "

(Investopedia.com, cited here Dictionary.com 2004)

Example 2:

"Main Entry: head of house·hold

: an unmarried individual who is not a surviving spouse and who maintains a household which for more than one-half of the taxable year is the principal abode of a person who qualifies as a dependent under section 2(b) of the Internal Revenue Code"

(Merriam-Webster's Dictionary of Law, cited her Dictionary.com 2004)

MOT (4.0 and 4.7) does not give any Finnish equivalent for the keyword *head of (the) household* nor for *head of (the) family*, although it can be retrieved when seeking with Finnish word *perheenpää* (MOT: "1 head of the family 2 householder 3 Big Daddy").

According to the studied dictionaries *head of household* is not commonly used in everyday language, and thus not a synonym for *breadwinner* in terms of use and style. No changes during the researched period (2002-2009) were found. The question therefore is, why is it so commonly used in social science thesauri? Although not common in everyday language, is it a social science term? (See later chapters 9.1.2 *Breadwinners and Heads of Household in thesauri*; 9.2.3.2 and 9.3.3.2 *Breadwinners and Heads of household*.) The above dictionary samples represent *heads of household* as a term of administration and statistics.

It transpired that as a term of statistics *head of household* has a defined meaning, as presented in the "Principles and Recommendations for Population and Housing Censuses" produced by the United Nations:

> "Definition of: head of household [code 181]
>
> Long name head of household or reference member of households
>
> That person in the household who is acknowledged as such by the other members. Countries may use the term they deem most appropriate to identify this person (head of household, householder, household reference person, among others) as long as solely the person so identified is used to determine relationships between household members. The notion of head of household assumes that most households are family households and that one person in such family households has primary authority and responsibility for household affairs and is, in the majority of cases, its chief economic support. Where spouses are considered equal in household authority and responsibility and may share economic support of the household, the concept of head of household is no longer considered valid even for family households. In order for the relationship among members of the household to be determined under these circumstances, it is essential that either (a) the members of the household designate one among them as a reference member with no implication of headship or (b) provision be made for designation of joint headship where desired." (UN 2003)

The definition is rich in its aspects, and thus also the pragmatic meaning is clarified, and the term is seen as a historical term rather than as a current practical term. Its meaning as a statistical term is very close to the common language meaning of *breadwinner* as a family role and especially as a family's "chief economic support".

7.1.4 Homemakers and Housewives in dictionaries

The studied keyword ***homemaker*** was found in all the online dictionaries studied. The definitions of *homemaker* given in the studied online dictionaries are rather similar, e.g.:

- "one who manages a household especially as a wife and mother" (Merriam-Webster 2002)
- "A housewife, esp. one in charge of the domestic arrangements (as opp. to a paid housekeeper); also, one who manages a household. So home-making n. and a." (OED Online 2002)

In the table below the relevant semantic characteristics used in the definitions of the dictionaries studied (component analysis) are illustrated. (Symbol ++ refers to a static, invariable characteristic, + to a dynamic characteristic, which is in definition expressed with words like "often", "especially".)

HOMEMAKER characteristics:	Merriam-Webster	OED Online(a)	OED Online(b)	Cambridge International Dictionary of English	Dictionary.com	Word Net	NHD
woman	+	++				++	
mother	+						
wife	+	++				++	
housewife	+	++				++	
family member	+	++	+	++		++	++
manages a household	++	+	++	++	++	++	++
non-paid	+	++		+		++	++
takes care of children	+			+			++

Table 9: *Homemaker's characteristics in dictionaries (component analysis)*

The characteristic common to all of the studied definitions was 'manages a household'. The second most common characteristic was family member' – managing a household is often done within own family. The third most common character was 'being unpaid': taking care of household is usually not a paid duty or work. The fourth most common characteristic was 'gender': four out of the six expressed in some way that a homemaker was a married woman. The fifth most common characteristic was 'taking care of children'. Homemaker is thus often a parent taking care of her/his own children, but not necessarily. Only one definition was totally neutral in the gender issue while the other ones referred to a female in its example of use.

- Most gender-neutral:

a person who manages a home and **often** raises children instead of earning money from employment

(Cambridge International Dictionary of English 2002)

- Second most gender-neutral:

person who takes care of his or her family's house and children: Zora is a homemaker, and her husband is a teacher.

USAGE NOTE: The term homemaker describes anyone who creates and cares for a home. It is a more contemporary term than housewife or househusband.

(Newbury House Dictionary of American English 2002)

As synonyms for *homemaker* in dictionaries the words provided were *housewife, lady of the house, woman of the house*.

In most of the studied dictionaries (except OED Online, and as later seen in chapter 9.1.3, in every thesauri) the keyword *homemaker* is represented as an unambiguous word, although it is clearly in its nature polysemous. No large changes were found

during the study period 2002-2009: Dictionary.com resulted in one more example, originating from Random House Dictionary, and in line with previous examples studied, but also expressing its polysemous nature and relatedness to housekeeper:

> "1.   a person who manages the household of his or her own family, esp. as a principal occupation.

> 2.   a person employed to manage a household and do household chores for others, as for the sick or elderly.

> Origin:

> 1885–90; home + maker

> Usage note:

> 1. See housewife."

> (Dictionary.com 2009)

When using dictionary translations from one language to another the results are often (in the case of family role terms) misleading – as in the case of *homemaker*. In the Finnish-English-Finnish dictionary (MOT Englanti 4.0 suomi-englanti and also version 4.7) homemaker is translated as 1. *perheenemäntä* (*~homemaker/housewife*) and *kotiäiti* (*stay-at-home mother*). Its near-equivalent housewife is translated as *kotirouva, perheenemäntä, kotiäiti*. According to these equivalents, the concepts that the terms represent are to a great extent the same, but contrary to other studied discourses *housewife* would be broader at its extension of meaning:
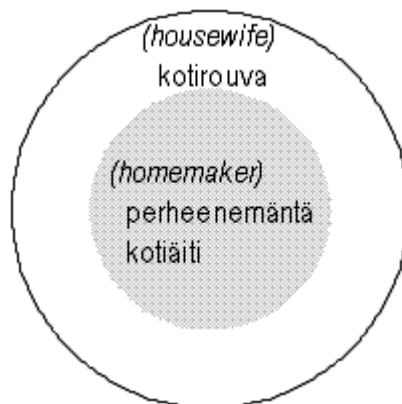
Figure 10: *Homemakers vs. housewives in dictionary translations*

Of all the studied keywords *homemaker* emerges as the most complex one, with regard to gender, in translations. How should the problem be treated according to **language usage guides?** How is the word represented **in sociological surveys?**

In the British guides, for non-discriminatory language use for academic staff, *homemaker* is considered a modern and a gender-neutral equivalent for the word *housewife* (see e.g. UCL 2000; Brunel 1999; School of Psychology 2002; University of Southampton 1997).

"Gender has traditionally been associated with the words for particular roles, for example 'foreman' and 'housewife'. Wherever possible it is best to use gender-neutral terms such as 'supervisor', 'carer', 'homemaker' or another suitable alternative." (Brunel 1999)

Thus, the word *homemaker* has very different connotations – strong, neutral, feminine etc. For example, the Guide to non-discriminatory language usage (UCL 2002[51]) of the University College London advises (as in the previous example of Brunel, 1999) to avoid the term *housewife* and to use instead *homemaker*, *shopper*, and *consumer* - depending on the context. *Homemaker* is then considered to express neither a gender nor a marital status.

Is the term *homemaker* commonly used in surveys? Despite of (previously discussed) recommendations in the large international surveys studied (from the turn of the millennium) such as the World Value Surveys and Eurobarometers, the term *homemaker* is not used. Instead, in the original English versions the term *housewife* is used, which is translated into Finnish as *kotiäiti* (stay-at-home mother).

En: Being a <u>housewife</u> is just as fulfilling as working for pay. (WVS 2000(a), V116)

Fi: <u>Kotiäitinä</u> [stay-at-home mother] oleminen on aivan yhtä antoisaa kuin ansiotyössä käyminen. (WVS 2000(b), Q46D)

En: If no paid employment: ---

<u>Housewife</u> not otherwise employed (WVS 2000(a), V229)

Fi: -   <u>Kotiäiti</u> [stay-at-home mother], ei muuta työtä (WVS 2000(b), Q96)

How close to the word *homemaker* is the word *housewife* in style and in meaning?

The keyword **housewife** was found in all the six dictionaries studied. In two dictionaries *housewife* has other meanings than relevant in family role context, and additionally it was explained to also mean "a pocket-size container for small articles (as thread) (Merriam-Webster OnLine, see also Dictionary.com). It is also an adjective meaning "[t]o manage with skill and economy, as a housewife or other female manager; to economize" (Webster's Revised Unabridged Dictonary, 1996, 1998, via Dictionary.com). In all the dictionaries the main meaning was understood as a family role (see example 1 below), and thus also in a relevant way to the study. Its main characteristics were then 'woman', 'family member', and 'manages a household'. In four dictionaries it was also considered as a married woman and in one case "usually"

---

[51] Regarding the purpose and scope of the guide it is stated:

"This guidance supports UCL's commitment to equal opportunities and is for use by everyone who produces teaching and learning materials or any written material such as letters, memos, minutes and reports, in the course of their work. Language reflects the values of our society and its use can perpetuate prejudice and discrimination. It is important therefore that we use language that is inclusive and that we are sensitive to the risk of patronising, offending or excluding colleagues or students through our use of language. In the setting of examinations, it is particularly important to avoid any ethnic, gender, religious or inappropriate cultural bias in questions. This guide has been endorsed by the Academic Committee which hopes that its principles will be reflected in UCL communications, both internal and external." (UCL 2002)

such. In most cases housewives were not considered to have any other job, or at least this was the normal assumption (see example 2 below). In three cases, the studied keyword co-occurred with the word *homemaker*, which was represented as a synonymous word or as a near-synonymous word, particularly in style it was seen as a more up-to-date version (see example 3 below).

Example 1:

"housewife /haswaf/ n. -wives /wavz/ a woman who does not work outside of the home: She is a housewife who is very active in civic affairs. See: homemaker, USAGE NOTE."

(Newbury House Dictionary of American English 2002)

Example 2:

"a woman whose work is inside the home, doing the cleaning, cooking, etc., and who usually does not have any other job"

(Cambridge Advanced Learner's Dictionary 2002)

Example 3:

"Usage note:

1. Housewife is offensive to some, perhaps because of an implied contrast with career woman (just a housewife) and perhaps because it defines an occupation in terms of a woman's relation to a man. Homemaker is a common substitute."

(Based on the Random House Dictionary, Dictionary.com 2002)

Thus, *housewife* usually refers to a married woman, who takes care of her household instead of earning money from employment. How close is it to *homemakers* according to semantic characteristics represented in dictionary definitions and explanations? - Is caregiving as a role as common? According to the dictionaries the difference is small. – *Homemakers* were not necessarily mothers, and taking care of a household was much more central than taking care of children. *Housewives* were not understood as stay-at-home mothers – the care-giving aspect was completely lacking.

In MOT *housewife* has two meanings:

1 *kotirouva*, *kotiäiti* (homewife, homemother)

2 *ompelurasia*, *ompelu*[*tarvike*]*laatikko* (a small-size container for small needlework articles) (MOT 4.0 and 4.7)

But how does the dictionary keywords *housewife* versus *kotiäiti* (homemother) versus *kotirouva* (homewife) differ?

	woman	work at home	mother	married	unmarried
housewife	+	+	0	+	-
kotiäiti (stay-at-home mother)	+	+	+	0	0
kotirouva (homewife)	+	+	0	+	-

Table 10: *Housewife vs. kotiäiti vs. kotirouva; semantic characteristics matrix*

As shown in the above sample (table 10) the essential delimiting characteristics are 'motherhood' and 'marriage'. In practice, the English word *housewife* has a broader meaning than its direct Finnish translation (*kotirouva*) –it is used also when referring to stay-at-home mothers.

Thus, the overall discourse of dictionaries shows that as a word in common language usage *homemaker* is normally understood as a woman, who takes care of home and children instead of earning money from employment. In most of the dictionaries studied (as in the thesauri) *homemaker* is represented as an unambiguous word, although it evidently is not. It is still noteworthy, that the discourse of the studied dictionaries represents a kind of consensus and does not reflect the different practices in real language usage. In addition, the previously mentioned analysis of the overall discourse of large international surveys shows that *homemaker* is not considered a usable gender neutral word. Instead, more precise words referring to mothers who stay at home taking care of children, are used. Fathers are included in "others".

7.1.5 Equivalence in dictionaries

The dictionaries studied often provided formal word-for-word translations or lists of (near)synonymous words, without reference to the context being used. At best they provided brief encyclopedia like definitions for the keywords. Therefore, it is difficult to judge their nature in a coherent way compared with the other discourses studied in the following chapters 8 and 9. This is because examples of use are often lacking or very brief, the idea is clearly to provide information about semantic correspondences and not that much concerning the pragmatic correspondences. Nevertheless, the function of the explanations, word lists and/or provided equivalents stayed unclear for the most part (cf. Reiss & Vermeer 1986 and Koskinen 2001 in chapter 3.4.2.2 *Skopos-theory*) and it was

not apparent from whose perspective the information provided was given (cf. Nida 1964/2000; Venuti 1998; Vehmas-Lehto 1999).

The dictionaries do not provide information about the level of equivalence (cf. ISO 5964-1985), nor information about the selected translation strategies. When the dictionaries provide several equivalents, it is not clear whether it is a question of synonymy or near-synonymy as in exact or partial equivalence, or of single-to-multiple term equivalence (cf. ISO 5664-1985 in chapter 3.3.3.2 *Equivalence in multilingual thesaurus construction*).

Apparently, all the dictionaries studied strive for denotative equivalence and occasionally also for connotative equivalence, but to make a distinction between denotative versus connotative equivalence (cf. Koller 1989) is difficult. Moreover, text-normative, pragmatic and formal equivalence (cf. ibid.) is often impossible to judge. At the more general level it is rather obvious that the general strategy is foreignisation (cf. Lindfors 2001).

- **Summary and Conclusions**

The online dictionaries studied were gathered via link-pages of the Finnish university libraries and translation science departments and were thus considered to also represent the common linguistic tools for academic use. The generally known weaknesses of dictionary definitions - such as ambiguisity, circular definitions and lists of words - were still present also in this study (cf. Wierzbicka 1996, 240-241, TSK 1989, 41), and impaired the analysis. Additionally, it is noteworthy that the dictionaries studied were not coherent in their style and scope as regards the information provided.

According to the studied online dictionaries *family role* and *head of household* are not parts of the most commonly everyday language, whereas *breadwinner*, *homemaker* and *housewife* are common words with rather general and solid meanings. *Homemaker* is also regarded as a gender-neutral equivalent for the word *housewife*.

Dictionaries provided information about the denotative meaning of the keywords studied, whereas the connotative aspect was for the most part not visible (cf. Nida 1975). Some dictionaries provided examples, which also illustrated the pragmatic meaning, i.e. the style and context in which the words are used (cf. Reiss 1981/2000; Vehmas-Lehto 1999).

The online dictionaries studied – which were typical of their genre - do not provide enough information to evaluate the equivalence properly (cf. Nida 1964/2000; Reiss & Vermeer 1986; Venuti 1998; Vehmas-Lehto 1999; Koskinen 2001). It is apparent, though, that dictionaries aim more often at semantic equivalents and foreignisation strategies, than at e.g. pragmatic equivalence and domestication (cf. Koller 1989; Lindfors 2001).

7.2 Social scientists

In the next section, the same terminological case (as previously discussed) is explored as well as is how the problematics are understood by the potential users – that is, of the online thesauri studied, namely British and Finnish social scientists.

7.2.1 British social scientists

7.2.1.1 Background

A total of seven British social scientists were contacted by e-mail. Previous to this the potential respondents had been detected on the basis of searches and browsing in several databases (Finnish, British, American and international), and further selected by studying their research profiles via university home-pages and by viewing the potential informants' publications. Most of the potential respondents answered the first inquiry within two days; six agreed to participate in the study. One informant was going to be abroad during the time when the interviews were planned to be conducted.

The British social scientists interviewed represented four universities, located in three cities. Four of the respondents were women and all the informants were conducting research on post-doctoral level. The aim was to also include both sexes in the group, but the primary emphasis was on other criteria, such as, subject knowledge and experience. The respondents represented five different departments or institutions (and one respondent had a position in two): 1) sociology, 2) education, 3) economic geography, 4) gender studies and 5) social policy.

The interviews took place at the informants' place of work and were tape recorded. On the average the interview lasted one hour (at minimum 40 minutes and at maximum 85 minutes). As with the other groups studied, in comparison with the Finnish informants the difference in duration is due to the fact that the British informants were not asked for any translation work and the reasoning behind the translations. (See interview guide in appendix 2 and also chapters 7.2.1 and 7.2.2.)

Four out of six respondents were used to make their own searches in databases and all six were used to giving keywords to their articles, but it was often considered difficult.

> "*Do you keyword your own articles?*"

> "Only when I'm forced to --- I do it but with difficulties." (BR SS1)

Most of them did not use a thesaurus for either indexing or information seeking and they were not very familiar with thesauri designed for indexing and information retrieval.

"*Do you use any tools like these controlled vocabularies like thesauri when you give keywords?*"

"Not any specific --- most of the research is cross-disciplinary and that don't meet these categories." (BR SS1)

"Well, that never occurred to me. Do people do that?" (BR SS4)

"I think for me this is just a question of what are the key themes here. And they probably appear in the text, but I wouldn't do anything like which are the words that appear most often, because you can then, you can miss the important themes if you do that. So I don't think I'm very scientific about that [laughing]. So what seems to me to be the kind of the main themes and I think with the view to, well… what kind of areas would I … If people were exploring particular areas in what areas would I want them to find this piece." (BR SS5)

The information seeking and keywording practices influenced the interview situation, leading to less emphasis being placed on matters of descriptor equivalence in comparison to the original idea of the study.

7.2.1.2 Family roles as regarded by British social scientists

Among the six British social scientists interviewed the stimulus word *family* **roles** resulted in 45 word associations, out of which 37 were unique:
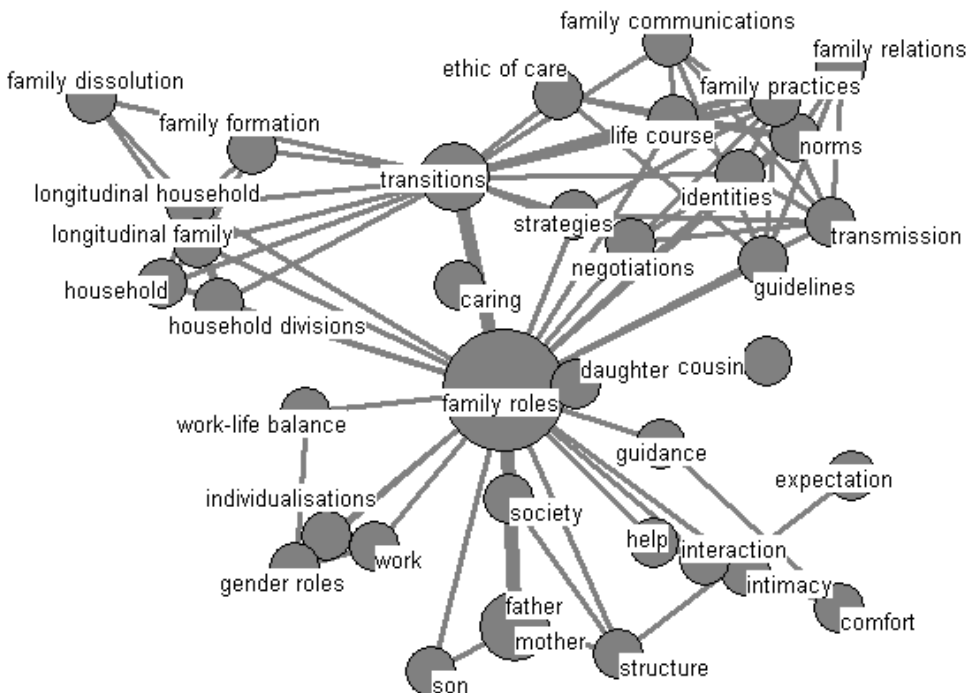
Figure 11: *Word associations for "family roles" by the British social scientists*

In this group, the word family roles connoted mainly as a social role and often together with words expressing transition (such as *transition, transmission*) and potential or clear conflict (such as *negotiation, family dissolution*).

The participants understood family roles usually in a more broad sense than represented in ELSST (examples 1 and 2 below), and also used to talk about them with broader concepts. Gender perspective was often considered as an organic part of this (examples 1 and 2 below). When family roles were discussed in the sense they are represented in ELLST, they were considered out-dated (example 2 below). In addition, the ideological reasons were often discussed (example 3 below, cf. Venuti 1998).

Example 1:

"With family roles I am principally looking at economic or closely related to economic roles, so I'm looking at work roles, for instance, but by work I don't just mean only paid work, but also unpaid work in the households… and in particular I'm interested in gender differentiations with such roles. And connected with those are activities that are on the margin of work, one of those margins might be childcare, at home for instance, another might be a shopping and consumption in general. --- And in each of these cases I'm interested in how those differentiate, not just by gender, but on three levels, by gender, by economic structure ---"

"And how would you divide them into sub-concepts?"

"I suppose domestic division of labour is going to be one set off, and a household consumption is going to be a second, and childcare would be a third and leisure would be a fourth." (BR SS1)

Example 2:

"*How would you divide them into sub-concepts? What kind of roles you would name?*"

"It is not easy, because things are changing. There used to be a male breadwinner, a male provider, and a female homemaker. Now this is not the case. The families are often dual-breadwinner families, which leaves the care-issue open. ---" (BR SS4)

(Later after asking about ELLST terms)

"They are old-fashioned and a bit, you know, kind of functionalist. In gender terms old-fashioned and out of date and don't reflect the reality. And functionalist for a society, which is depended on kind of oppressive system for women, really. Feels like - role theory feels like right now, you know, it was sociology kind of invented a kind of view of the family in way, which probably never existed." (BR SS4)

Example 3:

"There is a problem about when you put all these words in a thesaurus, explaining their use, their common use, it's difficult to do it in a way that is actually critical rather than reproducing the problems of it." (BR SS5)

Both *family* and *role* were constantly considered difficult concepts and terms, but the idea of roles was often criticised on a more principle level.

"I think role is a problematic concept and I try not to use it. --- One slips sometimes into the language of roles, but for me they are too prescriptive, they suggest too much, that they are gloves that you can put on, you know, garments that are waiting for

people to climb into, rather than different practices and identities, and the way one negotiates, different relationships in society. And, you know, there used to be old debate and conflict on roles, you know the domestic roles and work roles, but you know, if you have an identity of a working mother, you're moving between different contexts and you put your practices and the kind of relations that you're engaged in, are much more in an interactive process between you as an agent, as an individual, negotiating with various constrains and constructions and relationships. I don't talk about family roles as such. I find them as rather static, old-fashioned concepts, really. And I would find a way round it using the other concepts, really." (BR SS4)

As with the other keywords used, the British informants were not asked for translations. Most of them brought up cultural differences, and some also linguistic differences between different cultures. The most influencing factor considered was that in different cultures, families consist of different generations – e.g. respondent BRSS1 explained: "in Italy there is three generation, whereas in Finland only two". Respondent BR SS3 further clarified: "and in some languages it is important to make a distinction whether relatives are from the mothers or the fathers side".

7.2.1.3 Breadwinners and Heads of household as regarded by British social scientists

The British social scientists gave 53 associations in all, out of which 40 were unique, to the stimulus word **breadwinner**:
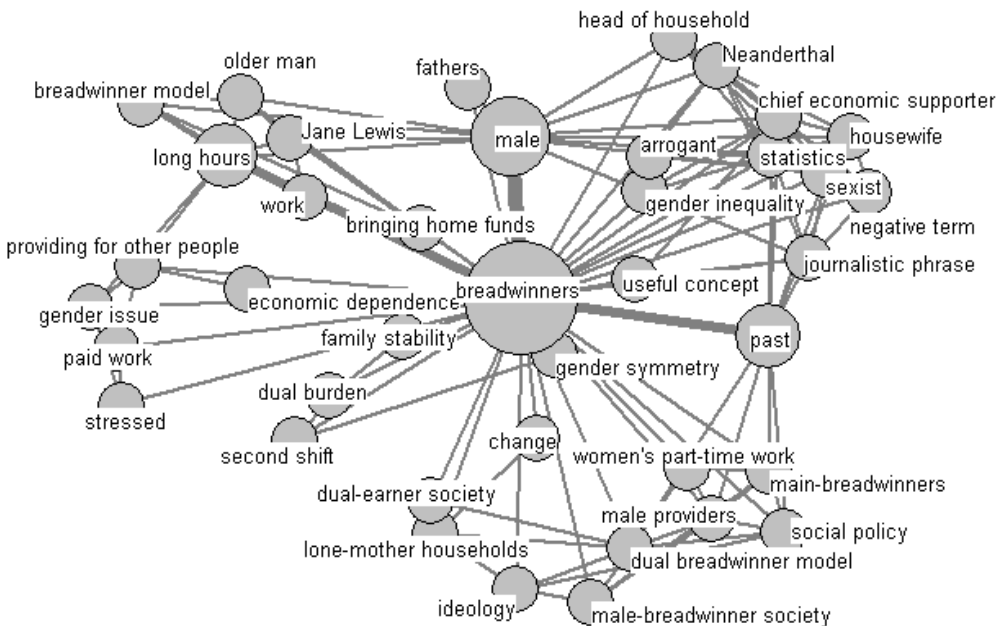
Figure 12: *Word associations for "breadwinners" by the British social scientists*

Breadwinner was commonly understood to refer to men, to represent the past (examples 1 and 2 below), and also connected negatively to long working hours. The term was not necessarily considered useless, but its content was considered to have changed. The semantic meaning was understood similarly and considered to have stayed relatively much the same, but the pragmatic meaning was considered to be in flux (cf. Koller 1989).

Example 1:

"That is a journalistic phrase, I suppose. That has never been used in… I suppose in most of my work we would use something like chief economic supporter[52]. So it's a useful concept, which as with housewife reflected very clear distinction with the past, because there was clearly one breadwinner, there was one housewife, and that was the way family structure was organised, clearly that's no longer the case. "(BR SS2)

Example 2:

"But I mean in a social policy sense you can say we are moving from a male-breadwinner society, in which men were deemed to be the main breadwinner, to a society in which we have a dual-worker household. And indeed we have very many lone-mother households, where the women increasingly are the main breadwinners." (BR SS4)

Among the British socials scientists, the associations for the **heads of household** were more heterogeneous than for *breadwinner*. The stimulus word resulted in altogether 59 associations, out of which 49 unique.

---

[52] Cf. Glossary of CeLSIUS (Centre for Longitudinal Study Information and User Support), which defines "Chief economic supporter" as:

"(CES) In the 1971 census, a household member aged 15 or over who is the head of household, or is related to the head, and who is the highest ranked in the household according to the following criteria, in this order:

 * Economic activity - those in full-time employment ranked highest, followed by those out of employment, part-time, retired, and others.

 * Family position - married ranked highest, followed by widowed or divorced in a family, and others

 * Sex - male ranked above female

 * Age - older ranked above younger" [2005, last visited 8.7.2009, URL: <http://www.celsius.lshtm.ac.uk/glossary/glossary.php?selectphrase=Chief+economic+supporter>]
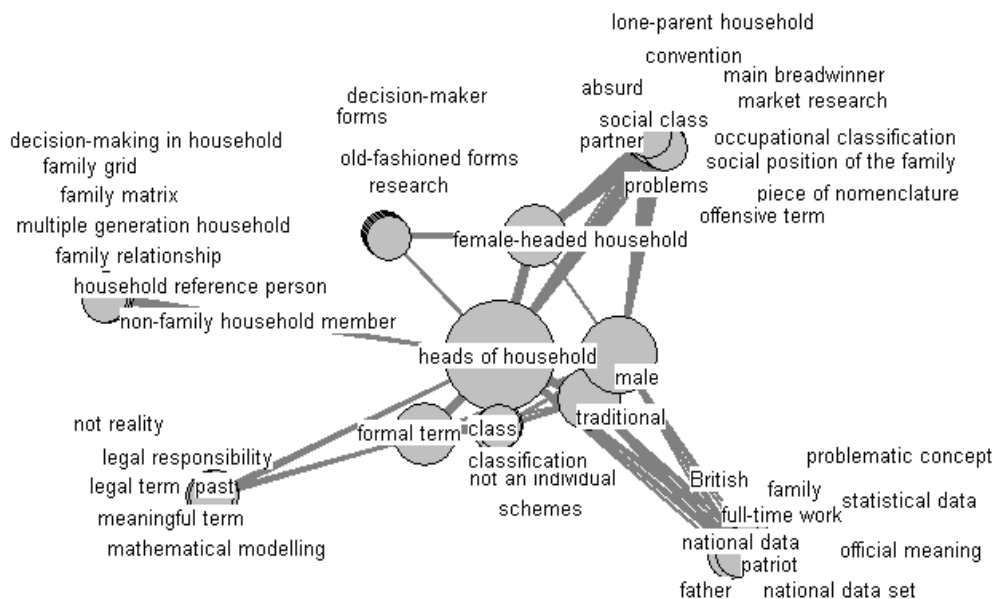
Figure 13: *Word associations for "heads of household" by the British social scientists*

Heads of household was thus strongly associated as a male-related and as a formal or traditional word. Moreover, issues related to gender division were commonly brought up.

On the other hand the word is considered a statistical term and as such more accepted, but on the other hand more sexist and as thus even more negatively regarded than *breadwinner* (see example below). The semantic meaning was understood similarly among the informants, and the differences were thus more on the pragmatic level. Respondent BR SS5 said:

> "Hah! [laughing] Yes, well. I once threw somebody out of the house, a researcher, market researcher, because he wanted – he asked my partner's occupation, my partner wasn't there. And I said you don't need to know that, and he said I have to have the occupation of the head of the household. And I said I'm the head of household. And he said, no no, if you're cohabiting, you know, we have to put the man down here as the head of the household. And I said excuse me, I own this house, I'm the main breadwinner, I'm the head of household. If you are not prepared to call me as the head of the household you can go now. He said, you know, it's not sexist. I said oh yes it is. [laughing]" (BR SS5)

7.2.1.4 Homemakers and Housewives as regarded by British social scientists

How did the British social scientists consider the concept and term ***homemaker***? First they were asked to give about five word associations (response words) for the stimulus word *homemaker* and later to define the concept and evaluate its usability as a descriptor.

The picture below illustrates what associations the word *homemaker* gave:
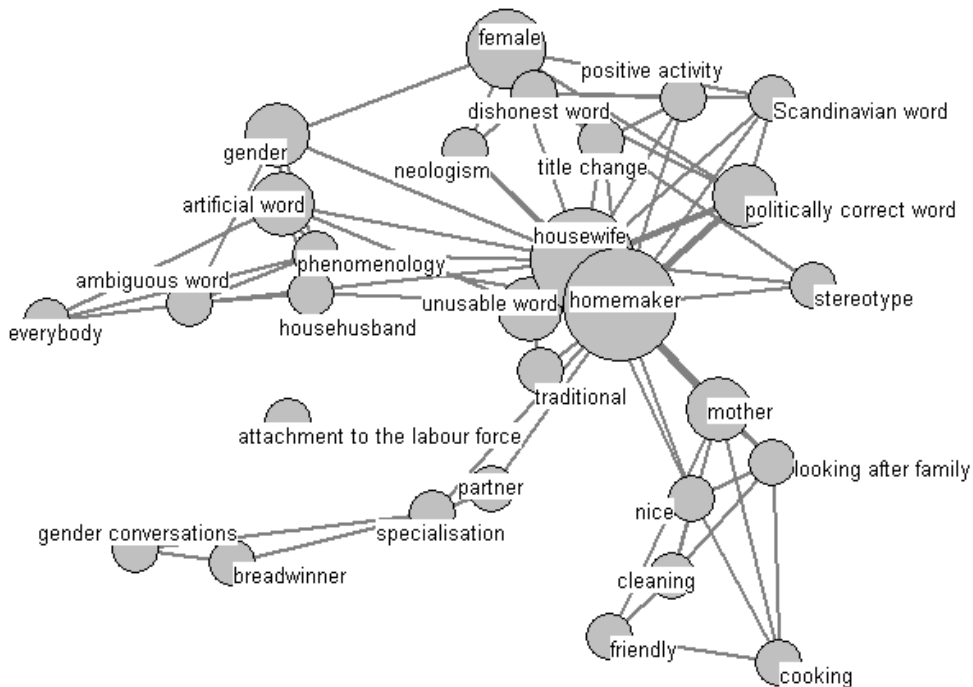
Figure 14: *Word associations for "homemaker" by the British social scientists*

The stimulus word *homemaker* resulted in 56 response words, out of which 29 were unique. Three clear clusters were found: issues related to 1) usability of word (such as *politically correct term, dishonest word*), 2) activities (such as *looking after family, cooking*) and to 3) gender (such as *househusband, female*). It associated commonly as an unusable, ambiguous and artificial word. As an activity it was associated usually positively, so the negativeness occurred mostly on a term level, not on a concept level. In practice *homemaker* was considered to refer to a person, who is female and looking after home and children. *Homemaker* was thus commonly understood to be in practice a synonym for *housewife*, but *housewife* produced still more negative associations both as a word and as a concept (see the next figure below). Either of these words was considered to be usable as a descriptor and the interviewees would rather avoid using both of them.

The stimulus word ***housewives*** resulted in 56 response words, out of which 41 were unique:
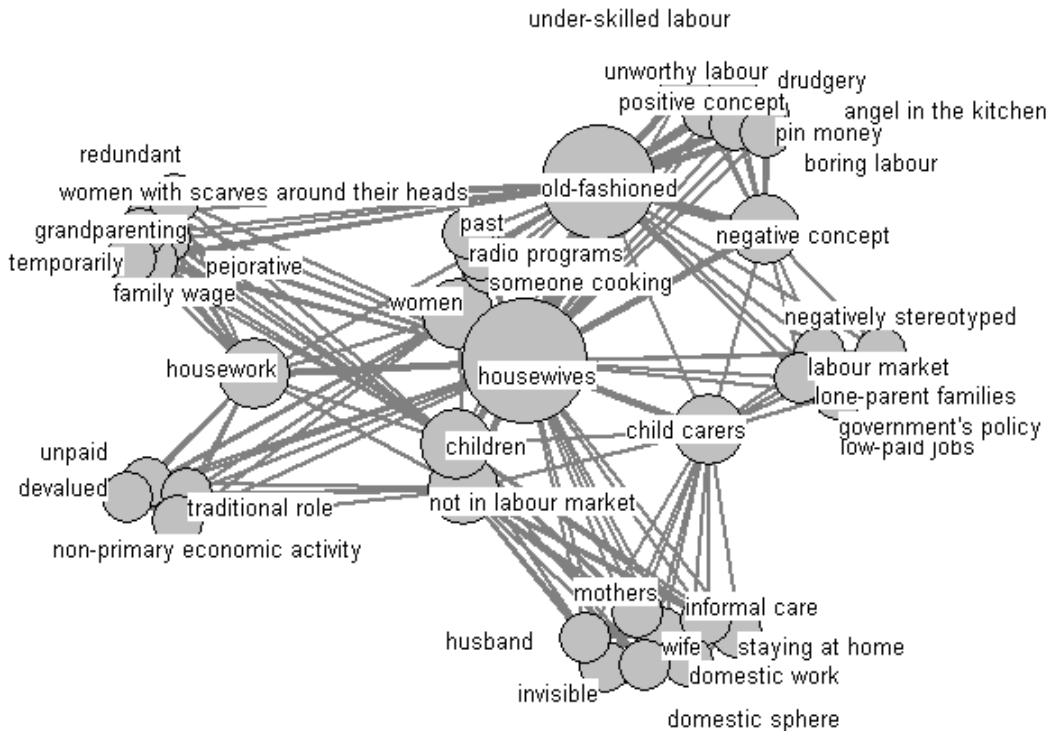
Figure 15: *Word associations for "housewives" by the British social scientists*

Housewives was constantly considered as an old-fashioned word, and often to represent a negative concept, although acting as a housewife could also be seen as a positive action (cf. pragmatic versus connotative meaning and Koller 1989 in chapter 3.4.1.2). As a common language word it was still seen as being very British, whereas its near-equivalents *homemaker* and *stay-at-home mother* were often said to be American words.

The overall discourse of British social scientists shows that on the denotative level *homemaker* and *housewife* is understood similarly as in the dictionaries, thesauri and databases. However, there is a significant difference on the connotative level and due to that difference the discourses might not meet in a possible information seeking situation.

7.2.1.5 Equivalence by British social scientists

As reported previously, the British participants were not asked to translate from English to any other language. The equivalence issues were discussed together with other issues – e.g. would the family roles terminology be similar in Scandinavian and/or European countries at the concept level. Most informants referred to literature they read in English, but which is translated from other languages, and/or referred to their experience in multinational research projects.

The informants within this study group considered there to exist differences mainly for two reasons: in Scandinavian countries the dual-breadwinner model is more common than in the UK and the role of children is stronger (example 1 below). The dominance of English in international co-operation was also often discussed (example 2 below), and also regarded as harmful for English-speaking societies (example 3 below, cf. Katan 2004).

Example 1:

[*In the context of ELSST family roles content*] "Childrens rights are far greater issue in Scandinavic countries than in UK." (BR TS2)

Example 2:

"Family and household mean different things in different countries." (BR SS1)

Example 3:

" - - - So there are gaps between the conceptual meaning and the epistemology I suppose, with the common sense everyday language in the cultural and national context. So there are many more layers the problems to deal with in a cross-national research. And together with, you know, with the dominance of English language that everybody has to work with because it has become increasingly the common language, everybody has different expertise in English, and English has its limitations, you know, it doesn't necessarily reflect all the cultural diversity of meanings that you were having in other languages." (BR SS4)

Some participants also talked about translation strategies and – although not directly saying it – critised the idea of internationalisation (cf. chapter 3.4.1.3 *Domestication or foreignisation – or internationalisation*; Venuti 1995).

"--- I think it is a problem in European social research, which is a mess due the dominance with English language, in a sense that with the same word must be used to refer to different things and it gives the folks idea that things are becoming more similar, whereas the actual fact is there are important differences between different European states. --- This means that English-speaking speakers use a poorer English, as the other languages are becoming narrower, the English language is also becoming narrower, because people, instead of using five different words to describe the same… You know you may have five slightly different words, which mean similar things, but slightly different … But you use one, because that one could be understood… So I think it is quite dangerous to lots of words… ---" (BR SS3)

Equivalence and translation strategy issues were not only problems to be considered at the cultural or linguistic level, but also at the discourse level (cf. Carmen 2004; Nida & Reyburn 1981). The aims were considered important to bear in mind (cf. Skopos-theory, in Reiss & Vermeer 1989 and example 1 below), and it was also recognised that the familiarity and similarity could be greater among individuals representing specific discourse than language (example 2 below). The informants also made it clear from whose perspective they were speaking (example 3 and 4 below, cf. Venuti 1998; Nida 1964/2000; Vehmas-Lehto 1999), and it was much more than just geographic-cultural approaches.

Example 1:

"It is a matter of principle, to aim at gender-neutral use. Gender-specific terms mirrors often the reality, but it is politically correct to recognise, that it is possible to men to stay at home as well as to women." (BR SS3)

Example 2:

"I think perhaps there is a kind of bigger difficulty about it, which is that when you combine a thesaurus you take a term and then you try kind of attach the definition to the term and if possible to say what would be the equivalent term in a different language. But actually because, because terms don't, most terms don't have fixed meanings and then in a sense the kind of connotations depend on the context. But because you necessarily take them out of the context you loose all those sort of subtleties of meaning, that are there when it's kind of embedded in its discourse or is different discourses. So that would be, you know, that's a sort of problem about translation, isn't it. That unless you can translate the whole kind of discourse and cultural framing you are bound to loose something of the meaning of the word." (BR SS5) --- "--- So there are those difficulties. I mean the problem for me there is that I don't actually read any other language, so I only ever read things in translations and I'm aware of that, you know, one has to be very… You can't do the kind of discourse analysis on a translation, that you can do on a document that was already in English, because you have to be aware that the sort of connotations of the original were probably rather different. So in a sense of what I'm saying is, you know, I'm aware that there is a problem and my linguistic skills are not such that… I'm kind of engaged in it very much, because the kind of work I do is so much about looking at discourses and you can't actually, you know, I couldn't actually do that in other language." (BR SS5)

Example 3:

"My feeling is that the same discourses, the same variety of discourses, the conceptual discourses exist in both in Scandinavia and in Britain. So I would say, I mean I would find very easily people, who would think like me in Scandinavia, in fact because they have a history of a strong welfare state and a history of active citizenship, that ideas and concepts are in a sense more elaborated or they are more self-conscious about some of these things and so I would find … I think I would say I would find very much more sympathetic and easy to communicate with people from Scandinavian countries. I can't say that compared to Britain, because Britain is huge social science community, and actually, you know, you only know rather small part of it. Of course you know there are people who I can communicate and discuss with the same conceptual level in Britain as well, of course. But I don't think the North sees such a barrier really, not on a country. And I think we have learned a lot from Scandinavia, because certainly in a welfare state research and in a sociology of childhood for example, where your policies and ideas and the public discourse has been so much more… so much … I can't say in advance, because that sounds like some kind of national progress, but so much more interesting and sympathetic." (BR SS3)

Example 4:

"What kinds of sub-concepts would you give, how would you divide these family roles?- If you think of British society?"

"Well, again... I'm always speaking as a sort of quantitative social scientist. I mean in terms of family I would tend to use a sort of hierarchy, which would be --- (BR SS2)

The British social scientists considered translation and equivalence issues to represent two-way problems between target and source languages and cultures. The idea of dynamic equivalence was clearly prior to formal translations (cf. Nida 1964/2000; Vehmas-Lehto 1999), and the idea of internationalisation (cf. chapter 3.4.1.3 *Domestication or foreignisation – or internationalisation*) was criticised. According to the majority the translations should reflect variety of differences as well as at the geo-cultural as at the discourse level. The connotative level of the words was considered important, since it influences the way users react to the words and whether they accept to use the given descriptors or not (cf. Ulrych 1992).

- **Summary and Conclusions**

For the six British social scientists interviewed the studied terms represented quite a narrow way of understanding family roles. Time and gender issues were often discussed together with them (cf. Suojanen 1993; Engelberg 1993 and 1998), as well as the ideological reasons were often discussed (cf. Venuti 1998). The time aspect, gender issues and ideological motivations were considered as factors considerably influencing the interpretation of the terms studied.

In general, the word *family roles* was considered to refer to tasks and roles within a family or household. *Breadwinners* and *households* were considered as masculine related words, although they do not express the sex clearly. They were considered as formal and partly out-dated. *Homemakers* and *housewives* were commonly considered as synonymous expressions. In all the cases the connotative aspects were also discussed and they were considered significant (cf. Ulrych 1992).

The British social scientists considered translation and equivalence issues both from the perspective of the target and the source languages and cultures. Dynamic equivalence was usually aimed at and the focus was on the receiver and the usability of the translation (cf. Nida 1964/2000; Vehmas-Lehto 1999). The idea of internationalisation (cf. Venuti (1995) was criticised.

7.2.2 Finnish Social Scientists

7.2.2.1 Background

The Finnish social scientists were identified via searches and browsing with family role terminology in the Finnish, British, American and international databases and further selected via studying information concerning each individual on university web-pages and via their publications. Altogether eleven Finnish social scientists representing

four universities were contacted, out of which seven were female and four male. Most of them answered to the first inquiry and within two days. One informant had just travelled abroad for a year, two informants said they were too busy to participate in the research interview and two did not answer at all. No more inquiries were sent, as the six social scientists were considered to be sufficient.

The six Finnish social scientists interviewed represented three universities, located in three cities. Five of the respondents were women. The gender division was not considered a problem, since the emphasis in this study was not on gender differences among discourses and also while collecting material it became more obvious, that gender did not explain the differences in focus. Most importantly, the respondents represented four different departments: 1) sociology, 2) social policy, 3) sociology and social psychology and 4) gender studies. Most of the informants were conducting their research at a post-doctoral level (two were licentiates) at the time of the interview.

The interviews took place at the informants' work places and were tape recorded. On the average they lasted 95 minutes (a minimum 45 minutes and a maximum 150 minutes). The Finnish respondents were also asked further questions about translations into Finnish and consequently these interviews lasted longer. (See the interview guide in appendix 1.)

Three out of six interviewees were used to making their own searches in databases and all six were used to giving keywords to their articles, but similar to the British respondents these were not considered an easy task. It was also not common to think from the information seekers viewpoint, and the keywords selected could even be very unique or a neologism, in order to "profile the research" (FI SS4). None were very familiar with thesauri designed for indexing and information retrieval. They were all aware of YSA, but not used to using it. The general Finnish thesaurus YSA was considered either too general or just foreign. In the case of family role matters, they often preferred rather general terms in information seeking, such as motherhood, fatherhood, parenting, gender and family, and their combination with the Boolean operator AND.

> "What kinds of terms would you use in information seeking about family role matters in the Finnish databases?"

> "Well, family, single parent, nuclear family, dual-parent family, re-married family. With these kinds of different types of family concepts." (FI SS2)

> (After showing how the family role terminology was presented in ELSST)

> "I would consider these more as phenomena, I would seek as them, not as population categories, but… Or I think that would be more natural for me, since my research interests are more like researching rules and phenomena." (FI SS6)

It is still noteworthy that both the Finnish and the British social scientists tended to favour other information seeking strategies than keyword searches, and new relevant documents were most often found via journal circulation at the home institution, colleagues, conference proceedings and/or references in an already known article etc.

7.2.2.2 Family roles as considered by Finnish social scientists

When discussing the studied word ***family roles*** the Finnish equivalent "*perheroolit*" was used. First how the participants consider the theme was asked, after which what *family roles* (*perheroolit*) are about and what kinds of roles the participants would name was asked. The first introduction to the English terms studied was word association test, in which the participants were asked to give about five response words to the certain stimuli words. After that definitions and translations were asked and discussed. – The progress was the same with the other terms studied.
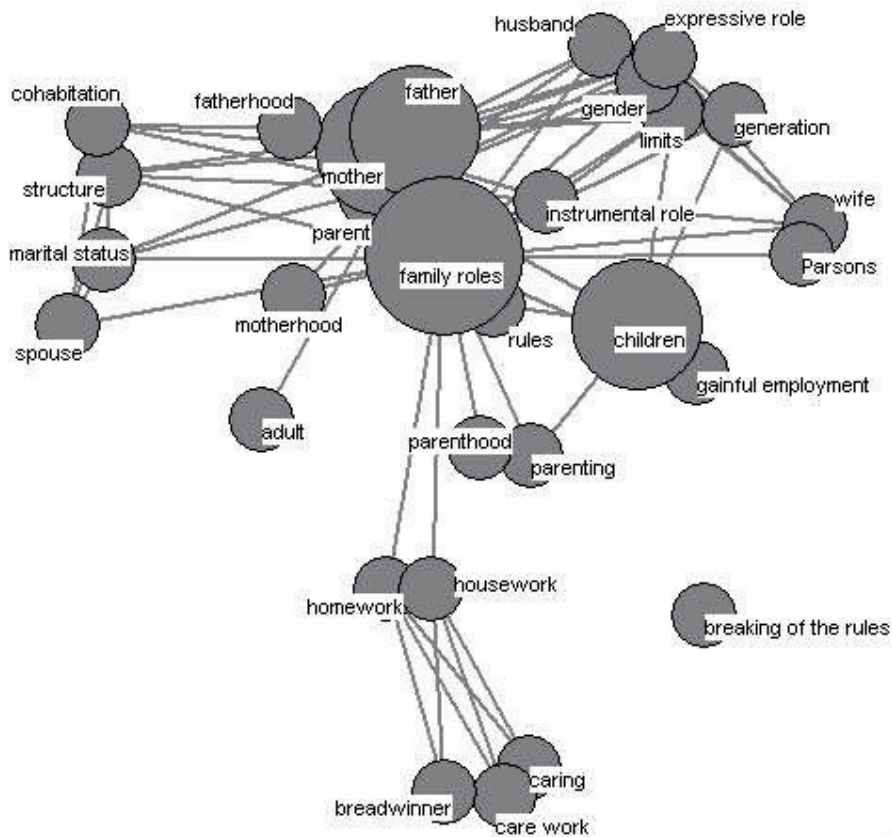
Figure 16: *Word associations for "family roles" by the Finnish social scientists (transl. author's)*

The Finnish social scientists gave altogether 46 response words, out of which 30 were unique. The sample was thus slightly more consistent than the British one.

The associations of the Finnish social scientists are clearly related to family members and family was understood as a nuclear family with parents and children. As a research object it associates more with terms of statistics. *Family roles* is clearly a foreign concept and also connotes more strongly as an old-fashioned concept (even as

"Briticisms"). It is not understood as part of a modern common language concept in Finnish, although some of the informants mentioned, that it is coming back, but in a different way from the 1950's Talcott Parsons functionalist theory (women with an instrumental role, men with an expressional role). Children and parents are emphasised, but in a more gender neutral way than e.g. in the Finnish database LINDA sample (cf. later in chapter 8).

In the case of *family roles* the informants emphasised the pragmatic and the connotative meaning in a rather similar and equal way to the more denotative meaning (cf. Koller 1989), and thus as a descriptor *family roles* was not commonly accepted. - The idea of roles was considered too rigid, and nowadays in the information environment even misleading (cf. Venuti 1998). The dynamic nature of the studied terminological case has also impact on knowledge management (cf. Forsman 2005), since, for example, family refers today to a broader variety of phenomena and processes, than it used to do some decades ago.

> "But it is in actually with all of these --- [terms], that they are specifically roles in the meaning of gender roles. They create different tasks, duties and responsibilities for men and women. And these are, at least from the Finnish perspective, from the most rude perspective, which is strange for us, which makes us disturbed. But I don't know how it is with all the other, in which is an assumption of a specific kind of problematic gender-neutrality…" (FI SS1)

The participants were used to giving criteria or functions according to which to divide and name different roles instead of actually spelling out the roles. The themes mentioned were parenthood (*vanhemmuus*), parenting (*kasvattaminen*), grandparenthood (*isovanhemmuus*), caregiving (*hoivaaminen*), participation in working life (*työelämään kuuluminen*), sisterhood (*sisaruus*), sexuality (*seksuaalisuus*), financial relationships (*taloudelliset suhteet*), relatives/kinsfolk system (*sukujärjestelmä*), adulthood (*aikuisuus*), childhood (*lapsuus*) and marital status (*siviilisääty*).

The roles named were parents (*vanhemmat*), children (*lapset*), grandparents (*isovanhemmat*), stay-at-home mothers (*kotiäidit*), stay-at-home fathers (*koti-isät*), working mothers (*työssäkäyvät äidit*), working fathers (*työssäkäyvät isät*), person who takes care of child's welfare (*lapsen hyvinvoinnista huolehtija*), family provider (*perheen elättäjä*), father's role (*isän rooli*), mother's role (*äidin rooli*), child's role (*lapsen rooli*), and "in a broader context" also grandparent's role (*isovanhemman rooli*).

The Finnish translation would constantly be *perheroolit* [direct, literal translation, familyroles], but it was not generally considered appropriate to use this word. - The Finnish social scientists would often prefer to talk about tasks and responsibilities within a family and they emphasised that each person in a family has several roles and for example the caregiver and supporter role is often shared. The concept of family roles was thus understood differently from the thesaurus sample, and the meanings were only partly overlapping.

> "*How would you translate family roles into Finnish*?"

"Perherooli [familyrole]. But if I would really translate, I would like to use another word, and if I would refer to roles, I would talk about duties of family members ["perheenjäsenten tehtävistä]. (FI SS3)

Thus, the concept 'family' has already become very problematic. Roles are considered to guide thoughts towards something old-fashioned. - It was popular in 1950's and now it seems to be coming back, but in a different way from the past. Translation would be "perheroolit", but its relevance and meaning as referring to roles is often questioned.

7.2.2.3 Breadwinners and Heads of household as considered by Finnish social scientists

For the stimulus word **breadwinner** the Finnish social scientists gave a total of 40 response words, out of which 28 were unique. Words that co-occurred (coc) more than once with "breadwinners" were male (4 coc, *mies*), supporter of the family (3 coc, *elättäjä*), dual breadwinner family (2 coc, *kahden elättäjän perhe*) and traditional model (2 coc, *perinteinen malli*).
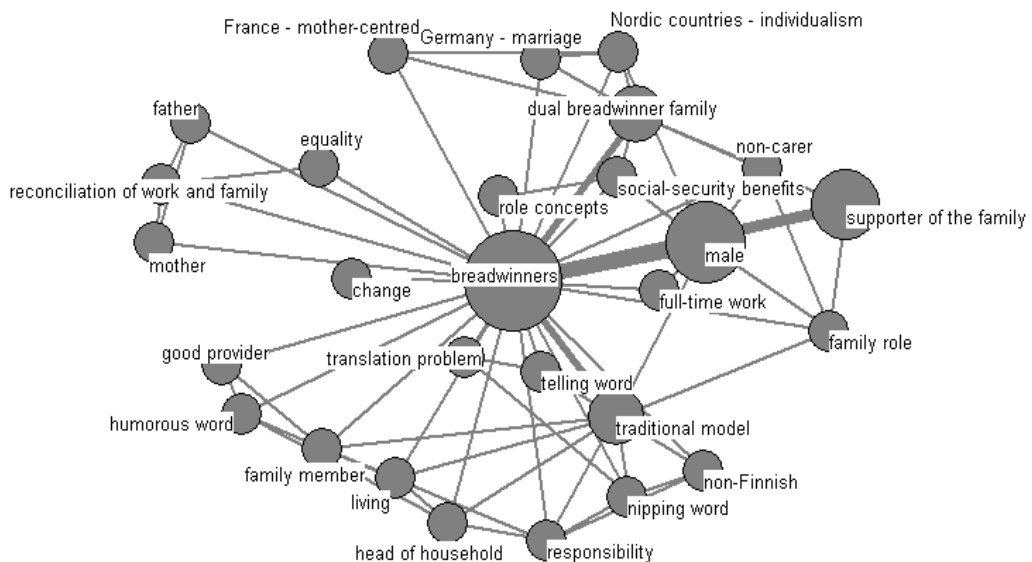
Figure 17: *Word associations for "breadwinners" by the Finnish social scientists (transl. author's)*

Similar to the case of family roles, the connotative and pragmatic aspects were clearly raised by the participants. The semantic aspects were discussed, but the pragmatic aspects were much more emphasised.

"Well, it is connected to this kind of traditional division of roles, in which men go to work and women are stay-at-home mothers, so in the conception is one and only

person taking care of economic living. "Good provider"... It is the one that is
sometimes used with a smile on your face." (FI SS5)

Four of the six Finnish social scientists interviewed would translate *breadwinner*
with the meaning of provider (*elättäjä, perheenelättäjä*). Two would translate it as
employed, (*työssä käyvä*). All of them considered it to refer, first and foremost, to men
and that even working single mothers would rather be named as employed than as
breadwinners.

For the Finnish social scientists interviewed the studied word **heads of household**
was associated with 43 different words with a total of 53 occurrences:

Figure 18: *Word associations for "heads of household" by the Finnish social scientists
(transl. author's)*

The two most common associations were male (*mies*) and head of the family
(*perheenpää*). Clearly this term denotes as a statistical term and as a common language
word it has connotations of being a foreign and negative concept.

Head of household was translated by all the informants as *perheenpää* (head of the
family), although it was also noted, that it is question of households, which are not
necessarily families. The informants thus made a difference between semantic and
pragmatic meaning, but preferred the form, which is well-established and a common
term in statistics, and prioritised pragmatic aspects to semantic ones (cf. Koller 1989;
Vehmas-Lehto 1999).

7.2.2.4 Homemakers and Housewives as considered by Finnish social scientists

How did the Finnish social scientists considered the concept and term ***homemaker***? As with the other terms studied, first they were asked to give about five word associations (response words) for the English stimulus word *homemaker* and later to define the concept, translate it into Finnish and to evaluate its usability as a descriptor.

Among the Finnish participants the other stimulus words seldom and only randomly resulted in response words in English, but *homemaker* also provided, in half of the cases, response words in English. - Two of the informants gave some response words in English and some in Finnish, one informant gave solely in English and three solely in Finnish.

The picture below illustrates the response words given for the stimulus word *homemaker* (here also the Finnish ones are translated into English by the author):
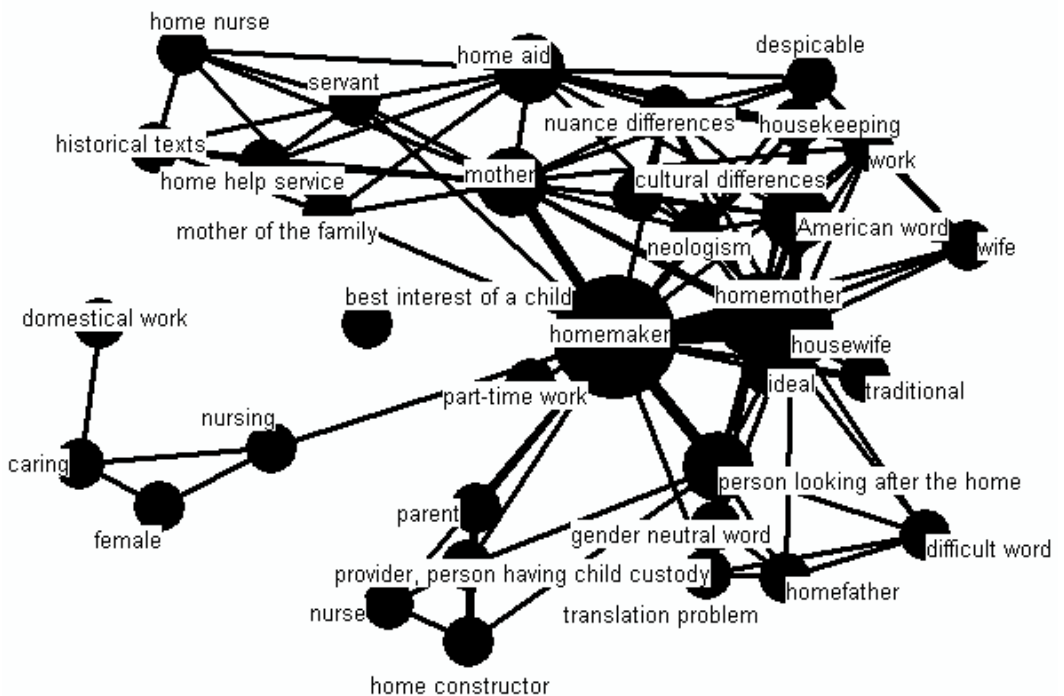
Figure 19: *Word associations for "homemakers" by the Finnish social scientists (transl. author's)*

As in the British sample, the associations occurred on two levels. *Homemaker* associated as a foreign word and as difficult to translate. It was usually not a negative concept, but the context was often considered to be out of date. As an activity it associated mostly as being a stay-at-home mother, but also as a paid caregiver or a home aid.

Although the word associated often as a mother (female), it was in principle considered as a gender neutral word. The concept was considered difficult to translate into Finnish due to two characteristics: 1) gender neutrality and 2) home making ('take care of home', not necessarily look after children as a mother or father).

"It is difficult. It has always been difficult for Finns." (FI SS5)

Two of the informants considered the pragmatic Finnish equivalent to be a descriptive phrase ("*kotona työskentelevä/lapsia kotona hoitava vanhempi*" – a person working at home/a parent taking care of children at home), one a term "*kotiäiti*" (stay-at-home mother), one "*kotirouva*" (housewife), one "*kodinhoitaja*" (housekeeper, as an occupational title) and one did not give any kind of equivalent.

Among the Finnish social scientists the stimulus word ***housewives*** resulted in 38 different response words, with totally 47 occurrences:
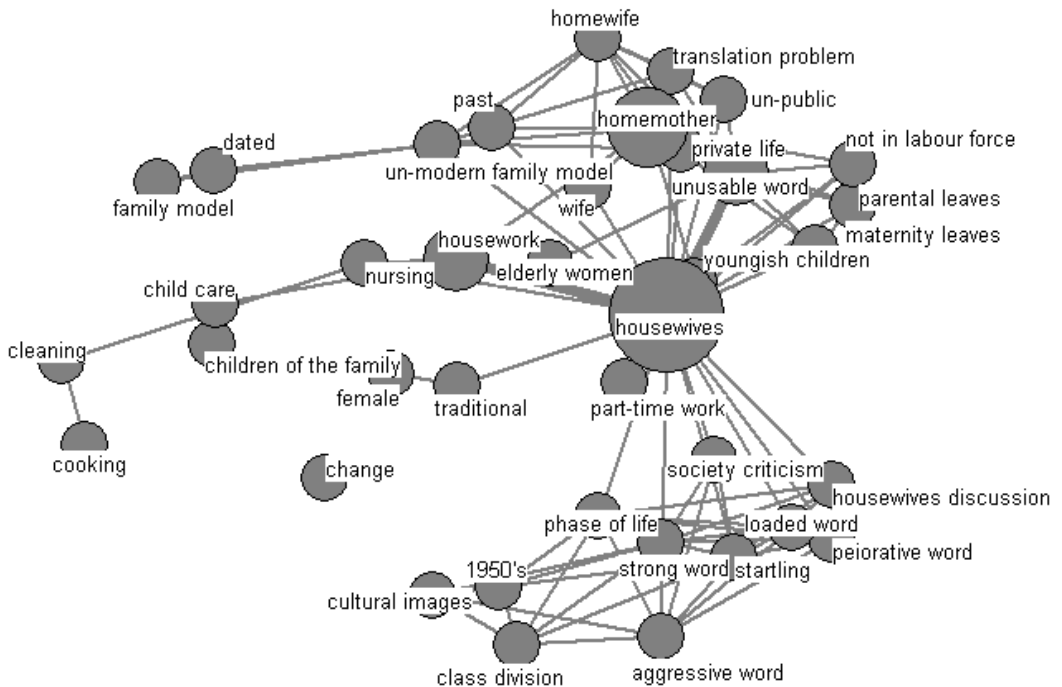
Figure 20: *Word associations for "homemakers" by the Finnish social scientists (transl. author's)*

The words often associated were stay-at-home mothers (*kotiäidit*), homework (*kotityö*) and unusable word (I would not use it, *en käyttäisi*).

Half of the informants would translate it as *kotirouva* (homewife, ~ a wife staying at home instead of earning money from employment), and half as *kotiäiti* (stay-at-home mother). It was generally considered an unusable word with negative connotations. The Finnish equivalent's denotative meaning was thus often narrower than its English source term's scope, and the connotative meaning clearly differed. The pragmatic meaning is thus also different in Finnish than in English. (Cf. Koller 1989)

The overall discourse of the Finnish social scientists interviewed shows *homemaker* to be a foreign, old-fashioned and ambiguous word in a Finnish context. It was not so much in consistency with other discourses studied as the British one was. *Housewives* was considered to be out-dated and to connotate negatively, and its modern equivalent would be stay-at-home mother.

7.2.2.5 Equivalence by Finnish social scientists

The way *family roles* were represented in the ELSST material was considered foreign, out-dated and/or too narrowly understood, and it caused problems in translation process at the concept, term and descriptor level. Ideological reasons also influenced the way the Finnish social scientists interviewed wanted to use and translate the terms (cf. Venuti 1998). Sometimes prior to translating certain content was to translate in a certain style, i.e. in this case the aim was to follow the values or gender-neutrality in the target culture (cf. also Vehmas-Lehto 1999). Respondent FI SS3 described the relevance, usability and meaning of the studied terms:

"Would you think of seeking information with these terms?"

"Well no. These are foreign the way, that these base on divisioned roles and on a certain kind of family model, where the roles are divided particularly so, that woman is like inside the family and man somehow on the edge of the family, as a protector or provider, or so that he anyhow has his relations outside the family, and as a head of family would represent the family to outside of the family. In the Nordic countries we have very strongly an alternative model, we talk about shared parenthood, it is like "dual breadwinner", and we talk about two career families. These terms do not exactly meet… But what is essential is that in families both adults, who are most often man and woman, but they could be in principle also lesbian or gay couple, so both of them have a career or task in working life, and then they share the household duties and care, childcare I mean. So this is based completely to the asymmetric family model, which has existed sometimes in Finland at least at the level of ideal [laugh], but which cannot really be present. Of course we could have stay-at-home mothers and housewives like as according to studies as periodic identity, that we have somebody at home with home care subsidity, but it is for three years. Of course if three children are born, so it can last nine years. But these people plan to return to a dual breadwinner model. And we say that men would participate in domestic responsibilities and care, so that is at least the aim, although it clicks the most. The worst is that women have moved towards being providers, whereas men have not become more caring. But it is at least the ideal that is wanted to be used to conceptualise the matters and following that principle I would also seek for information. I don't think I would find very often a person in Finland, who is just taking care of household. So this is probably right in Germany and… I don't know about France, but in Germany and in South-European countries… It has a lot to do with Catholicism, so the role of religion steps also in here, but it is less studied. But in social policy there is a discussion about welfare regimes, how Europe is divided into welfare regimes, that is into models, there is great difference in this relation… So is the Nordic perspective involved here? --- And feministic movement and women opinion leaders in these Central-European countries

would be interested in other kind of definition of family, so why do these stay so much behind…" (FI SS3)

The participants favoured more dynamic equivalence than formal equivalence (Nida 1964/2000; Vehmas-Lehto 1999), and when considering the different translation alternatives the pragmatic meaning (cf. Koller 1989) was often considered. If exact equivalence was not possible, they tended to favour a narrower equivalence relationship. Thus in terms of degree, it was clearly more common to choose a narrower equivalent relationship, than a broader one, which is not consistent with thesaurus standards (cf. ISO 5964). Non-equivalence was also possible.

The more complex the terminological case, the greater the variety was in aimed equivalence degree. For example, with the case of the most problematic term, *homemaker*, which was generally considered a cultural word (Newmark 1998; Venuti 1995) among the Finnish as well as among the British participants, the semantic and denotative meanings were emphasised, and some strove for a more literal translation, and some aimed at dynamic equivalence. - The pragmatic meaning would be stay-at-home mother (*kotiäiti*), but due to operative reasons (aiming at gender neutrality and guidance of thinking and language use) it was not easy to translate, and also resulted in descriptive phrases as translation equivalents. Although the reasonings for the translation equivalents (cf. Reiss & Vermeer 1989) were discussed, the aimed for equivalence was not always stable. E.g. *breadwinner* was also considered to be a foreign concept, its meaning was understood more precisely and constantly, and it resulted in more similar equivalence types and degrees than *homemaker*. The most simple term to translate into Finnish was *head of household*, which was considered as a statistical term and as such it was similar to a universal (Newmark 1998), whereas the other terms were seen to reflect more qualitative research traditions and reflect the idea of the nuclear family.

The time factor was discussed a great deal (cf. Aitchison 1991; Suojanen 1993; Hirsjärvi and Hurme 2000), and it also influenced the translation strategy selected. In the case of *homemaker* and *housewife*, for example, half of the informants chose to reflect practices common in the past, and half of the participants preferred to adjust the translation to mirror current practices. The versions reflecting the past were more semantic translations, and the versions reflecting current practices aimed at mirroring pragmatic meanings. When considering thesaurus construction and use, it is thus important that thesauri mirror several kinds of time aspects, although only one would be selected as a preferred descriptor.

> "*Homemaker*" is American... It is probably this kind of housewife, which is considered as the most terrible type of people [laughing] and which is connoted with all negative things. I don't know exactly, but "*homemaker*" is more like… The association is that in Finnish we have housewife (*kotirouva*), which is similarly out-dated, it is related to wives, and stay-at-home mother has replaced it. So in this sense *homemaker* is equivalent to Finnish *stay-at-home mother*. But it [homemaker] includes so much more. I think it's important to keep in mind that there is this side of wife and mother, but also housekeeping, so that is wife, mother and housekeeper. And that it is work. And "*homemaker*" emphasises it, whereas *stay-at-home mother* concentrates on motherhood and on the relation to children. And *stay-at-home mother* excludes the

aspect of wife, and it did not come to mind, and it is of course one side of this. And in addition it also excludes pretty effectively the work that is done at home, or if it is considered from the perspective of work it is taking care of children. So in that sense there exists clear cultural and tone differences… And for me it was at some point a neologism, --- and now it is "homemother". (FI SS1)

"I would translate "homemaker" as housekeeper ("*kodinhoitaja*"), as a profession. --- And "housewife" as stay-at-home mother ("*kotiäiti*"), although it does not translate that directly..." (FI SS3)

We can thus assume that the aimed equivalence type and degree are influenced by the values of the target culture, which are related to the concept to be translated. In their translations the informants took into consideration several perspectives (cf. Koller 1989). Usually, they prioritised denotative and pragmatic equivalence, and together with pragmatic equivalence were also aiming for connotative equivalence. Text-normative aspects were considered very little, and never prioritised. Formal equivalence was sometimes discussed, but not aimed at.

- **Summary and Conclusions**

For the Finnish informants the concept *family* was already very problematic. *Roles* as a word was considered to guide opinion to something old-fashioned. – According to the participants it was popular in the 1950's, and now it seems to be coming back, but in a different sense than in the past. The translation was constantly "perheroolit", but its relevance and meaning referring to roles was often criticised, and thus there existed a clear gap between the denotative and pragmatic meaning (cf. Koller 1989).

In the case of *breadwinner* the semantic aspects were more frequently discussed, but as in the previous case *family roles* the pragmatic aspects received more emphasis. Four of the six Finnish social scientist interviewed would translate *breadwinner* with the meaning of provider ("elättäjä", "perheenelättäjä") and two would translate it as employed ("työssä käyvä"). All of them considered it to be a masculine-related word and not to be used when referring to women (cf. Engelberg 1993 and 1998).

Head of household denoted a statistical term and as a common language word it connoted as a foreign and negative concept. It was translated by all the informants as "perheenpää" (head of the family), although it was also noted that it is a question concerning the definition of households. The informants made a difference between semantic and pragmatic meanings, but preferred the form, which is well-established and common in statistical use, and prioritised pragmatic aspects over semantic ones (cf. Koller 1989; Vehmas-Lehto 1999).

Homemaker was associated as a clearly foreign term. Its connotation was considered to be negative, although its semantic and pragmatic meanings were considered rather positively (cf. Koller 1989; Vehmas-Lehto 1999). It was usually not a negative concept, but the context was often considered out of date. As an activity it associated mostly with being a stay-at-home mother, but also a paid caregiver or a home

aid. In principle it was seen as a gender-neural term, but in practice as female-related (cf. Engelberg 1993 and 1998). It was considered difficult to translate, and it produced the greatest variation in Finnish equivalents given. - Two of the informants considered the pragmatic Finnish equivalent to be a descriptive phrase ("*kotona työskentelevä/lapsia kotona hoitava vanhempi*" – a person working at home/a parent taking care of children at home), one the term "*kotiäiti*" (stay-at-home mother), one "*kotirouva*" (housewife), one "*kodinhoitaja*" (housekeeper, as an occupational title) and one did not give any kind of equivalent.

In the case of *housewives* there existed a division between two perspectives: Half of the participants would translate it as *kotirouva* (homewife, ~ a wife staying at home instead of earning money from employment), and half as *kotiäiti* (stay-at-home mother). It was generally considered an unusable word with a negative connotation. The Finnish equivalent's denotative meaning was thus often narrower than its English source term's scope, and the connotative meaning clearly differed. The pragmatic meaning is also different in Finnish than in English. (Cf. Koller 1989.)

The overall discourse of the Finnish social scientists interviewed shows *homemaker* to be a foreign, old-fashioned and ambiguous word in a Finnish context. It was not so much in consistency with other discourses studied as the British one was. *Housewives* was considered an out-dated word and to connotate negatively, and its modern equivalent would be stay-at-home mother.

We can assume that both the content of the concept and the values of the target culture related to the concept, influence the aimed equivalence as to type and degree. - The more complex the case, the greater the variety became in regards to the aimed for degree and type of equivalence. In their translations the Finnish social scientists took into consideration several perspectives (cf. Koller 1989). Usually they prioritised denotative and pragmatic equivalence, and together with pragmatic equivalence they were also trying to attain a connotative equivalence. Text-normative aspects were hardly considered, and never prioritised. Formal equivalence was sometimes discussed, but not aimed at.

The participants followed a rather consistent domestication strategy when trying to find relevant equivalents from the perspective of Finnish practices (cf. Venuti 1998; Lindfors 2001). They did not provide equivalents according to the existential equivalence (cf. Koskinen 2000), but followed the ideas of the skopos-theory (cf. Reiss & Vermeer 1986; Vermeer 1989; Koskinen 2001 in chapter 3.4.2.2 *Skopos-theory*) and thus considered the function of the translation they provided.

7.3 Similarities and differences in common and academic language use

Six English language online dictionaries and one Finnish-English online dictionary, all provided by Universities for academic use, were explored in order to study how they understood and represented the specific family role terms. It transpired that dictionary definitions are poor when considering the pragmatic meanings of the studied terms. The

participants interviewed tended to provide much more information about the connotative aspects, style and gender-relatedness of the studied terms. *Breadwinner*, for example, was commonly considered to be a male-related word, whereas only two of the studied dictionaries indicated (with dynamic variations) it to be a male-related word, and paradoxically, one a feminine-related word.

The twelve social scientists interviewed were equally British and Finnish. Their research areas represent different aspects of family roles (such as reconciliation of work and family, being stay-at-home father, socio-economic differences of family members in different countries etc).

In the social science discourses studied, the meanings are more precise, although not necessarily more narrow than in the common language discourse of the dictionaries studied.

In the terminological case of family roles the Finnish participants similarly and equally emphasised the pragmatic and the connotative meaning more than the semantic and denotative meaning, and thus – although aiming at communicative equivalence (cf. Newmark 1989) the studied terms were not commonly accepted as descriptors. This is partly due to the thematic case, but also with the barrier between (documentary) thesaurus and (natural) common and scientific language.

In the context of family roles terminology both the Finnish and the British interviewees mentioned many words that referred to family relationships and family members, such as parenting and children, but the British informants spoke more frequently about the words of social roles and emphasised the conflict aspect.

The ELSST family roles terms studied were generally considered to represent economic roles and statistics, but the British informants talked more about economic roles (providers etc), extended families and kinships than the Finnish informants. Both the Finnish and British respondents stated that the studied terminological case was lacking especially from the perspective of children – Finnish discussed more about this, and the British participants assumed the Finnish participants to do so. The case was also considered to represent the idea of the nuclear family and the practices of quantitative analysis too much. In general, the terminology was more acceptable to British informants as words, terms and descriptors, although the concepts were often regarded as problematic.

The researchers' qualitative versus quantitative research orientation influenced the selected translation strategy – the terms were more acceptable for respondents using quantitative material and the translations more semantic for the respondents using qualitative material. - It is noteworthy, that due to its origin and use of context ELSST reflects mostly quantitative material and aims to provide a tool for indexing and retrieving of this. The informants representing more qualitative research orientation especially tended to favour more general terms (also in indexing and in information seeking) such as motherhood, fatherhood, parenthood, children, family, home and work. Both the Finnish and the British informants thought, that it is better to use politically correct terms (cf. Venuti 1998) and thus, for example, to avoid genderism, although they

thought that it is in a way artificial – in real language usage, for instance, *homemaker* refers to women (cf. Engelberg 1993 and 1998).

Pragmatic and communicative equivalence was often prior to any semantic translations and meanings (cf. Koller 1989). In general the British interviewees were against internationalisation of terminology and some even clearly prioritised domestication in the sense that they considered the rich English terminology which is full of nuances to essential and preferable to a more general and easily translatable terminology. The Finnish informants followed rather consistently domestication strategies when trying to find equivalents relevant from the perspective of Finnish practices (cf. Venuti 1998; Lindfors 2001). They did not provide equivalents according to the existential equivalence (cf. Koskinen 2000), but followed the ideas of the skopos-theory (cf. Reiss & Vermeer 1986; Vermeer 1989; Koskinen 2001 in chapter 3.4.2.2 *Skopos-theory*) and thus considered the function of the translation they provided.

8 Equivalence and family roles in indexing practices

This chapter focuses on indexing term discourse and on equivalence and translation strategies from the perspective of indexers. The material that is used comes from datasets that have been retrieved from two Finnish and one British databases, interviews with Finnish and British indexers (tot. 6) and simulated indexing task situations (5 articles).

The British and Finnish indexers represent different social science libraries or information centres (cf. chapter 6.3.5) and were interviewed (similarly to other groups) in order to find out how the studied terms are understood and defined/translated in different cultures and subcultures, and in order to study how the given equivalents differ from each other. For the purposes of gaining knowledge on what is pragmatic equivalence in multicultural thesaurus construction, equivalence and the aims of translations were also studied from the perspective of indexers.

8.1 Databases

The databases used for analysing the studied terms as title words and descriptors are (in discussion order) the Finnish databases ARTO and LINDA, and the British database COPAC. The **general aim is to find out, how these specific terms are found and understood in the databases and what is the indexing term context**. The general **research questions** in this part of the study have been operationalised as:

- How are the terms studied used in the databases? What is their frequency? To what degree are the terms studied used in indexing today? Which are the most co-used descriptors?

- Do the terms studied represent a documentary language rather than a natural language in the studied samples?

- What is the semantic lexical network of the studied terms? What are the studied terms about according to the indexing?

- Are the semantically gender-neutral terms neutral in indexing practice?

- To what extent do the indexing terms of the samples represent a transliteration problem versus a potential translation problem?

In the analysis of the database material discourse analysis, co-word analysis and a bibliometric tool box, Bibexcel, are used.

In this process the terms that occurred most frequently are listed first, and then their co-occurrence is counted. With the help of the MDS-technique (multi-dimensional-scaling) the co-occurrences can be represented as a map, where the distance between terms is the opposite of the relation to the number of terms which occurred

simultaneously. – The more the terms co-occur together the closer they are placed in the map. Because the frequency of a single word strongly influences co-occurrences, the effect of word frequencies have been eliminated via normalising the distance unit (the observed frequency divided by that of the expected frequency). (Persson 1991, 56. More about co-word analysis in chapter 6.2.3.)

The terms used in the databases are mainly *family roles* terms from ELSST (European Language Social Science Thesaurus) and their Finnish equivalents (English/Finnish: *family roles/perheroolit, breadwinners/perheenelättäjät, heads of household/perheenpäät, homemakers/kotitaloustyötä tekevät, housewives/kotiäidit+koti-isät*).

8.1.1 Finnish databases

8.1.1.1 The Finnish database ARTO

ARTO is a reference database for Finnish articles (an index of articles from approx. 1000 Finnish journals). Altogether about 40 institutions have been indexing the ARTO database. (Helsinki University Library 2005) A common indexing tool used is The General Finnish Thesaurus YSA, but special vocabularies are also used when needed. The use of special vocabularies is organisation-depended.

The search in ARTO covers descriptors (subject terms) and titles. The depth of any further analyses is dependent on the sample. **The general aim is to find out, how these specific terms are found and understood in the ARTO database and what is their indexing term context**. The emphasis is thus mostly at the term level, while a deeper level of analysis would be too great a task in this context.

First, the English terms studied are discussed, then their Finnish equivalents. In the study of Finnish practices, the terminological starting point is the Finnish equivalents provided in the working version of ELLST (European Language Social Science Thesaurus). Some Finnish near-equivalents are also evaluated from the perspective of semantics.

The searches were conducted on the ARTO database at the beginning of the study (2002-03) and updated in October 2004 (with only minor changes).

The material showed that pre-coordinated indexing was rare, and when it occurred the descriptors were handled (as was also found later in LINDA and COPAC databases) as separate. Thus e.g. the two given descriptors "*kotityöt*/homework – *miehet*/men, *koti-isät*/stay-at-home fathers" were analysed as three: homework, men, stay-at-home fathers (- the original Finnish terms are written here in italics), and the analysis occured in those cases more on the term level than on an indexing term level. Possible record duplicates (different editions etc.) were also eliminated and typing errors corrected.

8.1.1.1.1 Family roles in ARTO

The search in the ARTO database with the string **famil? AND role?** in the descriptor or title field resulted in 36 records (cf. in 17.6.2003 34 records), published in 1972-2003. The 25 records from the years 1990-2003 were further analysed.

Only eight of the 25 most recent records were considered relevant, i.e. they referred to *family roles* as a single concept and in the way it is commonly understood in thesauri (see chapter 9).[53] These eight were all published between 1991-2003.[54] Six of them were articles in books published in English, one a periodical article in English and one a periodical article published in Finnish with an English summary. Three of the records do not have any descriptors, the other five have altogether 31 in Finnish (YSA-descriptors) and three have, in addition, English equivalents for the Finnish descriptors and one has the Finnish descriptors also partly in Swedish[55]. The **final sample** is thus only five records concerning family roles which include descriptors.

In three cases, the searched concept was found in the titles and in four cases as a descriptor. The concept was used rather evenly by the indexers and authors. The used descriptors of the sample are:

NO	Finnish (original)	English
3	arkielämä	everyday life
3	elämäntapa	way of life
3	perhe	family
3	roolit	roles
3	tutkimus	research
3	Viro	Estonia
1	naiset	women
1	naistutkimus	women's studies / gender studies
1	perheet	families
1	perhetausta	family background

[53] What were the other most recently retrieved records like? Twelve of them were not at all relevant: in each of them the search string did not appear as a concept (e.g. in a title "*Role of sensory and cognitive information in the enchancement of certainty and liking for novel and familiar foods*"). In four of the retrieved records the search string did not appear as a concept, but the records were partly relevant: these also concerned gender roles and family members but were not about them (e.g. in record titled as "*The joke as an instrument of power? : sexual relations in jokes told by family members*).

[54] One in 1991, three in 1994, one in 1996, two in 1997 and one in 2003.

[55] The Finnish descriptors were eight, Swedish only three.

1	sairastavuus	morbidity
1	sosiaalinen tausta	social background
1	sukupuoliroolit	gender roles
1	Suomi	Finland
1	terveys	health
1	työelämä	working life
1	työnjako	division of labor
1	venäläiset	Russians
1	yhteiskuntaluokat	social classes

Table 11: *ARTO family roles –sample (5 records, 1990-2003, descriptors)*

Two of the descriptors used referred to women, none to men. Genderism was thus not a clear phenomenon, and neither was grandparenthood, as in the COPAC database (cf. chapter 8.1.2).

The illustration of the indexing term network of the five records, which were considered relevant and had descriptors[56], shows how *family roles* are understood in the Finnish context as part of everyday life and in geographical contexts close to Finland (cf. descriptors *Finland*, *Russians*, *Estonia*):

---

[56] In all the five records precoordinated indexing was used, but in the analysed sample the precoordinated descriptors were separated and handled as individuals.
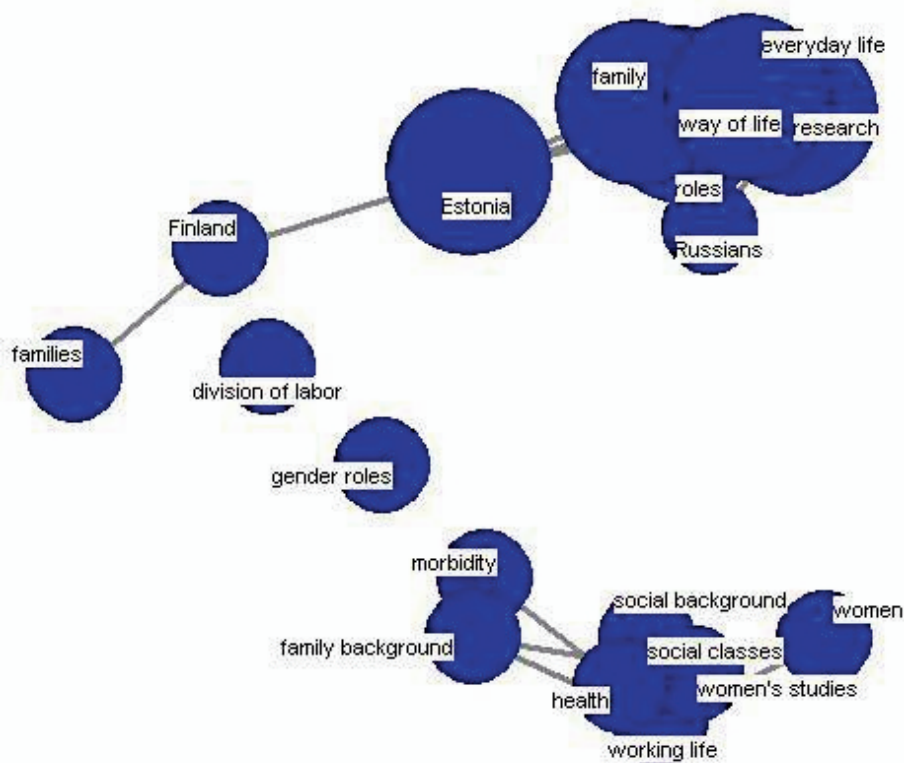
Figure 21: *ARTO family roles -sample (5 records, 1990-2003)*

(Three of the five records studied also had equivalent descriptors in English, and the Finnish descriptors of the two others were translated by the author.)

NOTE: the same records are included in the Finnish sample (perhe? AND rooli?): four records had equivalent descriptors also in Finnish (*perhe – roolit*) and in one record an equivalent was represented in the title ("Sosiaaliluokka ja perheroolit naisten terveyden tutkimuksessa" / Social class and family roles in the research of women's health).

The English concept *family roles* is thus used both by Finnish authors of English articles and by Finnish indexers. The usage is not common, but it is still used in the same way as it is represented in commonly used thesauri. Due to YSA practices the indexers refer to it with separate terms (*family – roles*).

The search (in 12.10.2004), with the ELSST Finnish equivalent *perherool?* in the title or descriptor field (TI/DE) resulted in three records published in 1992, 1991 and 1974. They all had the studied concept in the title ('*family roles*'), but the aspect was described very differently in the indexing.

The search (12.10.2004) **perhe? AND rool?** in TI/DE resulted in 43 records (in June 2002 42 records) from the time period 1960-2003. 25 records from the years 1990-2003 were further analysed. Two of the retrieved records referred to the same

publication, published in two different forms (periodical articles in paper and electronic formats). Except the publication form of the publication they were otherwise similar and are therefore counted in the further analysis as one record.

The **final sample** covers 20 records, which emerged as relevant: they have the studied concept either in the title and/or as one of the descriptors. They were published in 1991-2003.[57] Four of the records referred to articles in books, one to a book, and fifteen to periodical articles.

16 of the 20 records studied referred to documents written in Finnish and according to the indexing four of them had a summary in English. Three of the studied records referred to documents written in English and one in Swedish. Thus, in this sample the studied concept occurred mostly as an original Finnish concept.

Altogether the 20 records, which were considered relevant ones, have 158 descriptors, that is on average 7.5 descriptors per record. The variation is remarkable: the minimum is two and the maximum 52. There were 99 different individual descriptors. The descriptors used more than once were:

---

[57] One in 1991, one in 1993, three in 1994, two in 1995, two in 1996, two in 1999, three in 2000, two in 2001, three in 2002 and one in 2003.

NO	Finnish descriptor (original)	English
17	roolit	roles
11	perhe	family
5	perheet	families
5	perhe-elämä	family life
4	vanhemmuus	parenthood
3	arkielämä	everyday life
3	elämäntapa	way of life
3	isät	fathers
3	miehet	males
3	sukupolvet	generations
3	tutkimus	research
2	äidit	mothers
2	isyys	paternity
2	perheterapia	family theraphy
2	sukupuoli	gender/sex
2	uusperheet	stepfamilies/re-married families
2	vanhemmat	parents
2	kotikasvatus[58]	parenting
2	Viro	Estonia
2	työ	work

Table 12: *ARTO perhe (family) and rooli (role) –sample (20 records, 1990-2003, most used descriptors*

Eight descriptors referred to time (each occurred once, e.g. *1800-luku*/ the 19th century, *myöhäiskeskiaika*/ The Late Middle Ages), one to a geographical name (*Viro*/Estonia, occurred once). Five descriptors referred to research, material or methods (with altogether 7 occurrences, e.g. *tutkimus*/research, *diskurssianalyysi*/discourse analysis). Six descriptors were masculine related (with altogether 11 occurrences, e.g. isät, miehet) and eight were female related (with altogether 9 occurrences, e.g. *äidit*/mothers, *äitiys*/motherhood).

---

[58] Depending on the context kasvatus and kotikasvatus was translated either as parenting, education, or home-education.

For the co-word analysis the plural form *perheet* was replaced with the more common form, singular form *perhe* (4 replaces). In one record the both forms were used as descriptors within the same record, and the plural form was eliminated.
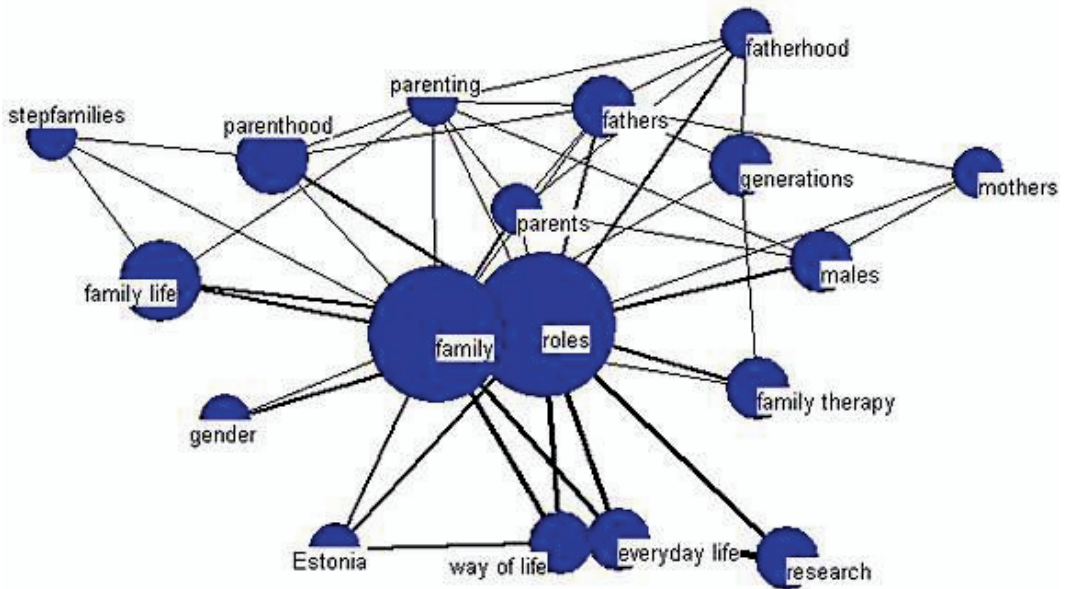
Figure 22: *Family (perhe?) and Roles (rool?) in DE/TI-search, 20 records published 1990-2003, co-occurrences of the descriptors with two or more occurrences*

In most cases (~3/4), the searched strings occurred in the descriptor field. Although the searched terms were not common, they still resulted in relevant material about parenthood and family roles as a research object. A family was understood to consist of two generations, but is not necessarily a nuclear family.

Children were only implicitly visible in the results, although e.g. in the discussion and studies of the interviewed Finnish social scientists the child-aspect was strongly – and explicitly – emphasised (cf. Chapter 7). The term *family roles* is not so strongly connected to working life and gender roles as in the COPAC database (cf. later in subchapter 8.1.2).

8.1.1.1.2 Breadwinners and Heads of household in ARTO

The search (12.10.2004) **breadwinner? OR "bread winner?"** resulted in four records. In two cases the term was used as a metaphor and referred to a company. In two

cases it referred to a concept in the way it is understood in ELSST. The concept was not represented in the descriptor field, but occurred solely as a title word. Three of the four retrieved records referred to a document published in English, one to a document including both Finnish and English versions. The search with the Finnish equivalents **perheenelättäj? OR "perheen elättäj?"** resulted in no records (13.10.04). The term *breadwinner* was thus in the Finnish ARTO database clearly a foreign concept.

The success with its near equivalent *heads of household* was rather similar. The search "**head? of household?" OR "household head**?" resulted on 13.10.2004 in no records.

The search (13.10.2004) with the Finnish equivalent **perheenpä? OR "perheen pä?"** in titles or as descriptors resulted in five records, published in 1992-2002. All the records referred to periodical articles, written in Finnish, and had the search string in the title field.

In two of the records the search string appeared in the sense commonly understood in thesauri. The records were titled as "Naisesta on tullut perheenpää ja vahvempi sukupuoli" (Women has become the head of the family and the stronger sex) and "Sansibarin nainen ottamansa omaansa : mies on yhä perheen pää, mutta naiset ahkeroivat itselleen itsenäisempää asemaa" (The woman of Sansibar is taking her own : man is still head of the family, but women are striving for a more independent position). In one case, the term *perheenpää* (head of household) was used metaphorically to refer to husbands: "Kiivaat perheenpäät, nalkuttavat äkäpussit : väkivalta uuden ajan alun englantilaisessa avioliitossa" (Hot-tempered heads of families, nagging cranks : violence in English marriage in the beginning of the new era). In one case, the search string occurred in a title in a different compound word than studied, but the connotation and thus metaphoric usage can be the same – "Perheen pääministerit" (The prime ministers of families). In one case the string did not refer to the studied concept, but occurred in a title "Yhden perheen päivän jätteet : pakkausten karnevaali" (One family's one day garbage : the carnival of the packages). The Finnish ELSST equivalent is thus not understood nor used similarly in ARTO as its English origin usually is in the other discourses studied.

8.1.1.1.3 Homemakers and Housewives in ARTO

The search (13.10.04) **homemaker? OR "home maker?"** resulted in five references, published in 1984-1997[59]. In all the records the search string occurred in the title field. Three of them were written in English and two in Finnish with an abstract in English.

In the older cases the word *homemaker* occurred in a compound word *communal homemakers* and *municipial homemakers* and referred thus to an occupational title (in

---

[59] PY= 1997, 1996, 1994, 1991, 1984.

Finnish *kunnallinen kodinhoitaja*). In three of the more recent cases the word was used to refer to housewives and/or to homemothers as the individuals who decide about the nutritional issues in families. The studied concept had no equivalents in the descriptor field.

No references were retrieved (13.10.2004) with the Finnish ELSST equivalent **"kotitaloustyötä tekev?"**. It is thus not used as a descriptor or in titles.

The search (13.10.04) with the strings **housewife? OR housewives?** resulted in only one record (1980).

A natural near-equivalent for both *housewives* and *homemakers* in the Finnish context is stay-at-home mothers (*kotiäidit*, cf. especially chapter 7) and also stay-at-home fathers (*koti-isät*, cf. esp. 7.2.2.4). They are used in ELSST as a Finnish translation for *housewives* (*kotiäidit+koti-isät*). Their use in ARTO was also evaluated in order to analyse how close the context of their pragmatic indexing term is to those of the English variants studied.

The TI/DE-search **homemother? OR "stay-at-home-mother?"** (+ other spelling variations) resulted in no references (13.10.04). The Finnish equivalent search (13.10.2004) DE/TI kotiäi? resulted in 46 references, published between 1978-2006. The **sample** further analysed consists of 29 records and presents the years 1990-2003. (For the co-word analysis the plural form *perheet* was replaced with the more common form, singular form *perhe*, one replace.)

12 records were retrieved solely on the basis of descriptors (*kotiäidit*), 11 on the basis of both title word and descriptor, 2 solely based on title words and in addition 4 due to title word and descriptor, but the string appeared in a title as a part of a compound word (e.g. *kotiäitivuosien*/of years of homemotherhood, *uuskotiäitiys*/neo-homemotherhood, *kotiäitiys*/homemotherhood). **Linguistically** the sample was very Finnish – 23 published in Finnish and 6 in Swedish.

The sample consists of 59 descriptors with altogether 114 occurrences (on average 4 descriptors per record). The most used descriptors within the sample (with two or more occurrences) are:

NO	Finnish (original)	English
22	kotiäidit	stay-at-home mothers
8	lastenhoito	childcare
6	kotihoito	home care
5	äidit	mothers
4	perhe	family
4	naisen asema	women's status
3	äitiys	motherhood
3	naiset	females
3	arki	everyday life
2	kotihoidontuki	home care subsidy
2	jaksaminen	endurance (well-being)
2	hyvinvointi	welfare
2	kotitaloustyö	domestic work
2	ajankäyttö	time utilisation
2	kotityö	home work
2	hoitovapaa	parental leave

Table 13: *ARTO kotiäidit –sample (29 records, 1990-2003)*

When including all the descriptors for the co-word analysis, the richness of topics is clear:

Figure 23: *ARTO kotiäidit –sample (1990-2003, all descriptors)*

Limiting the analysis to the descriptors with two or more occurrences the main topics become more observable and clear:

Figure 24: *ARTO kotiäidit –sample (1990-2003, descriptors with more than 1 occurrence)*

Stay-at-home mothers are here discussed together with issues related to childcare and home work (with descriptors such as childcare, home care, domestic work, home work, time utilisation, everyday life), motherhood and women's status (motherhood, women's status, womanhood, gender roles) and mothers' well-being (mental health, burn-out, welfare). Issues of labour force were more minor.

In the searches (14.10.2004) the English equivalents **homefather** OR **home father** OR **stay at home father** (+ different spelling variations) resulted in no records.

The search with the Finnish equivalent **koti-isät** (13.10.2004, DE: koti-isät OR TI: koti-is?) resulted in 11 references, published in 1994-2002. Four were considered irrelevant: in two of these cases the search string appeared as two separate descriptors and did not refer to one concept (*koti, isät – home, fathers*[60]) and in two cases it was in different title words and concepts than those studied (*koti-ISDN - home ISDN, --koti, isänmaa-- - --home, fatherland---*).

In the 7 cases that were **studied further**, the studied word occurred as a descriptor and/or in title: in three cases solely as a descriptor, in three cases both as a descriptor and title word and in one case solely in title[61]. The used descriptors are (all descriptors translated by the author):

---

[60] Titled as: "Taikinasta tavoitteisiin : isiä odotetaan mukaan koulujen vanhempainiltoihin" (From pastry to aims: fathers are expected to participate in the schools' parental evenings), kw: *koti*/home, *koulu*/school, *yhteistyö*/co-operation), *vanhemmat*/parents, isät/fathers.

[61] This record was the only one in this sample that had no descriptors.

NO	Finnish (original)	English
6	koti-isät	stay-at-home fathers
3	isät	fathers
2	hoitovapaa	parental leave
2	kotityö	home work
2	miehet	men
1	äidit	mothers
1	isyys	fatherhood
1	isyysloma	paternity leave
1	kotihoito	home care
1	lapset	children
1	päivähoito	day care
1	vanhemmuus	parenthood
1	vanhempainloma	parental leave
1	Yhdysvallat	United States

Table 14: *ARTO koti-isät –sample (7 records, 1990-2002)*

The topics about *stay-at-home fathers* were more concerned with leave and family benefits, and less concentrated on childcare, parenthood and gender status than the topics of stay-at-home mothers in the ARTO *kotiäidit*-sample.

Common dictionary equivalents are *kotirouvat* and *perheenemännät* (cf. chapter 7). Their use in ARTO was also evaluated.

The search (13.10.2004) with the Finnish equivalent **kotirouv?** in the title or descriptor field resulted in eight references from the years between 1960-1996[62]. The word *kotirouvat* occurred once as a descriptor (1992) and seven times as a title word. All the retrieved references were about periodical articles written in Finnish. The word was used similarly in polemic sense in the 1960's, 1980's and 1990's. The aim was clearly to act as a signal word - to seize the readers interest (cf. Forsman 2005), as in a title that plays with connotations - "Ammattilaisia, kotirouvia, kukkahattutätejä, työttömiä... : kotipalvelu maailmalla monenlaisissa käsissä" (Professionals, housewives, aunties with flowery hats / fussy aunties, the unemployed… : home care abroad/overseas in many kinds of hands).

---

[62] One reference from 1960, one from 1961, one from 1981, one from 1985, two from 1989, one from 1992 and one from 1996.

A search (14.10.2004) with **perheenemän? OR "perheen emän?"** in the title or as a descriptor resulted in eleven references, published between 1960-1997[63] and in Finnish. Ten were periodical articles and one an article in a book. Only in one of these did *perheenemännät* occur in a title (record from 1986) and thus it was mainly used by the authors of the articles.

The five records from 1990 onwards were studied further; three were book reviews and two periodical articles. All were concerned with 1920's and 1930's homemakers and the word *perheenemännät* was used as a near-synonym for homemaker/lady of the house. The motivation to use the word seemed to be its connotation – that of expressing something historical, out-of-date today but current eighty years ago (cf. Venuti 1995 and 1998, Engelberg 1993 and 1998; Lehtonen 2000).

- **Summary and Conclusions**

The family roles terminology as understood in the working version of ELSST and in the other thesauri studied (cf. 9.1) was in general rather foreign and unusable in the Finnish ARTO database. *Family roles* (FI: *perheroolit*) resulted in some relevant records, in which family was commonly understood to consist of two generations (parent(s) and children), but the types of families varied (traditional nuclear families, remarried families etc.). The Finnish aspect (cf. un-Finnish databases in later chapters) had clearly a strong emphasis on parenthood. *Breadwinners* (FI: *perheenelättäjät*) and *heads of households* (FI: *perheenpäät*) were foreign concepts in ARTO and seldom met, but understood in a euphemistic sense and thus differently from other discourses studied. *Homemakers* (FI: *kotitaloustyötä tekevät*) was also a foreign concept as well as the English version *housewives*, which Finnish (direct) translation *kotirouvat* was seldom used and in an aggravate sense. The Finnish translation *kotiäidit* and *koti-isät* were very Finnish, but the topics related to them had a stronger emphasis on caregiving and well-being issues than in the English versions, where labour force issues are respectively emphasised more than in ARTO.

The number of terms representing a transliteration problem rather than a translation problem was less than one tenth in ARTO, and the indexing style was rather gender-neutral. There were still great differences among the indexing of different records especially in terms of the number of given descriptors. The common vocabulary tool used was The General Finnish Thesaurus, YSA.

---

[63] Two were published in 1960, two in 1961, two in 1986, one in 1991, one in 1994, one in 1995, one in 1996 and one in 1997.

8.1.1.2 Finnish database LINDA

The second Finnish database studied is LINDA. It is the union catalogue of Finnish University Libraries and also includes the National Repository Library, the Library of Parliament, the Library of Statistics and Lahti Science Library. It contains references to monographs, serials, cartographic materials, audiovisual materials, electronic resources, multimedia and archives. (Helsinki University Library 2005, last updated 15.2.05) As in the case of ARTO, institutions conducting indexing for the LINDA database are truly many, and a common indexing tool (YSA thesaurus) is used, and in addition to which special vocabularies are used.

The variations in the number of indexing terms in LINDA was even greater than in ARTO and a remarkable maximum number of 53 descriptors was given for one document.

(The research questions concerning this part of the study, the used search strategies as well as dataset harmonisation procedures are the same as reported in the context of the other Finnish database studied, ARTO.)

The searches were made during June and July 2003.

8.1.1.2.1 Family roles in LINDA

The descriptor OR title search[64] with the string **family role(s)** resulted in four records, referring to English books published in 1996 (USA), 1976 (author Finnish, published in USA), 1974 and 1968 (same publication, different edition, USA). The first one had descriptors from YSA and LCSH (Library of Congress Subject Headings)[65] and the others from YSA.

The search[66] with the separate words **family and roles** (strings cut in order to include both plural and singular forms) resulted in 112 records, out of which 53 were from the studied time period 1990-2003.

Twenty-nine out of 53 were found to be irrelevant and/or duplicates and 9 were without keywords. The irrelevant ones were retrieved usually on the basis of the title

---

[64] Haku: yhdistelmähaku = (family role?)[in Asiasana/Ämnesord/Subject]OR(family role?)[in Nimeke/Titel/Title]

[65] YSA: perhe/family, historia/history, keskiaika/Middle ages, Eurooppa/Europe, avioliitto/marriage, naisen asema/women's status, sosiaalihistoria/social history, perhehistoria/family history. LCSH: Family, History, Marriage, Social history, Medieval, 500-1500, Family in literature.

[66] Haku: yhdistelmähaku = (famil? role?)[in Nimeke/Titel/Title]OR(famil? role?)[in Asiasana/Ämnesord/Subject]

field[67], some were retrieved because of the note field. The **sample studied** further includes 17 records, having the concept of *family roles* in title and/or descriptor field. The search strings also occurred in other words than family/families and role/roles. Relevant records were also retrieved with narrower descriptors including search strings, such as *sex role* (sukupuolirooli) or *traditional roles* (perinteiset roolit) and *re-married families* (uusperheet) or *American nuclear family* (amerikkalainen ydinperhe). More specific terms, as in "Family and changing gender roles"[68] were also often used in the titles.

Three of the records have descriptors solely in Finnish (YSA terms) and one solely in English (no vocabulary source named), in twelve cases the sole English vocabulary tool mentioned was The Library of Congress Subject Headings (LCSH), in one case the Medical Subject Headings (MeSH) was also used. The English vocabulary tools mentioned are both American. Additionally, in two cases geographical descriptors in English were provided – these were in a separate field.

For the further analysis the English indexing was complemented with the Finnish one and the Finnish descriptors were translated into English (conceptual duplicates were removed). The 49 additions to the English material are: Men's status (4 times), Women's status (4 times), Employment (3), Home work (3), Sex role (3), Division of labor (2), Families (2), Family (2), Labor market (2), Re-married families (2), Women (2), Cohabitation, Divorge, Families with children, Family planning , Gender, Identity, Interaction, Life cycle, Marriage, Mothers, Northwest Europe, Parental relationship, Power, Sexual behavior, Sex life, Sexual harassment, Sexuality, Social conditions, Working life, 1990's.

On average, the additions were in their nature more general words than the original English descriptors. When comparing the English language indexing, mostly based on LCSH, with the Finnish language indexing, mostly based on YSA, it was obvious that the differences are mostly due to the vocabulary control tool used. LCSH, for example, makes it possible to use specific terms *Dual career families* and *Work and family* and *Working mothers*, while the user of YSA covers the same concepts with more general terms such as family (*perhe*) and työelämä (*employment*) and women (*naiset*).

The sample has 75 different descriptors with altogether 131 occurrences (on average 7,7 per record). The picture below illustrates the co-occurrence of the ones most used (with 2 or more occurrences):

---

[67] Examples of titles, in which family and roles together do not form a concept in the sense relevant for the study: "Balancing work and family : the role of the workplace"; " Discovery of a new family of soluble pyrophosphatases and the functional role of quaternary structure in pyrophosphatases"; " The nerve growth factor family of neurotrophic factors : emphasis on their role in the central nervous system".

[68] In the English language indexing based on LCSH was constantly used form *Sex role* to refer to gender roles.
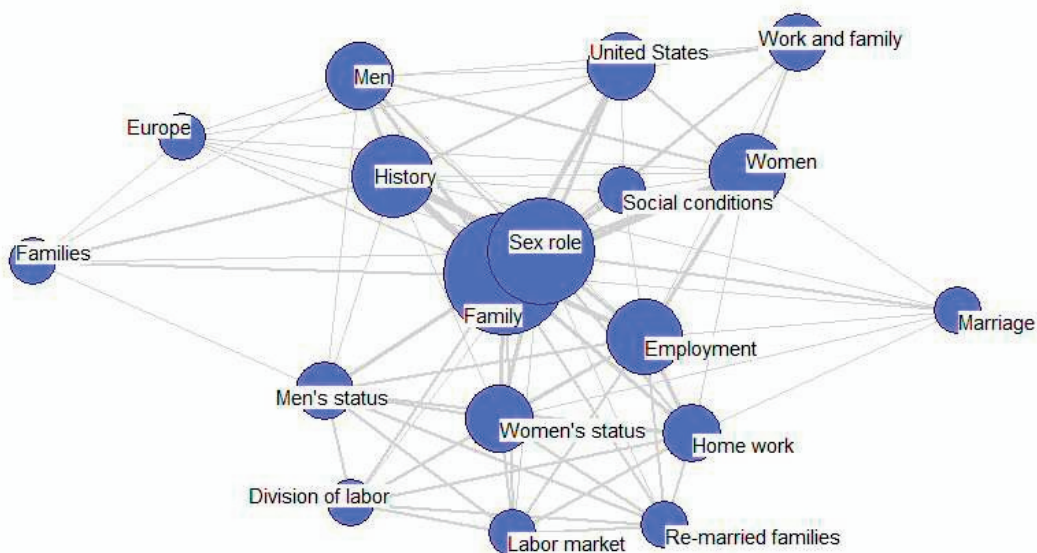
Figure 25: *LINDA famil* AND rol* sample (17 records, 1990-2003, descriptors with more than one occurrence)*

Together the words *family* and *roles* and their narrower words as a title and/or descriptor word resulted in material where together they are equivalent with the way *family roles* are understood in the other discourses studied (cf. esp. chapters 7 and 9) and thus understood as a social role within a family and with a stronger gender aspect. The problem in seeking is that the strings also result in considerable amounts of irrelevant material, in which they do not correspond with *family role* as a social science concept and term.

The Finnish equivalent **perheroolit** (singular and plural forms included in the search[69]) resulted in five records from the years 1993, 1992, 1982 and 1973 (the last one with a duplicate). The three oldest have no keywords. All the titles clearly indicated that the topic concerned concepts or practices of family roles, and the records with descriptors are entitled "Visioita arkielämästä: lukion päättöluokkalaisten käsityksiä työ- ja perherooleista sekä niiden yhteensovittamisesta osana arkipäivää vuonna 2010" / *Visions of everyday life: the last year highschool students' ideas about work and family roles and their integration as part of everyday life in 2010* and "Työ- ja perheroolien yhteensovittaminen vammaisen lapsen vanhempien kannalta" / *Integration of work and family roles from the perspective of parents of handicapped child*.

In the first, (from year 1993) the concept was covered with YSA term gender roles (*sukupuoliroolit*) and in the second with terms from MeSH: *family, work, daily activities*.

---

[69] Haku: yhdistelmähaku = (perherool?)[in Asiasana/Ämnesord/Subject]OR(perherool?)[in Nimeke/Titel/Title]

Hakutulos: Viitteet 1 - 5 (yhteensä 5).

Perheroolit (family roles) is thus a Finnish concept and used by Finnish authors, although it is very rarely used and never used as a documentary language term.

The search[70] with the separate strings **perhe* (famil*) and rooli(t) (role(s)** resulted in 57 records, out of which 33 were from the studied time period 1990-2003. After removing records without any keyword (3 hits), duplicate and irrelevant ones (altogether 8 hits), the **sample** was reduced to 22 records from the years 1991-2002.

Out of the 22 records studied further, 19 were published in Finland, 18 of them in Finnish and one in Swedish. Two were published in English (one in United States and one in Sweden), and one in Norwegian in Norway. **Linguistically**, therefore the sample had a rather Finnish bias. In all the cases, the common indexing vocabulary tool used was YSA, and in addition to this, in five cases Allärs (- the Swedish language version of YSA) was also used. Other indexing tools used were; in one case LCSH, in one case MeSH, and in one case also "free words" in Finnish and in one case "free words" in English. The indexing with several languages was not equal between different language versions, but the Finnish indexing was in each case clearly more specific both in terms of the number and depth of descriptors (e.g. in extreme case Finnish *mieheys; sosiologia; itsetunto; tunne-elämä; seksuaalisuus; minäkuva; sukupuoliroolit; identiteetti; pojat; isyys; vanhempi-lapsisuhde; isät; miehet; roolit; elämänhallinta; nuorisotutkimus; miestutkimus; nuoriso; sukupuoli; elämäntapa; elinolot; kasvatus; perhe; mies; lapsi; ikäryhmät* versus in English *men; psychology; gender indentity; fathers; adolescence*.)

In five cases, the indexer had used the concept of family roles with pre-coordination, and thus used family (*perhe* 2 hits) or families (*perheet* 3 hits) together with roles (*roolit*). It was more common to use other terms, i.e. more narrow or related ones. The term and concept of roles was found in most of the records as an indexing term, but it was also covered by the term gender roles (*sukupuoliroolit*). Family or families (*perhe* or *perheet*) were also common descriptors, but sometimes the concept was included in a narrower or related term such as in family life (*perhe-elämä*), family members (*perheenjäsenet*) or family education (*perhekasvatus*).

The sample has 148 different descriptors with altogether 259 occurrences (on average 11.8 per record). The most common descriptors are listed in the table below (original Finnish ones and English translations):

---

[70] Haku: yhdistelmähaku = (perhe? rool?)[in Asiasana/Ämnesord/Subject]OR(perhe? rool?)[in Nimeke/Titel/Title]

NO	Finnish (original)	English
22	roolit	Roles
14	perhe / perheet	Families
11	isät	Fathers
8	perhe-elämä	Family life
7	vanhemmuus / kasvatus[71]	Parenting
6	miehet	Men
6	isyys	Fatherhood
5	sukupuoliroolit	Sex role
4	kotikasvatus	Home education
4	lapset	Children
4	sukupuoli	Gender

Table 15: *LINDA Finnish version of family + roles sample (1990-2003, most common descriptors)*

The topics are varied, but the clear and common themes are parenthood (especially fatherhood), family life, values and family education, division of labor especially in the context of farming:

---

[71] *Kasvatus* was translated in two cases as "parenting" (based on its context).
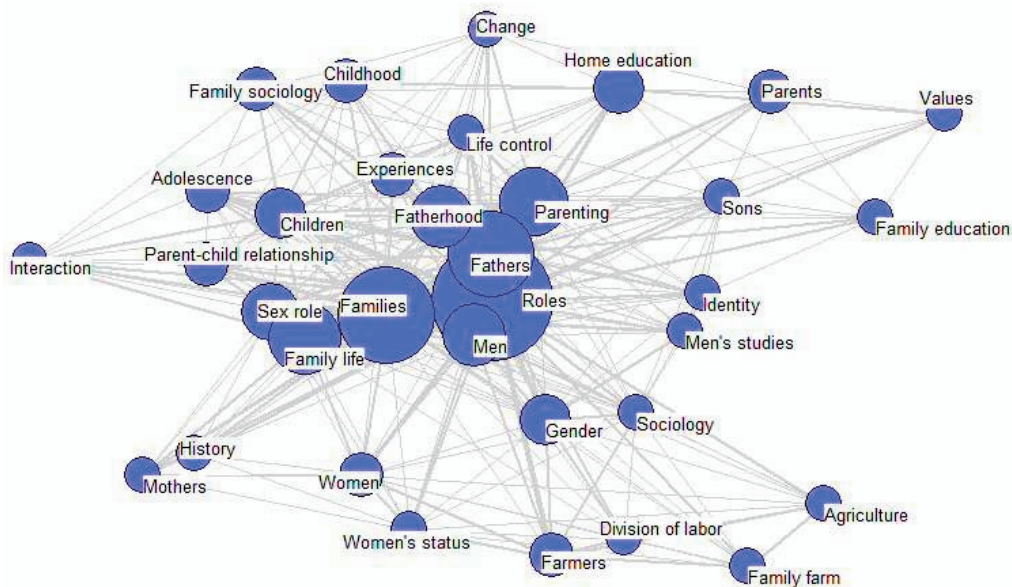
Figure 26: *LINDA family + roles (perhe+roolit) –sample (22 records, 1991-2002, descriptors with two or more occurrences)*

The result is somewhat similar to the result in ARTO and COPAC. In this case, the material indexed and the perspective it represent is clearly Finnish. The material was found in the fields of sociology, education and psychology, and no fiction or general reports etc were retrieved within this sample as (cf. later) in COPAC. This sample shows that as a concept family roles (*perheroolit*) already exists in LINDA, although it has no single descriptor as can be found in YSA. The concept is also covered with other, narrower or related terms, e.g. *sukupuoliroolit* (gender roles) and *perhe-elämä* (family life).

As seen previously with the other samples studied, the use of geographical descriptors was not in relation to the document content. Most of the documents were about Finnish practices, but only in two records was this revealed by the descriptors (*Finland, Lapland*). One document was about American practices, and in that case the descriptors used were reliable (*United States*, 1 occurrence). (The German and Norwegian records did not provide enough information to judge the original documents' geographical relatedness.)

8.1.1.2.2 Breadwinners and Heads of household in LINDA

The searches[72] with ***breadwinner*** (including different spelling variations) in the title or descriptor field resulted in eight records.

The form "bread winner" resulted in two records from the years 1934 (as title word) and 1952 (in both used as title word, no descriptors). The form **breadwinner(s)** resulted in six records (incl. one duplicate), all from the studied time period (1994-1999). The indexing was very heterogeneous: in two cases no descriptors were provided, in one case only in Swedish, in one case in Finnish (YSA) and English (LCSH: *Women; Employment; Europe; Men; Work and family; Sex role in the work environment*) and in one case in both Finnish and Swedish. The five first mentioned were relevant on the basis of their title and/or descriptors, but in the last one the word was used as a metaphor for a company (title: "Kuvia leivän isästä : Kemiyhtiön taipaleelta 1893-1993 = The breadwinner : a pictorial history of the Kemi company 1893-1993").

The Finnish ELSST equivalent ***perheenelättäjät*** (searches perheenelättäj? OR "perheen elättäj?") resulted in no records, and is thus not an established metadata term in Finnish.

The success with the related term ***heads of household*** was similar. The search[73] resulted in two irrelevant records from the years 1974 and 1975[74]. The Finnish ELSST equivalent ***perheenpäät*** (search[75] also incl. the form "perheen pä*") resulted in five records, out of which three were from the studied period. One was irrelevant (titled as "*Lapsen ja perheen päivystyspoliklinikkakäynti –tutkimus*" / A study of child's and family's visiting in emergency department of a hospital), one partly relevant ("*Perheen päätöksenteko auton hankinnassa*" / Family's decision-making in purchasing a car) and one relevant ("*Kun joku lapsista ryhtyykin perheen pääksi*" / When one of the children decides to be the head of the family). The indexing was in this sample also very heterogeneous: in the two most recent cases English MeSH descriptors and Finnish YSA descriptors were provided, in one case in Swedish and in one case in Finnish and English (vocabulary sources of the two last mentioned are not recorded).

---

[72] Haku: yhdistelmähaku = ("bread winner?")[in Nimeke/Titel/Title] OR ("bread winner?")[in Asiasana/Ämnesord/Subject]

 Haku: yhdistelmähaku = (breadwinner?)[in Asiasana/Ämnesord/Subject]OR(breadwinner?)[in Nimeke/Titel/Title]

[73] Haku: yhdistelmähaku = ("head of household?")[in Asiasana/Ämnesord/Subject] OR ("head of household?")[in Nimeke/Titel/Title]

[74] Both titled as: "Oral hygiene among youths 12-17 years, United States : estimates of the simplified oral hygiene index (OHI-S) for noninstitutionalized youths aged 12-17 years in the United States, by age, sex, race, family income, education of head of household, and geographical region, and a correlation analysis of the interrelation of oral hygiene, periodontal disease, and selected demographic characteristics"

[75] Haku: yhdistelmähaku = (perheen pä?)[in Kaikki sanat/Fritext/Keyword Anywhere]OR(perheenpä?)[in Kaikki sanat/Fritext/Keyword Anywhere]

The terms *breadwinners* and *heads of household* are thus not usable in the Finnish LINDA context. In order to seek information about them in English, different terms have to be used.

8.1.1.2.3 Homemakers and Housewives in LINDA

The search[76] ***homemaker*** (also the form "home-maker?" tested) in the title or subject field resulted in 24 records published in 1934-2002[77], out of which 13 were from the years 1990-2002.

How were the records indexed? Two records (out of thirteen retrieved from the years 1991-2002) had no descriptors. They were both published in the United States and in English. The searched string appeared in the title field in the sense of "home aid".

Linguistically the sample was rather heterogeneous, and therefore a further analysis is needed. Nine records were published in Finland: seven in Finnish, one in English and one in Swedish. The meaning of the word *homemaker* was rather homogeneous, referring in most cases to *home aid* (as a title word or as in MeSH descriptor *homemaker services*). In one case it was understood as a family mother (document written in English, but published in Finland). If Finnish descriptors were provided, the concept was commonly covered with YSA terms referring to home care, home care services and home nursing. In the table below the publishing country versus the language is summarised, (per record) as well as the field and meaning in which the string occurred; these being the YSA descriptors which were provided to cover the concept, and their translations:

---

[76] Haku: yhdistelmähaku = (homemaker?)[in Asiasana/Ämnesord/Subject]OR(homemaker?)[in Nimeke/Titel/Title]

[77] The search "home maker?" resulted in three records. One was the same as that retrieved with "home-maker" and the other two were eliminated as irrelevant, since they included the string as separated (home AND maker) in titles such as: "Home movie makers handbook", "Home banking the decision maker's guide".

Record	PY	Publishing country /Language	in the meaning of	YSA descriptos	YSA translation
D11	1991	Finland / Finnish	DE MeSH: homemaker services	nearest YSA: kotihoito, kotipalvelut	home care, home care services
D13	1991	USA / English	TI: kodinhoitaja (home aid, occupational title)	NO DE	
D10	1992	Finland / Finnish	DE MeSH: homemaker services	nearest YSA: kotihoito, kotipalvelut	home care, home care services
D12	1992	USA / English	TI: kodinhoitaja (home aid, occupational title)	NO DE	
D9	1993	Finland / Finnish	DE MeSH: homemaker services	nearest YSA: kotipalvelut	home care services
D8	1995	Germany/German	TI: kodinhoitaja (home aid, occupational title)	nearest YSA: kotisairaanhoito	home nursing
D6	1996	Germany/German	TI: kodinhoitaja (home aid, occupational title)	nearest YSA: kotisairaanhoito	home nursing
D7	1997	Finland/English	TI: perheenemäntä (family mother)	nearest YSA: kotitaloudet	households
D5	1998	Finland/Finnish	DE MeSH: homemaker services	nearest YSA: kotihoito, kotipalvelut, sosiaalityöntekijät	home care, home care services, social workers
D3	2001	Finland / Finnish	DE MeSH: homemaker services	nearest YSA-Allärs: hemservice; barnfamiljer	home care, families with children
D4	2001	Finland / Finnish	DE MeSH: homemaker services	only DE MeSH	
D1	2002	Finland / Finnish	DE MeSH: homemaker services	nearest YSA: kotihoito, terveyspalvelut	home care, health services
D2	2002	Finland / Swedish	DE MeSH: homemaker services	nearest YSA: kotihoito, terveyspalvelut	home care, health services

Table 16: *LINDA homemakers-sample (13 records, 1990-2003, summary)*

The Finnish ELSST equivalent **kotitaloustyötä tekevät** (search string "kotitaloustyötä tekev?") resulted in no records. The ELSST term *homemaker* is thus

foreign in Finnish, and when used in English, in the Finnish LINDA database it is understood mainly as a home aid instead of a family role.

The keyword search "**housewi?**" resulted in 73 records (53 hits in titles and 12 hits in descriptors). The 15 records from 1990-2003 were futher studied. Nine were published in the United Kingdom, four in the United States (+ one in Russia and one in Holland). 10 had housewife or housewives as a title word, two had both as a title and descriptor word, and three as a descriptor. Two had no descriptors, one had a descriptor in English, six in Finnish (three of them also in Swedish), and six in Finnish and English. (Additionally some had descriptors in Swedish.) How was the concept covered in the descriptors? Four had no reference to it – two had no descriptors and two referred to cookery books. In the other seven the studied concept was referred to in various ways – as women (*naiset*, 6 hits), and/or stay-at-home mothers (*kotiäidit*, 3 hits), as *housewives* (5 hits) and/or as wives (*aviovaimot*, 1 hit). The topics were mostly related to women's status (*naisen asema*/women's status 4 hits) and the emphasis was evenly placed on recent practices and on past (*naishistoria*/women's history 4 hits).

The title or descriptor search with the Finnish equivalent (direct translation for housewife) **kotirouvat**[78] resulted in 12 records, published in 1944-2002, and all on the basis of their titles. Three of them were from the studied period 1990-2003, and all referred to fiction (two were translations, one from English and one from Dutch). In the older records it was mostly used in the name of fiction novels. The Finnish word of *kotirouvat* does not represent modern language and especially not academic or documentary language.

The search[79] with the potentially equivalent **perheenemännät** as a descriptor or title word resulted in 87 records, out of which 11 (incl. 4 reprints) from the time period 1990-2003. It occurred in seven cases as a title word – four times as a historical word referring to past practices (cf. ARTO result) and three times as a humorous word referring to unskilled women (e.g. in a title "Ralliautoilua... perheenemännille ja muille vasta-alkajille" / *Racing… for familymothers and other beginners*). If a descriptor was used to refer to *familymothers*, it was women (*naiset*) or motherhood (*äitiys*). Similarly to the other Finnish database studied ARTO, the motivation to use the word seemed to be its connotation – to express something historical, out-of-date today or humorous today (cf. Venuti 1995 and 1998, Engelberg 1993 and 1998; Lehtonen 2000).

The search with the Finnish (equivalent for homemothers) **kotiäidit**[80] resulted in 68 records published in 1968-2003. Most of the records (52 hits) were from the study period 1990-2003. The use of the studied term had increased remarkably in the mid 1990's, but as a descriptor it was used for the first time in 1974. Out of the 52 records 13

---

[78] Haku: yhdistelmähaku = (kotirouv?)[in Asiasana/Ämnesord/Subject]OR(kotirouv?)[in Nimeke/Titel/Title]

[79] Haku: yhdistelmähaku = (perheenemännät)[in Asiasana/Ämnesord/Subject]OR(perheenemän?)[in Nimeke/Titel/Title]

[80] Haku: yhdistelmähaku = (kotiäi?)[in Asiasana/Ämnesord/Subject]OR(kotiäi?)[in Nimeke/Titel/Title]

were eliminated as duplicates (different editions etc.) and 5 as having no keywords, and thus the **sample** used for further analysis consisted of 34 records.

Sixteen cases were retrieved because of the used descriptor *kotiäidit*. In thirteen cases, the string occurred in both title and descriptor fields, and in five cases only in the title field. It was thus used more often by indexers, although it was also a common signal word in titles provided by authors, such as e.g. in the title "Ammattina kotiäiti : diskurssianalyyttinen tutkimus nykypäivän suomalaisesta kotiäitiydestä" / *Stay-at-home mother as occupation : discourse analys of modern Finnish stay-at-home motherhood*.

Linguistically and also culturally the sample was very Finnish – 26 out of 34 records referred to documents published in Finland and in Finnish, and in addition to one published in Finland but in English and one published in Finland but in Swedish. One document referred to was in Swedish from Sweden, two in English from the United States and three in English from the United Kingdom.

In the sample, 126 different descriptors were used with altogether 250 occurrences (on average 7.4 terms per record). The most common descriptors were:

NO	Finnish (original)	English (translation)
29	kotiäidit	stay-at-home mothers
12	äidit	mothers
12	naiset (incl. 1 singular form)	women
11	äitiys	motherhood
9	naisen asema	women's status
6	lapset (incl. 1 singular form)	children
5	kotihoito	home care
5	historia	history

Table 17: *Most common descriptors in LINDA Finnish version of the stay-at-home mothers sample*

According to the indexing the records were about motherhood, parenthood, care giving and education issues. When emphasis was on past, the topic was discussed together with women's participation in paid employment:

Figure 27: *LINDA Finnish stay-at-home sample (kotiäidit) –sample (34 records, 1990-2003, descriptors with 2 or more occurrences)*

The coherence on a conceptual and broader term level was more unified (cf. Iivonen 1989), e.g. the parenthood aspect was covered with several terms (*parenthood, parenting, motherhood* and *parent-child relationship*).

Descriptors referring to time or place were used 17 times. It is notheworthy (cf. other Finnish samples), that the cultural origin of the record becomes distorted – e.g. in the sample of Finnish equivalents for stay-at-home mothers (*kotiäidit*, cf. picture above) the emphasis is according to an even indexing between Sweden, United States and Finland, whereas in fact the cultural origin of Finland is superior to the others. As similar phenomenon occurred with the time aspect – terms referring to the past are visible, and the factual emphasis on recent or current practices becomes invisible. – The reader of the descriptors and bibliometric maps should thus be aware of "transparent matters" such as domesticity and modernity, which are commonly "hidden" in the indexing of Finnish samples (cf. Chen 2003; Venuti 1995; Yule 1996).

The search with the male equivalent for stay-at-home parent ***koti-isät*** (**stay-at-home fathers,** search string *koti-is?*) resulted in only 11 records, which is nearly one fourth of its female version. The first one was from year 1979, the second from 1982 (with no descriptors), and the third one, eliminated from the sample, was not relevant (a company name including the search string occurred as a title word and keyword). In addition there was one duplicate (different edition, 1997). The seven relevant records

from 1990-2003 were somewhat Finnish – in five of the documents the records referred to were also originally Finnish, one a translation from English into Finnish, one (geoculturally) Finnish but published in Swedish. The sample represented a rather even distribution of natural and documentary language: in three cases the word occurred in both descriptor and title field, in three cases a descriptor and in one case as a title word – thus four times as a title word and six times as a descriptor.

The sample has 43 descriptors, out of which 28 different. The most used descriptors are: stay-at-home fathers (FI: *koti-isät*, 6 hits), fatherhood (*isyys*, 4), parenthood (*vahemmuus*, 3), fathers (*isät*, 3), men (*miehet*, 2), family (*perhe*, 2) and homework (*kotityö* 2). As in the other LINDA samples also here the consistency was greater on theme level than on term level (cf. Iivonen 1989).
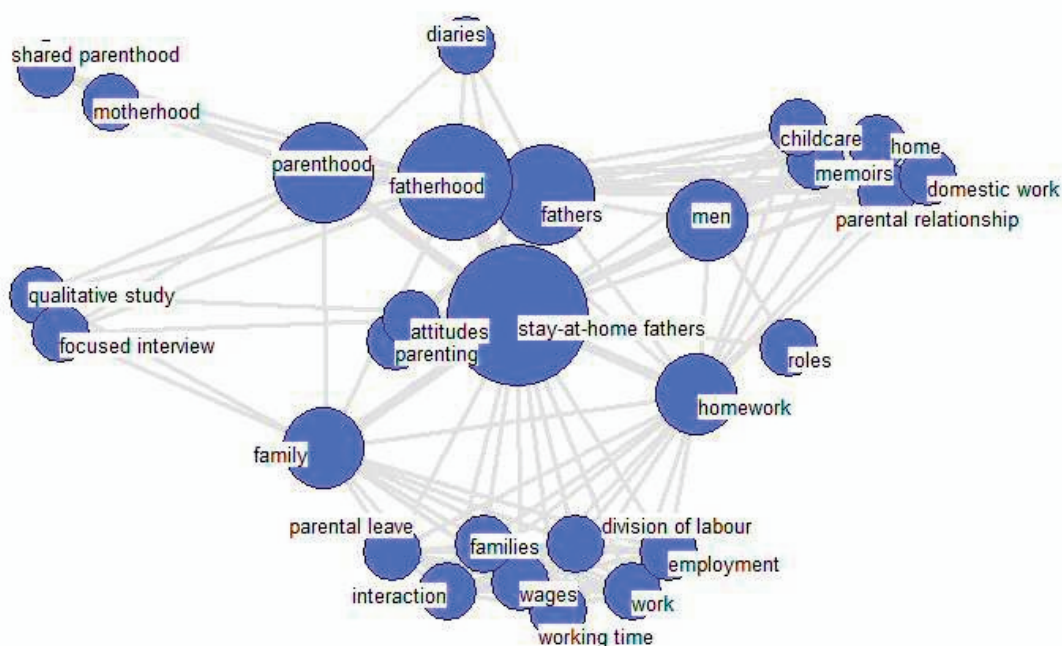
Figure 28: *LINDA koti-isät – sample (stay-at-home fathers, 1990-2003, 7 records)*

Stay-at-home fathers were often discussed together with issues related to domestic responsibilities and fatherhood and sometimes with employment. The word was understood similarly to stay-at-home mothers as a (gender-specific) parent staying at home in order to take care of own children (instead of earning money from employment). (It is thus semantically only partially equivalent to *househusbands*, as is *housewives* with stay-at-home mothers.)

- **Summary and Conclusions**

The indexing of the LINDA samples was also quite heterogeneous (cf. ARTO). In this case, the number of descriptors varied as well as, the different language versions.

Family roles was in general understood in similar way to that in ARTO, but in a more general sense and the emphasis more on the agricultural context and the division of labour. As a concept it clearly existed in LINDA.

Breadwinners and *heads of household* were clearly foreign terms and concepts, whereas *homemakers* was little used, but usually understood differently than in thesaurus discourse (cf. chapter 9) and mainly a home aid (occupational title). The Finnish ELSST equivalent was not used. *Housewives* was used more often as a title word than as a descriptor, and the 3Finnish reference to the concept was often expressed with the descriptor women (*naiset*). The Finnish housewife's direct translation *kotirouvat* did not represent academic or controlled language. The words commonly represented as modern equivalents *kotiäidit* (stay-at-home mothers) and *koti-isät* (stay-at-home fathers) became more popular during the 1990's. These were words represented in a somewhat even distribution between the documentary and natural language terms in LINDA.

The number of descriptors referring to names of time and place was less than one tenth, and usually current and domestic practices had no descriptors (as in Finland in the 1990's). Issues such as domesticity and modernity are thus commonly "hidden" in the indexing of Finnish samples and consequently in bibliometric maps unobservable (or biased) matters (cf. Chen 2003; Venuti 1995; Yule 1996).

8.1.2 British database COPAC

This chapter discusses how the study of family roles terminology is used in the COPAC database. COPAC is a British union online catalogue of the Consortium of University Research Libraries. At the time of data collection (September 2003 and September 2004) 26 institutions were contributing records to COPAC (University of Manchester 2004). The indexers do not conform to any established or common indexing guidelines nor controlled vocabularies. The retrieved material is more heterogeneous than e.g. the LINDA samples and also includes more fiction.

The search in COPAC covers descriptors (subject terms/keywords) and titles, and the search system recognises both singular and plural forms. The depth of the further analyses is dependent on the samples retrieved. **The general aim is to find out, how the terms are found and understood in COPAC and what is their indexing term context**. The emphasis here is mostly placed at the level of the term.

It became apparent that the indexing style of COPAC records varied very much both in terms of depth and style. Sometimes no descriptors (subject terms, keywords)

were given, and when given, they varied from one (e.g. "*Home economics*") to more than twenty (e.g. "*Women; Children; Family; Great Britain; Homemakers; Child care; Women's Health; Society; Role of, Women; Health services; Public health; Sociology; Housewives; Social change; Social policy; Social work with women; Economic conditions; Poor women; Social conditions; Mother and child; Home economics; Social aspects; Mothers; Health and hygiene; Social conditions; Maternal behaviour*").

The samples were slightly harmonised: if both singular and plural forms of one word occur, one form was selected in accordance to general thesaurus practices (cf. e.g. ISO 2788-1986). In addition, different spelling types were harmonised and clear typing errors corrected (KW *Housewifes* into *Housewives*). In addition, (very rarely occurred) names of persons and companies were eliminated, but (commonly occurred) names of geographical places were kept. The use of pre-coordinated indexing is common in COPAC, and it is replaced with single keywords. Other stylistic or conceptual changes are not made. Thus e.g. a record with *keywords* as:

> "KW‑ Child welfare ‑ Brazil
>
> KW‑ Poverty ‑ Brazil
>
> KW‑ Women heads of households ‑ Brazil
>
> KW‑ Brazil ‑ Economic conditions"

was coded for further co-word analysis as: "Child welfare; Poverty; Brazil; Women heads of households; Economic conditions".

When natural and relevant, titles were also further analysed from the focus of gender-neutrality. In the analysis the titles were evaluated as neutral, if they had no gender-related terms as title words or if they were used evenly (as e.g. in the title "The economics of women, men, and work"). The focus is especially on gender issues in three cases: breadwinners, heads of household and homemakers, since the gender aspect is known to cause challenges in the translation process (cf. esp. chapter 7).

8.1.2.1 Family roles in COPAC

The search (on 7.9.2004) in COPAC with the string "**family role***" and time limitation 1990‑2004 resulted in 45 records. **Linguistically** the sample was very homogeneous: 44 were published in English, one in Arabic. Ten of the 45 records were missing descriptors. According to the titles 30 of the records appeared relevant, 15 as non-relevant. The false hits were due to the search system and due to the nature of the term as being ambiguous and a general word.[81] Finally, only 23 out of the 45 records

---

[81] The search thus resulted in documents including searched string in titles such as "Neuronal cytoskeleton: roles in neuronal morphogenesis and organelle transport ‑‑ Cadherin cell adhesion receptor <u>family: roles</u> in multicellular organization and neurogenesis" and "The <u>family's role</u> in socializing youth".

fulfilled the two criteria and were thus considered as relevant and including descriptors. The 23 records were used as a **sample** for further analysis.

The 23 relevant records had in all 126 descriptors (on average 5,5 per record), out of which 71 were different:
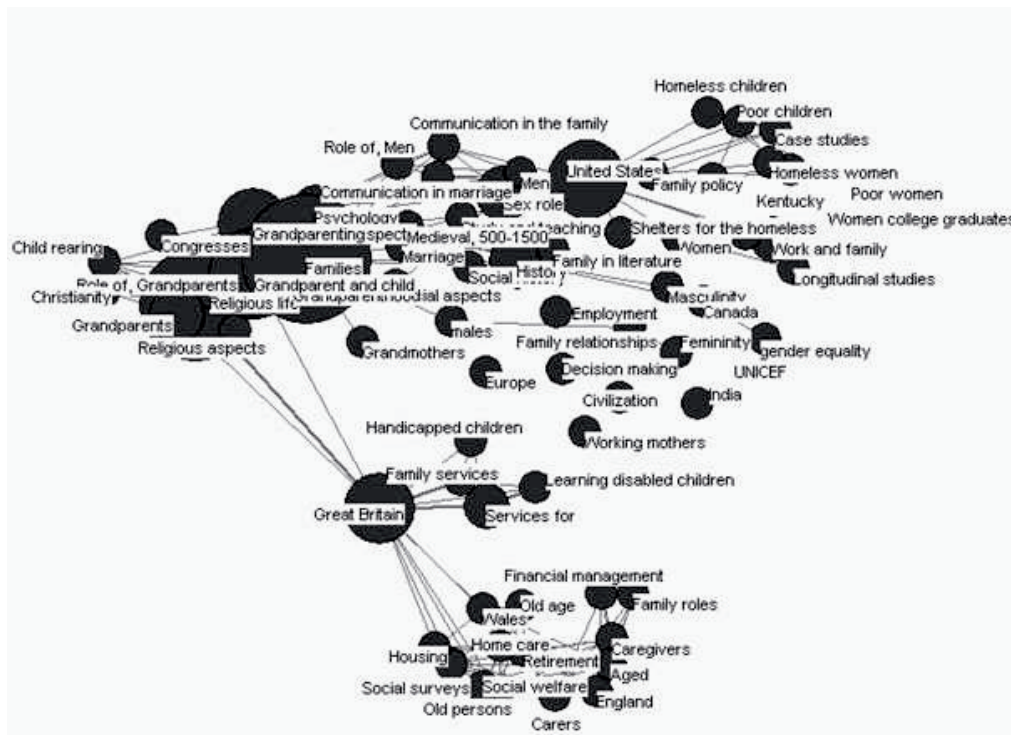
Figure 29: *COPAC family roles –sample (23 records, 1990-2004, all descriptorss)*

Only one record had "family roles" as a descriptor, and 22 records were thus retrieved in on the basis of their title (e.g. "Exploring variations in men's family roles"). The most commonly used **descriptors** of the sample were:

NO	Descriptors
16	Families
12	Role of, Grandparents
8	Grandparent and child
6	United States
5	Grandparenting
5	Great Britain
3	History
3	Sex role
3	Grandparents
2	Work and family
2	Services for
2	Psychology

Table 18: *COPAC family roles -sameple (23 records, 1990-2004, most common descriptors)*

Less than one fifth of the used descriptors represented a **transliteration** problem rather than translation problem – the sample included 8 place names with 17 occurrences (United States 6 hits, Great Britain 5 hits), 1 time name with one occurrence. The gender-specific keywording was slightly more female-related (7 terms, 7 occurrences) than masculine-related (4 terms with four occurrences).

The records with the string "family role*" and evaluated as referring to documents about family roles were mostly about grandparents role or influence within a family. As a topic it was commonly represented as American or British. Some records concerned working women, economic issues and family roles, and some family roles in families with handicapped children. The result is not clear, if all the descriptors are included, but if the number of descriptors is limited to the ones used more than once, the main theme is clearly apparent:
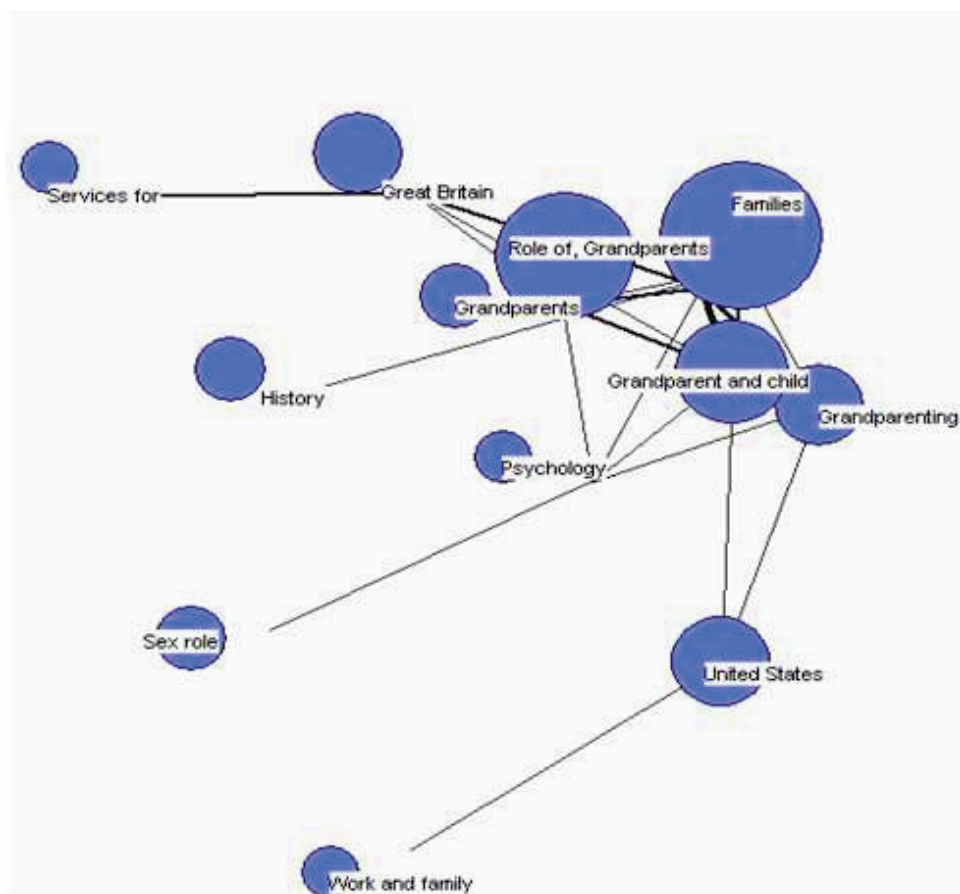
Figure 30: *COPAC family roles –sample (23 records, 1990-2004, descriptors with 2 more occurrences)*

The **discourse** of COPAC term *family roles* was thus very heterogeneous in terms of style, but as a topic it was rather more homogeneous resulting mainly in the role of grandparents in the family (cf. Iivonen 1989). Searching for *family roles* in the way it is commonly understood in the other discourses studied (cf. other subchapters in 7, 8 and 9) requires different strategies and terminology.

8.1.2.2 Breadwinners and Heads of household in COPAC

The keyword search (7.9.2004) in COPAC with the string **"breadwinner***" and a time limitation 1990-2004 resulted in 30 records, in all of which the string appeared in the title or subtitle field. The sample was **linguistically** homogeneous: 28 were published in English, 2 in Finnish and English. The sample was reduced for several (to a certain extent overlapping) reasons: 6 of the 30 records were missing descriptors, 7 were

reprints or different publishing forms or duplicates, and 3 were irrelevant[82], in which case the false hits were due to the term's euphemistic nature. Finally, only 16 out of the 30 records fulfilled the two criteria and were thus considered as relevant and included descriptors. The 16 records were used as a **sample** for further analysis.

In the sample, a total of 110 **descriptors** were used (on average 6.9 per record), out of which 68 were different and 14 used more than once:

NO	Descriptors
10	Employment
7	Women
6	Sexual division of labor
4	History
4	Social conditions
4	Work and family
3	20th century
3	Economic conditions
3	Great Britain
3	Men
3	Sex role
2	Economic aspects
2	Ontario
2	Single parents

Table 19: *COPAC breadwinners-sample (16 records, 1990-2004, most common descriptors)*

Descriptors referring to time were used 7 times (five times referring to the 20th century, and twice to the 19th century) and descriptors referring to place were used 18 times (5 referred to places in Great Britain, 3 in Europe, others world wide, and mostly at the level of country), and thus almost one fourth of the used descriptors in this sample represented a **transliteration** problem rather than a translation problem. In the picture below terms representing transliteration problem are removed in order to illustrate content which is more relevant from a thesaurus construction viewpoint. The semantic lexical indexing term network of breadwinner can be illustrated with the help of a co-word analysis map:

---

[82] Two records referred to a document titled "Kuvia leivän isästä : Kemiyhtiön taipaleelta 1893-1993 = The breadwinner : a pictorial history of the Kemi Company, 1893-1993", one to a document (letter, government publication) titled: "Planning and enforcement notice appeals by Mr & Mrs J McVey: The Breadwinner, 20 Runtsfield Place, Edinburgh". The descriptors given also revealed the records as irrelevant from the study's perspective.

Figure 31: *COPAC breadwinners-sample (16 records, 1990-2004, all descriptors)*

The word *breadwinners* was used and understood generally in the same way as in the thesauri studied (cf. chapter 9), that is, as a person bringing in family income gained from employment. It was used often with descriptors related to sex and family roles (e.g. *sex roles, work & family, sexual division of labor*). Clear themes (clusters) are sexual division of labour, employment and economic issues, single-parenthood and childcare, income distribution, social life and customs, identity issues in art and dual career families.

There were 6 different **gender specific** descriptors, used 14 times altogether:

NO	Descriptor
7	Women
3	Men
1	Girls
1	Married women
1	Single Mothers
1	Women in the theater

Table 20: *Gender specific descriptors in COPAC breadwinner-sample (1990-2004)*

Female related descriptors were clearly more common; only three out of 14 referred to men. They were used in 12 records, and thus 4 records were indexed in a gender-neutral way. All the records with the keyword *Men* also had its female equivalent *Women*. According to the information provided in COPAC (titles, subtitles, keywords, note field) two out of the four records without any gender-specific descriptors

referred to masculine emphasised and two to female emphasised document content. If used for women, *breadwinner* thus commonly needed a female qualifier (cf. Engelberg 1993; Venuti 1995 and 1998).

The starting point for the sample was thus **authors' discourse** and thus natural language (breadwinner(s) used in title or subtitle of the documents). Its indexing showed that in COPAC *breadwinners* occur most often in a modern female gendered context and in various geographical and cultural contexts. The records retrieved with the string breadwinner* were hits on the title (or subtitle) field and mostly about sexual division of labour and women's participation in family income either as a single-parent or as one partner in dual career families. The other common aspect was transition from past practices into a modern, dual career family model. It was most commonly used in the context of sociological studies (statistical as well as qualitative), but also a term used in literature studies, fiction and governmental documents. In order to study how documents about breadwinners are indexed, it is necessary to retrieve records about breadwinners by using other strategies (e.g. search terms like *Work & labor; Family & relationships; Income; Wages* etc).

The keyword search (7.9.2004) with the string "**heads of household***" and the time limitation 1990-2004, resulted 93 records, in all of which the string appeared in the title or subtitle or in the descriptor field. The sample diminished for several (in this case individual) reasons: 2 of the 93 records missed descriptors, 13 were reprints or different publishing forms or duplicates, and 3 were considered irrelevant, in which case the false hits were due to the search system's practice of also including other words between the searched words. Finally, 75 out of the 93 records fulfilled the two criteria and were considered relevant and these included descriptors. The 75 records are used as a **sample** for further analysis.

This sample was **linguistically** also very English dominated. Out of the 75 records 61 referred to documents published in English, 10 in Spanish, 2 in Portuguese, 1 in French and 1 in Swahili. The documents were mostly social science publications, but also a few fiction novels and governmental papers were also retrieved. Fifty-one records included the string in the keyword field, 5 in the title field and 19 in both fields.

In all 56 of the used descriptors with altogether 84 occurrences referred to geographical place name (such as e.g. 12 *United States*, 5 *India*, 4 *Chile*, 4 *Kenya*, 3 *Bangladesh*) and 8 (each used only once) referred to time (e.g. *20th century, Nineteen thirties, 1945*). Almost one fourth represented a potential **transliteration** problem rather than a translation problem. According to the records (judged by the metadata of the records) the **geographical** origin of the documents was rather global: 23 indicated to be from United States, 18 United Kingdom, 11 America (Brazil, Chile, Colombia, Costa-Rica, Honduras, Mexico, Paraguay), 9 Asia (India, Bangladesh), 8 Europe (Denmark, Sweden, France, Germany, Greece, Netherlands, Holland) and 6 from Africa (Mozambique, Republic of Zambia, Kenya, Lesotho). The indexing of COPAC is therefore rather multicultural at the content level, although very English at the language level.

The use of geographical names or words expressing a time period as a descriptor had no relationship e.g. to the publishing places or time of study which the records represent. It was also apparent that domestic issues were seldom expressed as domestic, and thus descriptors referring to the areas of United Kingdom as the place of study were used only twice, although there were many more which could have been indexed using this descriptor (*United Kingdom / England / Ireland / Scotland / London* etc.). This also especially applies to studies focussing on recent practices, such as, for example, of the 1990's and therefore they do not provide reliable information for co-word analysis, although in this case they do reflect the variety of geographical place names.

Fifty-three out of the 75 records were retrieved on the basis of the keyword field, 19 on the basis of keyword and title field, and only three solely on the basis of title field. Only 7 of the 75 records had *heads of household* as a descriptor, but even 65 had *women heads of households* as a descriptor. The gender-neutral form of the studied term seldom occurred.

In the sample, a total of 422 descriptors were used (on average 5.6 per record), out of which 176 were different and 53 used more than once. The most common ones were:

NO	Descriptors
65	Women heads of households
20	Women
20	Social conditions
18	Economic conditions
16	Poor women
12	United States
8	Statistics
8	Heads of households
7	Poverty
6	Single mothers
6	Household surveys
5	India

Table 21: *COPAC heads of household - sample (75 records, 1990-2004, most common descriptors)*

Mainly because of the lack of vocabulary control the co-word analysis of descriptors with two or more occurrences resulted in a very heterogeneous illustration:

Figure 32: *COPAC heads of household –sample (75 records, 1990-2004, descrpors with two or more occurrences)*

Heads of household was not much used in COPAC, but its female equivalent *women heads of household* was understood as a modern descriptor and as is generally the case in thesauri and dictionaries (cf. chapters 7 and 9) as a person having main economic responsibility in a family. In the female context it was often used for single mothers. Both words were used by the authors in titles and by the indexers in keyword field (the authors also used the form *female-headed households* and *women-headed households*). The gendered descriptors used in the sample were all female-related, except one, which expressed the absence of the masculine.

NO (tot. 131)	Descriptor	gender-relatedness
65	Women heads of households	female
20	Women	female
16	Poor women	female
6	Single mothers	female
4	Rural women	female
3	Urban women	female
2	Sex discrimination against women	female
2	Unmarried mothers	female
2	Women in development	female
1	African American single mothers	female
1	Afro-American single mothers	female
1	Feminist fiction	female
1	Matriarchy	female
1	Minority women	female
1	Mothers' pensions	female
1	Poor single mothers	female
1	Women immigrants	female
1	Women in agriculture	female
1	Women with social disabilities	female
1	Fatherless families	masculine

Table 22: *COPAC heads of household –sample (75 records, 1990-2004, gendered descriptors)*

The records retrieved with the search "heads of household" were usually not about role conflicts in Western societies (cf. especially the interviews in chapters 7, 8 and 9), however, a common context was developing countries, and single-mothers as maintenance providers of the family.

In the sample gender-relatedness occurred in various ways, and was thus analysed in detail. The table below summarises how gender-relatedness occurred in the title versus keyword field together with the *head of household* (hoh). The categories found are of eight kinds: 1) Symbol *fF* refers to the most common situation (40 cases), in which feminine related *women heads of household* is used as a descriptor, and also title expresses the emphasis being on women - but with other female related word(s) than *heads of household*. 2) Symbol *FF* refers to the second most common situation (17 cases) in which *women headed household* occurred both as a descriptor and title phrase (here also singular forms included and *female-headed households*). 3) Symbol *nF*

expresses that the descriptor *women heads of household* occurred together with gender-neutral title (6 cases). 4) Symbol ***nN*** refers to a gender neutral title being used together with gender neutral descriptor *heads of household* (5 cases). Exceptional cases were of four types. 5) Symbol ***mN*** expresses a combination of the gender neutral form of *heads of household* and masculine related word in title (2 cases). 6) ***Nn*** refers to a gender neutral form of *heads of household* as a title word being used with gender neutral descriptors (2 cases). 7) ***Ff*** describes how female related *women heads of household* have been used as a title word together with other female related descriptor (1 case) and 8) ***fN*** expresses that the title has been feminine related, but as a descriptor has been used gender neutral *heads of household* (1 case).

The eight categories are represented in the table below together with examples of titles and descriptors (signal words in italics):

NO (tot.75)	Types (gender relatedness in titles versus descriptors)	Example title	Example descriptors
40	female word in title field + female hoh descriptor (fF)	"*Women's* choices and the risk of poverty : case studies of breaking the cycle"	Poor women; Oklahoma; Case studies; Minority women; Economic conditions; *Women heads of households*; Economic conditions; Poverty; Research
17	female hoh descriptor + female hoh title word (FF)	"*Women-headed households* : diversity and dynamics in the developing world"	*Women heads of households*; Developing countries; Social conditions
6	gender neutral title + female hoh descriptor (nF)	"Finding the ties that bind : beyond headship and household"	Women heads of households; Developing countries
5	gender neutral title + gender neutral hoh descriptor (nN)	"Qualitative needs assessment of child-headed households in Rwanda"	Children; Rwanda; Social conditions; Heads of households
2	masculine word in title + genger neutral hoh descriptor (mN)	Poverty and family background in Greece : the role of the *father's* occupation and education	Poverty; Greece; Labor supply; Effect of education on; Greece; Heads of households; Employment
2	gender neutral hoh title + genger neutral hoh descriptor (Nn)	"The lands and people of Moray : population listings for the parish of Dallas, Morayshire : 1689 claims for recovery or compensation, 1777 heads of households and poor list, 1811 census"	Dallas (Scotland : Parish); Census; 1811; Genealogy
2	female hoh title + female descriptor (Ff)	"The Chilean 'social integration' approach to poverty alleviation : the case of the programme for *female heads of households*"	Poverty; Economic policy; Congresses; *Women*; Chile; Economic conditions; Social conditions; Social policy; 1988-
1	female word in title+ nender neutral hoh descriptor (fN)	"Del silencio a la palabra : un modelo de trabajo con las *mujeres* jefas de hogar"	*Heads of households*; Costa Rica; *Women*; Social conditions

Table 23: *COPAC heads of household -sample (75 records, 1990-2004 gendered term types)*

Further analysis showed that *heads of household* is generally considered to be masculine, and if the word is used for women, it is commonly expressed clearly with a feminine qualifier (such as *women-heads of household*, never used *men-heads of household*). Despite the apparent gender-neutrality of the term *heads of household*, it is in practice clearly masculine-related, and its semantic and pragmatic meanings partly differ (cf. Lehtonen 1992; Engelberg 1993; Yule 1996; Venuti 1998; Vehmas-Lehto 1999).

8.1.2.3 Homemakers and Housewives in COPAC

The search (25.9.2003) with "***homemaker***" as a title word and/or descriptor resulted in 154 records, published between 1869-2003. *Homemakers* occurred 154 times as a descriptor (1896-2003) and 76 times as a title word (author word, 1905-2003). It was used ambiguously both by the indexers and the authors. Nevertheless, the indexers were using the term in a more consistent way than the authors were doing in titles.

Twenty-one of the records were from the studied time period, that is from 1990-2003, but 9 of them were eliminated as irrelevant (3 without descriptors, 3 duplicates and 3 irrelevant and used as a metaphora e.g. for ceramic tableware). The **sample** that was studied further thus consisted of 12 records.

How was *homemaker* used as a descriptor between 1990-2003? It was used less ambiguously than in previous records and mostly (in nine cases of twelve) to refer to stay-at-home mothers and housewives. In two cases, it referred to a home aid and in one case to a proper name (a television programme).

How did *homemaker* occur in titles? During the time period 1990-2003 it occurred four times: Three times it was used to refer to a home interior decorator and once to a stay-at-home mother. In 1980-1989 it occurred more often, that is 11 times: Four times it was used to refer to a visiting nurse, three times to a stay-at-home mother, once to domestic aid/housekeeper, once to a bride (in future a homemaker) and twice it was used in a compound displaced homemakers (~divorced or widowed mothers who used to take care of the household and children and were economically dependent on their spouses).

The sample had 75 individual descriptors with altogether 109 occurrences (17 terms occurred more than once). The main topics are home economics, sexual division of labor, and *homemakers* are discussed together with women, household, family and women's roles. In the COPAC database the household aspect is strong, and the terms referring to roles (in this sample the used ones: *Sex roles; Role of, Women*) are relatively far more common than in the *housewives*-sample (see later).

The gendered descriptors used were different in all 11 cases having altogether 18 hits (*Women* 5 hits, *Mothers* 2 hits, *Housewives* 2 hits). No masculine-related terms were used. Potential **transliteration** problems were 9 with 14 hits (terms referring to time 3 each used once, and terms referring to geographical place 6 with 11 hits).

The diagram below illustrates how *homemaker* was used as a descriptor. The sample covers descriptors having two or more occurrences (12 of the most recent records 1990-2003 having descriptors):
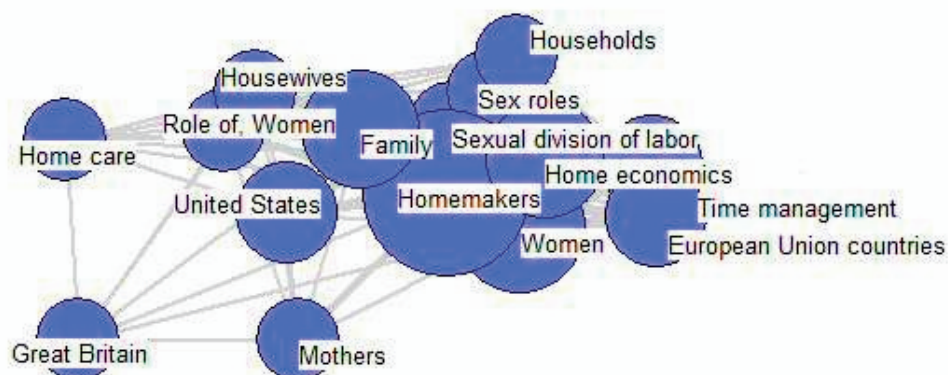
Figure 33: *COPAC homemakers –sample (12 records, as a descriptor, 1990-2003)*

The studied COPAC word *homemaker* is thus clearly a gender-specific, feminine word, and commonly understood as a social role. Its semantic and practical meaning in COPAC is similar to other discourses studied – refers to a person taking care of home and children and as a word is in principle gender-neutral, but in practice female related (cf. esp. chapters 7 and 9).

The search (7.9.2004) with "***housewives***" (singular form *housewife* included) as a subject term resulted in thousands of records, the oldest one published in 1623. Three hundred and one records were from the study period 1990-2004. It transpired that 128 records were duplicates (different editions), 15 having no descriptors, 29 were retrieved on the basis of their note-field, and three were published earlier than the time period studied (1758, 1860, 1986). This means that 175 records were eliminated from the sample, and the final **sample** includes 126 records published in 1990-2004 having *housewives* (or *housewife*) as a title word and/or descriptor.

This sample was also **linguistically** very much English dominated. Out of the 126 records 104 referred to documents published in English (out of which 3 in English and French, 2 in English and Dutch, 1 in English and Welsh), 9 in German, 4 in Spanish, 1 in Japanese, 1 in Swedish, 1 in French, 1 in Russian, 1 in Romanian, 1 in Portuguese, 1 in Marathi, 1 in Polish and 1 in Dutch. The documents were mostly social science publications, but also fiction novels, art books and cookery guides were retrieved. Eighty-five records included the string only in the descriptor field, 26 only in the title field and 15 in both fields. In the metadata (here titles versus descriptors) the use of *housewives* represented more the discourse of indexers' than authors'.

The 126 records further studied had 854 **descriptors**, out of which 331 different. On average the records had 6.8 descriptors. The most commonly used were:

NO	Descriptor
100	Housewives
37	History
35	Women
30	Home economics
26	Social conditions
22	20th century
17	United States
17	Employment
16	Great Britain
15	England
15	Wages

Table 24: *COPAC housewives –sample (126 records, 1990-2004, most common descriptors)*

In total, 58 descriptors with 136 occurrences referred to geographical place (most common ones *United States* with 17 hits, *Great Britain* with 16 hits, *England* with 15 hits and *Germany* with 9 hits). 22 descriptors with 52 occurrences referred to time (most common ones *20th century* with 22 hits, *19th century* with 8 hits, and *18th century* with 2 hits, others used only once). Thus over one fifth of the given descriptors represented a **transliteration** problem rather than a translation problem.

The documents indexed with *housewives* and/or having the term as a title word were mostly about women's employment, home economics, social conditions and sex roles. The perspectives were rather evenly on family and on societal level. Other clear sociological topics (see picture below) were technological innovations, consumerism and social life and customs.

Figure 34: *COPAC housewives –sample (126 records, 1990-2004, descriptors with three or more occurrences*

The discrimination of the term is weak, and it resulted in rather heterogeneous material. It was commonly understood as a social role (cf. other discourses in chapters 7 and 9), but also used as a general word referring to stay-at-home mothers. Thus other descriptors are much needed for describing the perspective and limiting the search (e.g. *Women consumers, Sexual division of labor*).

Within the context of the study it is also noteworthy that descriptors referring to roles were not used as much as the titles indicated. A document, for example, entitled "Motherhood at the crossroads : meeting the challenge of a changing role" had as descriptors *Mothers; Housewives; Psychology; Social conditions; Housewives; United States; Social conditions; Motherhood*.

- **Summary and Conclusions**

The emphasis in the analysis of the COPAC datasets was on the time period 1990-2004 and on descriptors and title words. The analysis clearly revealed the limitations of the system, various practices and the lack of a common vocabulary tool in indexing. Even today (3.9.2009) COPAC says about a subject search: "Not all records contain subject information, so a Subject search looks at the subject terms, plus the title and content notes in a record.", and about keyword search "Copac looks for Keywords in all parts of the record" (Copac Help 2009).

The family roles terms studied were used in COPAC, although not all of them were very common, and in some cases they were used in a different context than generally found in e.g. thesauri (cf. chapter 9.1). *Family roles* as a social role was linguistically a very English term in COPAC (publishing language of the documents the records retrieved in represented) and hardly used by indexers. Its lexical context was mainly related to grandparent role issues.

Breadwinners also represented natural language and English language material in COPAC and was used by the authors as a title word. It was used in the sense that is commonly understood in the other discourses studied. About one tenth of the descriptors which were co-used with *breadwinners* were gender-related, and female-related were clearly more common. The word was often used to refer to women and female-related words were used as qualifiers (male understood as a standard, cf. Engelberg 1993; Venuti 1995 and 1998).

Heads of household was used most commonly in the form of its female equivalent *women heads of household*. Its linguistic context was very English. It represented more documentary language than natural language, and it was commonly understood as a statistical and thesaurus term (cf. chapters 7 and 9), although its common context in COPAC was developing countries and single-mothers. As in the case of *breadwinner* the male was understood as the standard, and thus when *heads of household* was used to refer to women it was used together with words expressing the female content.

Homemakers was used rather ambiguously, but mostly in the meaning of stay-at-home mothers and housewives (cf. 7.1.1.4). It was clearly more common as a descriptor than as a title word. It was often used in the context of family and women's roles. It was clearly a feminine word and not used for men.

The term *housewives* was used a great deal in COPAC, and most commonly in the context of English language material. Although very general in nature, the term resulted mostly in social science material. It was clearly more often used as a descriptor than as a title word. The topics related to the word mostly concerned women's employment, home economics, social conditions and sex roles.

In most cases there was a clear difference between semantic and pragmatic meanings especially on gender-relatedness (cf. Lehtonen 1992; Engelberg 1993; Yule 1996; Venuti 1998; Vehmas-Lehto 1999).

The proportion of potential transliteration problem rather than translation problem, that is words referring to place, time and person names etc, was about one fifth of the used keywords. It is noteworthy, that when discussing potential transliteration problems (rather than translation problems) it is a question of generalisations. Within COPAC samples there also existed cultural time words (*Elizabethan* recipes, 1 hit), and a geographical reference with keyword *Domestic*. However, these kinds of descriptors that refer to proper names, time or place and need to be interpreted, are exceptional. The main challenge from the perspective of multilingual and multicultural thesaurus construction lies in other kinds of terms.

As discussed in the context of *breadwinners* the use of geographical names or words expressing time periods as a descriptor was not in relation e.g. to the publishing places or time of study which the records represent. It was also apparent that domestic issues were seldom expressed as domestic, and thus descriptors referring to the areas of United Kingdom as the place of study were seldom used. The same especially applies to studies with a focus on recent practices, such as those of the 1990's. Thus descriptors referring to time and place do not provide reliable information for co-word analysis, although in this case they do reflect the variety of geographical place names and time words (cf. Chen 2003).

8.2 Indexers

The focus of this chapter is to see how the interviewed British and Finnish indexers (tot. 6) represented the studied terminological case family roles and how they understood equivalence in multilingual indexing and thesaurus construction. First, the indexing practices, studied terminology and equivalence are discussed, and then the simulated indexing task by articles and phases. Last, the similarities and differences of the indexing discourse samples studied are summarised and concluded.

8.2.1 British indexers

This chapter will cover how the British indexers understand the terminological case (family roles terms), equivalence in multilingual indexing and thesaurus construction, as well as the results of simulated indexing task.

The topics discussed in the interviews were first the background of the indexer, secondly the indexing practices (material, languages, tools, principles), third equivalence, fourth the terminological case family roles (as concepts, sub-concepts, descriptors, word associations, definitions), and finally the simulated indexing task (see appendix 4 *Interview guide for the British indexers*).

8.2.1.1 Background

It was rather difficult to find British indexers who would be competent for the purpose of this study. Potential respondents were located by browsing the web-sites of relevant institutions, i.e. special libraries and information institutions including social science material. This resulted in contacting three people who unfortunately were not willing to participate in the study since they considered their work to be more "copying" than intellectual indexing. After this another strategy was used, and people known to be familiar with the field were contacted and asked for suggestions for potential

respondents. The strategy resulted in three suggestions, which were all considered relevant and which led to three interviews.

The informants are all female, and specialised into indexing of social science material. They represent four different social science institutions located in three different cities. (One respondent had two part-time positions.)

The indexed material was sent to the informants from one to two weeks before the interview took place. The interviews with the British indexers took place at the respondents' work place and lasted, on average, for 105 minutes (minimum 90 minutes, maximum 135 minutes). Notes were written during the interview, the interviews were tape recorded, and later transcribed.

8.2.1.2 Indexing practices

The indexed **material** was research reports and surveys, digital research data, articles and books. At their work, the informants indexed (i.e. gave keywords) in English, but two of three also indexed in languages other than English **language** material (Slavic languages, German, French, Spanish). Keywording of foreign language material was considered to also involve a little research, other than the use of dictionaries in order to find proper equivalents. The knowledge on subject domain was considered prior to language knowledge.

> "It is more important to have a deeper knowledge and understanding about the topic than a more advanced knowledge about the language." (BR IN2)

The **followed guidelines** were considered to be general (such as favouring plural forms and nouns - BR IN2, cf. Alaterä and Halttunen 2002), organisation-specific, and non-written. They were communicated via training and inhouse guidance.

> "--- no specific guidelines at the moment [for indexing practices] but we do have fairly specific training. When a new indexer comes she has the training and then in the beginning of her work is under the supervision of an experienced indexer. --- It can be quite hard to write --- It's more a question of getting the new indexer to think in the same way to know what to pick up and what to look for and so on. So it's probably quite difficult to produce anything written. But we do have certain rules that we follow." (BR IN1)

The main **tools** commonly used were organisation-specific keyword databases or thesauri. In two cases the controlled vocabularies were stable, but in one case there was (at the respondent's current work place) no vocabulary control.

> "[Tools used] depends very much on what you're doing it for. So for example when I worked in the [private sector] for five years we used --- when working in academic library there was at that time not any given set of subject headings - today they're probably using the Library of Congress Cubject Headings --- I tried to pick terms that have already been used and be consistent in that sort of way, but could add also our own new terms. There was not any overwriting control. That is similar to the current situation in [one of current work places] ---

> Main tool used is the keywords database. --- Other than that we use any number of reference books, we have a collection of dictionaries, we search on the Internet – we have a number of Internet sites we use regularly, for instance European Union online." (BR IN3)

The practices describing the nature of used descriptors varied (cf. Rowley 1988). In one case, for example, the indexer used only thesaurus terms, in other cases there were "own fields for controlled vocabulary terms and for other keywords" (BR IN2), and in the third case both controlled and free terms were given within one field.

In indexing the aim was to describe both the general and specific level of the topic, and it was also a clear guideline (example 1, cf. Rowley 1988). Different **discourses** were also considered (examples 2, 3, 4, cf. Buckland 1999; Lehtonen 1994; Venuti 1998; Vehmas-Lehto 1999).

Example 1:

> "--- [they] tend to tell people to use both specific and general terms, they often look keywords from the titles and subtitles, paragraph titles, subheadings, and there are own fields for controlled vocabulary terms and for other keywords" (BR IN2)

Example 2:

> "What we try to use to is to accommodate different vocabularies is to use a lot of leading terms or non-preferred terms and then make them lead to what we call a preferred term, which is what a study is indexed with and that way we can accomplish a little bit of free text language but still use the control." (BR IN1)

Example 3:

> "Keywording must be understood from the users' point of view and you got to use terms that the user is likely to look up or use. And if you don't, material is going to be missed and that's the great value of having a controlled vocabulary. --- "(BR IN3)

Example 4:

> "And another database that runs from here is [---] and that uses a controlled vocabulary and I'm sure that's a very good thing. And I think users can browse from the list what's being used, so again that gives people the chance to look different terminology. Because so many concepts can be represented by more than one word or phrase, and if you're not using a controlled vocabulary a lot of the material gets missed." (BR IN2)

Together with different discourses also the **time aspect** was often discussed (cf. Aitchison 1991; Suojanen 1993; Hirsjärvi and Hurme 2000) and it was seen as a construction problem (example 1) as well as an update problem (example 2).

Example 1:

> "And then language changes, doesn't it, as social perceptions change the use of word changes. If we again think of the term housewife, it was a perfectly accepted term thirty years ago. Everyone knew what it meant and was perfectly happy with it and didn't mind being called housewives, whereas now --- So it's all those sorts of issues as well coming to play, don't they. --- And different countries are moving at different rates, aren't they. --- the same kinds of changes are there, but not going at the same speed and some bits are going faster than other in different countries."(BR IN3)

Example 2:

> "--- new concepts like names of conflicts, e.g. September 11th. You don't know whether to include them into a thesaurus, if they are going to be known." (BR IN2)

Different discourses of users and ways of searching were acknowledged, as well as also **indexing inconsistency** (cf. Iivonen 1989 1995; Saarti 1999; Mai 2001) in several different occasions.

> "--- indexing is very subjective and if one person indexes the dataset from one year and then another person comes along and indexes the next year there is going to be differences because people just think differently. It's a very subjective thing and you know different people interpret concepts differently. So it keeps standardisation in that way. So that we know that if the person searching for the dataset finds one year they're going to find all the other years because of the same terms are in there." (BR IN1)

They considered receiving very little **feedback** (e.g. "what is the difference between assurance and insurance" BR IN2) from users. In some cases the feedback was considered insignificant.

> "We don't consistently add keywords to the records. We only add keywords if the authors of the books or articles etc have asked us to. And what I tend to do then is to edit the keywords they've suggested, because sometimes they are very unhelpful or I try to make the keyword in noun rather than in adjective and a plural rather than a singular and try to get a little bit consistency in that way. But we are really stuck with the kind of things that they want to add in, which we know is not over-helpful, but we haven't yet done what is necessarily to add a better thesaurus or controlled vocabulary." (BR IN3)

However, the co-operation with thesaurus constructors was considered to be natural and important part of indexer's work, and one indexer (BR IN2), also involved in thesaurus construction, described that thesauri are updated constantly, and when indexers come across a term they need they tell her. She then checks to see if a relevant term already exists that they have not found, and if not, she creates a new term. She also mentioned, that when they get a new subject specialist into the organisation (s)he usually suggests a lot of new terms to add.

The **usefulness of controlled vocabularies** was discussed via the needs of information seekers and indexers, and thesauri were considered useful, especially from the information searcher's point of view (cf. Kekäläinen 1999). Additionally, the thesauri were considered to demand constant updating in order to be useful.

> "Controlled vocabulary indexing is used, because it is easier for the user. --- The content of thesaurus is dynamic, has to be updated regularly. ---The whole thing has to be dynamic really to cope with the nature of social science research." (BR IN1)

The **future of thesauri** was considered to be needed and challenging. The major problem seemed to be the lack of common understanding of the role of thesauri in information retrieval (example 1). The rather new influence of web searching was also considered to have created new kinds of demands (example 2).

Example 1:

> "I doubt that most users would go to that effort of looking up one." (BR IN2)

Example 2:

> [*How you see usability of thesauri and their future*]
>
> "That's quite an interesting thing, because if we think of how people search information now… I mean most people go to the Internet and use a big search engine like Google and that sort of thing. --- I think that sort of search engine philosophy has very much led over into certainly this to the work we do here. We've done a lot of work on the websites trying to present the results in that sort of format, because that's what people are used to. And I think controlled vocabulary searching is really more and more important when you get more information, because otherwise the ways… It's a way of standardisation. With free-text searching it's very difficult to, I mean you put in the words that you want, but it's very unsorted mess, whereas if something is searched with the controlled, using the controlled vocabulary, thesaurus, and then, you know, that there is an element of standardisation that you're going to find things which are relevant. So I think it has been very much influenced by web searching, but it's definitely a big raw forest, though, very much so." (BR IN1)

Multilingual thesauri were considered as being even more necessary than monolingual ones (example 1), and often it was the same factors that were present in monolingual matters that became even harder to control or handle without multilingual thesauri (example 2, cf. Adler 1997; Hall 1981; Suojanen 1993 in chapter 1.2 *Objectives of the dissertation*).

Example 1:

> "Foreign material is more difficult, even if you're fluent in foreign language, because a lot of the terminology is very specific and there may not be a precise equivalent and then we have kind of approximate, to get as close as we can." (BR IN2)

Example 2:

> "I think they [thesauri] are even more needed in the Internet environment --- And it's particularly important if you're all doing international search, because the same term may be used differently in different countries. You know Britain and America use the same, well officially the language, but there is a lot of differences in what terms actually mean and when you're looking at it from a different language perspective, you know, if your main language is French say and you're looking at English language material, well how you will interpret the meaning of particular word will come very much from how your language deals with those concepts, won't it. (BR IN3)

8.2.1.3 Family Roles terms by British indexers

8.2.1.3.1 Family roles by British indexers

It is noteworthy, that when discussing the terminological case the indexers were in a slightly different position in comparison to the other groups studied, in the sense that before the interview situation they had been involved with the simulated indexing task (articles related to gender roles and caregiving issues within families, cf. chapter 6.3.5.1 *Indexed documents*) and its influence was sometimes very obvious.

[About family roles as concept]

"From the articles I've looked at it, they deal with division of labour within a home. There are various aspects that need to be covered, so economic aspects of working and bringing in money, housework and child care, those are the three main areas that I would see. And in the past there has usually been a very clear idea about which of the partners in the relationship would do which role, and that is less clear today. There are more opportunities for swapping roles about. And the articles I've read, you know, were quite interestingly showing the social consequences that arrive from that."

[*Later about subconcepts into a thesaurus*]

"None of the articles specifically talked about nursery provision, but that would probably also need to be a factor, and after school clubs and that sort of thing. And also some of the articles were hinting models of family life. I'm not familiar with that – with the terminology of that, but those various models that are being discussed today, they would need to be included." (BR IN3)

The stimulus word ***family roles*** resulted in 19 co-occurrences, out of which 16 were different. – As in the previous cases studied, the consistency here at the theme level was also greater than at the term level (cf. Iivonen 1989), and e.g. parents were covered by different kinds of words (*parents, mothers, fathers*).

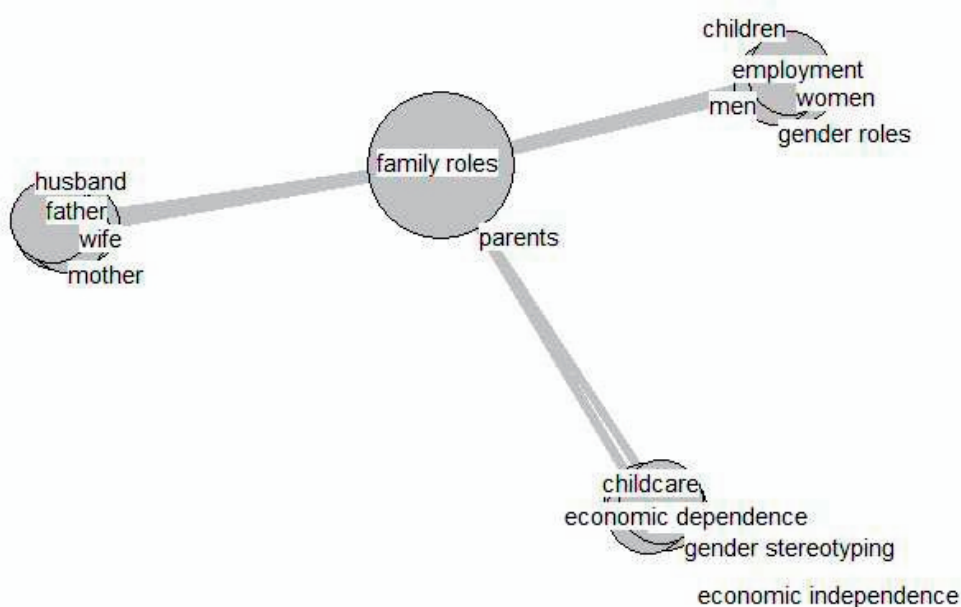**Figure 35:** *Word associations for "family roles" by British indexers*

Family roles resulted in rather neutral connotations expressing parents, gender issues (such as *gender roles, gender stereotyping*) and economics (such as *employment, economic dependence*).

In general, family roles were considered to be in flux, and this affects the terminology (cf. Aitchison 1991; Suojanen 1993; Hirsjärvi and Hurme 2000). British

and Scandinavian countries were considered to be rather similar – two thought Scandinavian to be more gender-neutral and gender-equal (see example 1 below), one (see example 2 below) to be in a rather same stage at least if compared with Southern-European and more Catholic countries. It was also considered more broadly than in the studied ELSST sample, and thus covering also other roles than just economic roles (see example 2 below, cf. chapter 4). The emphasis was still very much on the adult, and children were discussed more in the Scandinavian contexts (cf. example 1, cf. also chapter 7.3 *Similarities and differences in common and academic language use*). Changes in time were also considered to have impact on gender-neutral language use (example 2, cf. UCL 2000; Brunel 1999; School of Psychology 2002; University of Southampton 1997; Engelberg 1993, 1998).

[About how to define *family roles* in UK and later about differences and similarities between Britain and Scandinavian countries]

Example 1:

"I would mean the relationships that the members of the family have with one and other and the attributes they require in relation to another. One of you would always do the washing up or something like that, one would always be the main decision-maker --- and that may or may not be related to wider social roles, like gender role, or a relationship with childhood or that kind of thing. ---

My concept of Scandinavia may be very stereotyped, but I would imagine they have quite progressive nature of family roles and perhaps they are ahead of us in things like working mothers, and more of an equal divide between --- male and female roles. And I think they are, well the role of a child is slightly more… They treat as more equal I assume." (BR IN2)

Example 2:

"I would say it depends on how you define a family, of course, - speaking as a social scientist [laughing]. I mean a family unit, a nuclear family unit, was traditionally composed of an adult female, and adult male and one or more children. However that's not always the case now and I think you have to bear that in mind. And also the way that work has changed as well. It may be that … at one time we always used to, you know, the male was seen to be the head of the family and probably the highest earner, but again, that may not be the case now.

When you're talking about family roles nowadays you have to be gender non-specific and you have to be prepared to accept that traditional families, traditional family roles, might not be quite as they were. ---

Not necessarily the same in whole Europe. But what's happening, and has happened in Britain, is really happening in all over Europe really, and all over the world, as well. We've certainly this, certainly I would think possibly very large minority of families don't have a male householder anymore. It's often not a traditional marriage with two children anymore. And certainly in Scandinavian countries it's very similar. Maybe in countries and parts of Europe, which are more catholic like Italy or Spain and so on it's probably a higher proportion of traditional families. But I would think certainly in the cities that's changing as well." (BR IN1)

As descriptor *family roles* was considered useful and needed, although not necessarily to cover all the modern aspects of the concept - e.g. *mother*s should not be considered just biological (BR IN2). The informants thus made a distinction between the semantic and the pragmatic meaning (cf. Koller 1989).

> "I think so. Yes, because it's something that is Europe wide, although they may mean slightly different things in different countries. But certainly the nature of family is changing so much that a lot of research is going to be done in that area. And yes you do need terms to cover it, because these pieces of research are going to come up." (BR IN1)

The descriptors related to family roles in a thesaurus (as broader, narrower, related, and non-preferred terms) were various - beginning with more specific and concrete terms such as *father*, *mother*, *parent*, *child*, *husband*, *wife*, *breadwinner*, *homemaker*, *parent role*, *single parent*, *siblings* and ending to broader or more general terms as *socioeconomics*, *demographic change*, *gender roles*, *educational attainment*, *childcare*, *nursery provision*, *models of family life* and "perhaps" or "probably" *gender contract*, *role of the welfare state*, *maternity and paternity leave*, *part-time employment*. The suggestions tended to grow in depth, and usually no attempts were made to select a perspective (cf. Venuti 1998) or otherwise reduce the number of potential descriptors.

> "With siblings you might have the concepts to deal with their place like the middle child, the youngest child, those kinds of things are quite well defined. And there are these wider, I mean, like the mother-in-law has a quite well defined role." (BR IN2)

The descriptor mentioned also represented themes or concepts to be covered, and were thus often given to provide information about the context and possible themes around the studied descriptor instead of listing truly functional descriptors.

> "--- in some way educational attainments, because it seems that a higher education between women does have an effect on all the other things that we've talked about. There would have to be a term that would describe that. I'm not sure if the education attainment is the best word for that. ---" (BR IN3)

This case was not considered linguistically difficult, and the discussed challenges lie in the content, i.e. the concept level. According to British indexers there may thus be differentiation at the level of denotative, connotative, semantic and/or pragmatic meaning (cf. Vehmas-Lehto 1999 in chapter 3.1 *On Language, Culture and Meaning* and Koller 1989 in chapter 3.4.1.2 *The idea of equivalence*) between different languages and cultures.

8.2.1.3.2 Breadwinners and Heads of household by British indexers

The term ***breadwinner*** was already introduced by two of the informants within the context of defining *family roles*.

> "There used to be British phrase, which is not used that often anymore, called breadwinner, which basically means the highest income, the person who earns the highest amount." (BR IN1)

The stimulus word ***breadwinners*** resulted in 19 co-occurrences, out of which 16 were unique. Each informant associated it as a male and a worker (*men, worker, male-worker*), and here the consistency at the theme level was also greater than at the term level (cf. Iivonen 1989). It also associated with words referring to family income. In one case it also resulted in clear connotations, which were negative (*outdated, offensive*).
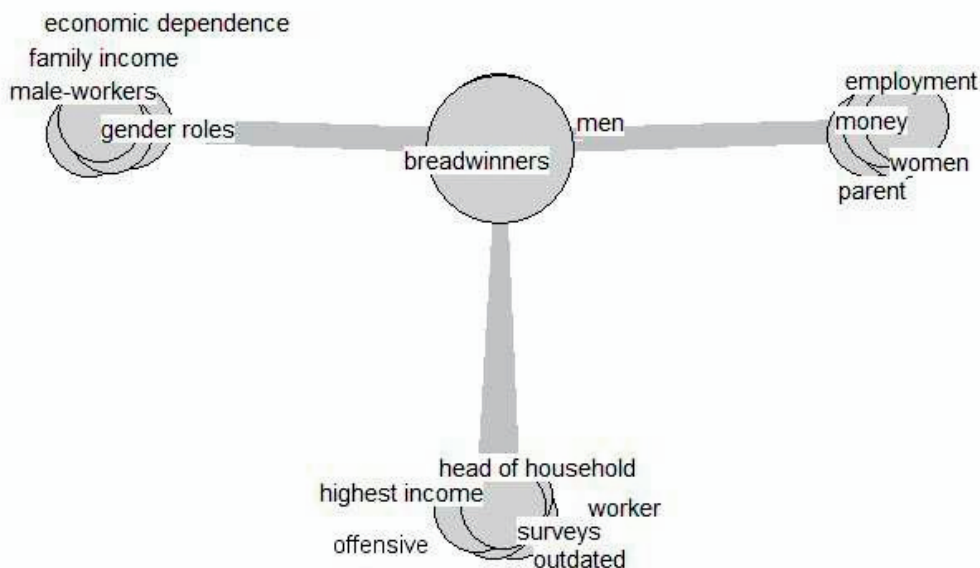
Figure 36: *Word associations for "breadwinners" by British indexers*

In general, *breadwinner* was not considered to be a good word to reflect modern practices. It was considered out-dated and gender-biased. It was also seen as a (near-)synonym for *head of household*. It was still considered a good descriptor, although there would also be better alternatives, such as *the highest income householder*.

> "The distinctive thing with breadwinners is that, they, whether they are male or female, their job pays the highest in the household. So they bring in the highest wage, effectively. ---
>
> There is a specific phrase here, which government surveys tend to use now, which is quite good: the highest income householder. Which is good because it's [gender] non-specific. ---
>
> [It is synonymous with] head of household. We tend to think that breadwinner is the head of the household, as well. Thought, of course, that might not be. It would be dangerous to call anybody with it, even to suggest that someone is the head of the household, nowadays." (BR IN1)

The semantic characteristics (cf. Nida 1975; Vehmas-Lehto 1999; Häkkinen 1994 in chapter 6.2.2 *Component analysis*) of *breadwinner* according to British indexers are *wage-earner*, *having the highest income in the household*, and often *male*.

The stimulus word ***heads of household*** resulted in 20 co-occurrences, out of which 15 were unique. *Heads of household* clearly associated with a man within a family, and with breadwinners and traditions:
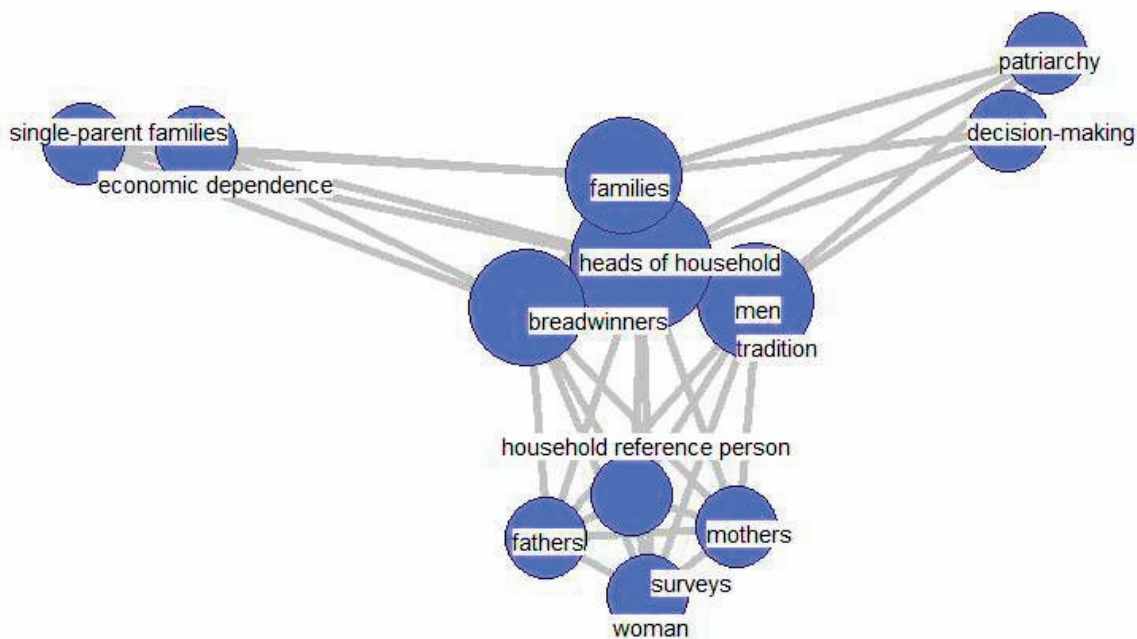
Figure 37: *Word associations for "heads of household" by British indexers*

The connotations of *breadwinner* were later discussed and considered as strong, which influence the usability of the term (cf. Ulrych 1992). It was considered as rude and to be commonly used to refer to men (see examples 1 and 2, cf. Venuti 1998).

Example 1:

"It's defined by income. That's also quite a loaded term and has a lot of connotations of a male head of the household, head of the family sort of thing, very patriarchal, sort of connotations.

I think with head of household it would be, I mean even if the wife earned more the man would still say that he is the head of the household. [laughing] And some of them still try, believe me. [laughing]

But I wouldn't say that it was the oldest man in the household, because, I mean, say if you get a family where so the two parents are in the work but they have either the wife's or the husband's father living with them, who is maybe retired so very depended on them. I wouldn't call him the head of the household. But by that definition [oldest man in the household] he would be." (BR IN1)

Example 2:

"--- But that's an interesting though that if you're moving towards a shared economic roles and household roles then how is it going to be possible to say who is the head of the household? I haven't thought that implication before." (BR IN3)

The term *heads of household* was still considered as being acceptable in a thesaurus, if it referred to past practices, but it should have also more modern equivalents.

8.2.1.3.3 Homemakers and Housewives by British indexers

Among the three British indexers *homemakers* (used as a stimulus word) resulted in 22 co-occurrences, out of which 18 were unique.
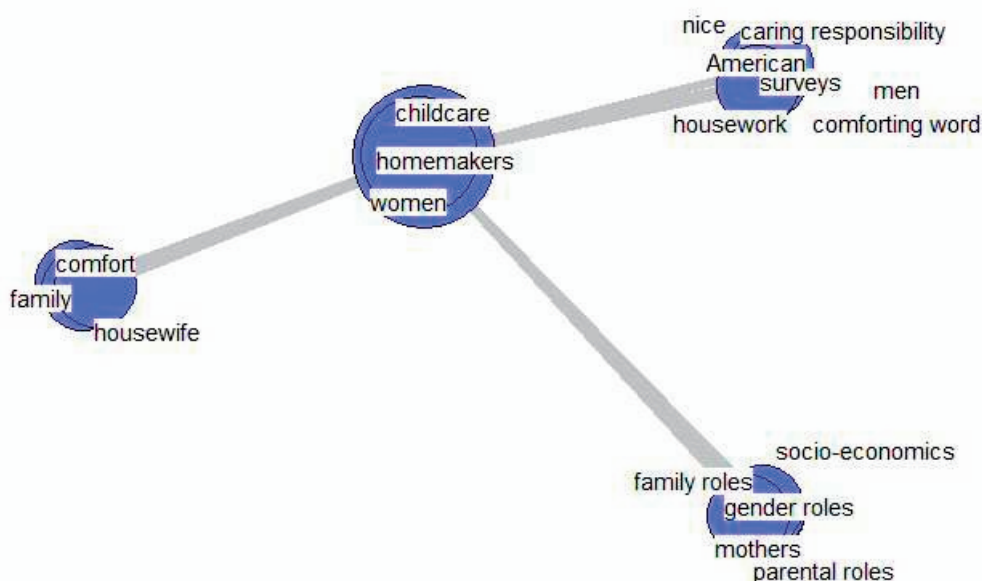
Figure 38: *Word associations for "homemakers" by British indexers*

It associated clearly as a woman taking care of children within a family. In this case the consistency was greater at the conceptual and at the theme level rather than at the term level (cf. Iivonen 1989). In principle it was considered to be a gender-non-specific term, and the connotations were positive (example 1 below) and it was also discussed as a metaphor (example 2).

Example 1:

> "In the old surveys about up until 1980's the surveys would use the term housewife. Homemakers is not gender specific, but it's very much an American word, I think." (BR IN1)

Example 2:

> "The person who is mainly in charge of looking after the home, seen that it is clean and things like decorating and who will design… and I suppose cooking, perhaps. But I think outside the family context as well I wouldn't necessarily see a homemaker as being good, just within a family. I would see myself as being quite a homemaker and single people to whom homelife is important can be homemakers." (BR IN2)

The main distinctive components discussed were marital status and motherhood – "homemakers are not necessarily wives or mothers or…" (BR IN2).

The stimulus word ***housewives*** resulted in 19 co-occurrences, out of which 15 were different. – As in the previous cases studied, the consistency at the theme level was greater than that at the term level (cf. Iivonen 1989), and e.g. employment was covered with different kinds of words (*employment of women, unemployment, part-time employment*). The connotation of *housewife* was neutral, if not slightly negative. *Housewife* was clearly understood as a woman in charge of housework at her home, and not belonging to full-time employment.
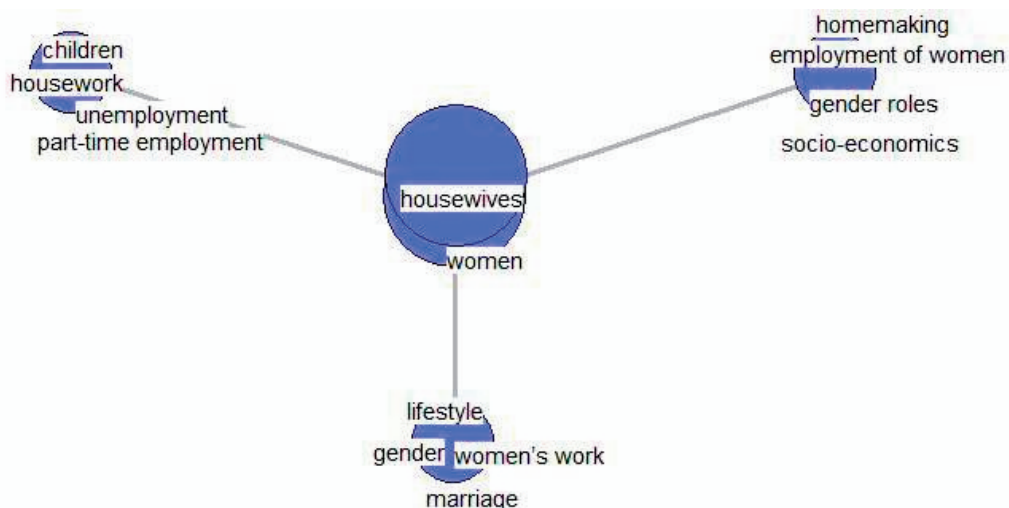
Figure 39: *Word associations for "housewives" by British indexers*

When discussing the meaning and usability of *housewife* it was constantly compared with the previous terminological case *homemaker*, resulting in more positive connotations concerning the term *homemaker*.

> "Until very recently I wouldn't really made any distinctions [between housewives and homemakers], because twenty years ago homemakers and housewives were the same thing. I think it's only in the last twenty years that it's been recognised that sometimes it suits a couple or the man to be the one who is the homemaker. I still think it's quite rare and I think what's more likely to happen is that a couple shares it more. I still think it's quite rare that men actually take the primary responsibility, but it is a possibility now. In fact I don't think that the term housewife is really used very much now and I can't see many modern women being very happy to be called as housewives, because a whole image goes with that, which most modern women wouldn't want to subscribe to. They might well say that well in this particular

circumstance I'm going to be one who stays at home, but they wouldn't want the longer term, that word housewife portrays." [*So would they call themselves homemakers or homemothers or stay-at-homemothers if they would today stay at home?*] I haven't really thought of that, but I would think they would think themselves as homemakers rather than as housewives." (BR IN3)

When *househusband* was brought up, it was considered to be an English word in various ways (see examples 1 and 2 below) - usually and especially when compared to *housewives* its content was considered positive, but it was not necessarily a positive word. (However, in the context of the simulated indexing task it was used frequently.)

Example 1:

"--- would use it [househusband] self-consciously, would not come naturally as housewife. It is very rare to find a husband who stays at home. --- The term itself is pretty widely used and certainly recognised. And I think in a sense it's a better term than housewife, because if someone is a househusband it's a positive choice because it's against the norm." (BR IN2)

Example 2:

[*What about this househusband, is it used?*] "Yes, I think it's used, because it does describe the situation very well." (BR IN3)

8.2.1.4 Simulated indexing task

8.2.1.4.1 Background

The British indexers were asked to index five articles, published in English, and written by different authors representing different kinds of users of English – natives and non-natives, men and women, written alone or in co-operation. The articles used were all social science papers and related to care giving issue within a family (i.e. different approaches of family sociology) in order to limit the topic and lexicology used. The approaches were still varied both in terms of research questions given in the paper, material and methods used etc. (See also chapter 6.3.5.1 *Indexed documents* and appendix 7.)

The used articles were:

1. Arnlaug Leira 2002: Updating the "gender contract"? Childcare reforms in the Nordic countries in the 1990s' NORA no. 2 2002, Volume 10, pp. 81-89.

2. Bridges Judith S. & Etaugh Claire & Barnes-Farrell Janet 2002: *Trait judgments of stay-at-home and employed parents: A function of social role and/or shifting standards?* Psychology of Women Quarterly, 26:2 (2002), pp. 140-150.

3. Nock Steven L. 2001: *The Marriages of Equally Dependent Spouses*. Journal of Family Issues, Vol. 22 No.6, September 2001, pp. 755-775.

4. Jalovaara Marika 2002: *Socioeconomic Differentials in Divorce Risk by Duration of Marriage*. Demographic Research, Volume 7, Article 16, Published 29 November 2002, URL: www.demographic-research.org, pp. 537-564.

5. Smith Calvin D. 1998: *Men Don't Do This Sort of Thing. A Case Study of the Social Isolation of Househusbands*. Men and Masculinities, Vol. 1 No. 2, October 1998, pp. 138-172.

It is notheworthy that the fifth article included keywords[83], but they were removed from the versions which were sent to the British and Finnish indexers. In the analysis of simulated indexing task different spelling types were modified according to the most commonly occurring (e.g. *Child care* and *child care* were replaced with the most common form *childcare* and *USA* with *United States*) and in cases when it was not considered to cause conceptual changes the singular versus plural form was selected according to the more commonly used (*social role* replaced with *social roles*), but no other harmonisation was made. The procedure was the same with the Finnish samples.

Together with the articles the respondents were briefly informed about the task:

"The simulated indexing task consists of two phases:

1. Please, index the articles using 9-12 descriptors, which describe what the article is about. Try to do it as it would be a part of your everyday job. Please use the sources and principles etc as you normally use and follow in indexing.

2. Please, index the same articles using again 9-12 descriptors. This time try to do it as if the end users were European social scientists and you were indexing for the European social science database.

In both phases please write down the terminology sources (like thesauri) if such are used." (Appendix 4)

8.2.1.4.2 On process

The British respondents experienced the task as interesting, and the more the respondent had experience (education) in the social sciences, the more positive was the attitude towards the task.

"Quite interesting, probably more interesting to me since I'm a social scientist, than to others." (BR IN1)

The respondents experienced the articles and selection of descriptors as unique processes (example 1), although they were considered to also form a theme and practice (example 2). They also noticed the different cultural backgrounds of the articles and this caused some feelings of insecurity (example 3).

Example 1:

---

[83] fatherhood, househusbands, masculinity, motherhood, domestic work, housework, child rearing, children, gender

"There was one, which was a little bit difficult. ---demographic… dry, other more interesting. Househusband – interesting, recognises experiences, which had heard from relatives and friends. --- For some the terms came very quickly and with others you had to think harder." (BR IN1)

Example 2:

"They were not difficult to index. --- When you're doing indexing to a journal and the articles form a kind of a theme as you're going through, something you didn't think of perhaps with the first article occurs you in the second one that whether that would have been a relevant one so it's that kind of thing. But by the time you get to the last one you've already kind of built up a more of a vocabulary in your head to deal with those kinds of articles. But I didn't find that there was any particularly difficult. For some it was difficult to looking up the - to find keywords for the methodology of it, things like data analysis, which is quite generic…" (BR IN2)

Example 3:

[*Do you think any of these articles were somehow different from others, more difficult or do you recall that any was somehow…*]

"I think that one [2. Bridges Judith S. & Etaugh Claire & Barnes-Farrell Janet 2002: *Trait judgments of stay-at-home and employed parents: A function of social role and/or shifting standards?*] I found particularly difficult. And I wondered whether it was very much from an American perspective. I wondered whether the terminology of trait judgements I have never come across… Trait used in that sort of way. So I think it was probably because it was within an American bias, that I found that one quite tricky. They've talked a lot of models, in this one, and I wasn't comfortable with that, because I didn't know whether the, I mean they had shifting standards models and role based models, and I didn't know whether shifting standards was a commonly used model, so I indexed it under models, shifting standards, rather than shifting standard models, for example, whereas if I knew more about it and I was confident that shifting standard models was a, you know, internationally used term, then I'd put that." (BR IN3)

The use of **vocabulary control** and sources varied as well as **domestic versus European practices**. One respondent used the same organisation-specific thesaurus in both phases, but with variation according to the focus of the task (example 1). The other respondent also used an organisation-specific thesaurus, but did not make any distinction between the two phases, since they considered the normal user group to be international. One respondent did not use any controlled vocabulary, but aimed at constructing a coherent terminology for all the five articles (example 2), and in two cases it was considered that there would be some differences if the indexing was for British users as opposed to European users.

[*What tools did you use*?]

Example 1:

"This is the sort of thing that we would have done ordinary work-wise and that was using [organisation-spesific thesaurus]. And with the European social science database I tried to put in a higher terms to describe the discipline as well, because I though giving the big database that's the first sort of thing people are going to search for. But I mean it's going to be obvious that the choices are still influenced by [the

organisation-spesific thesaurus], because that is what I use and which I know. And also I think [the organisation-spesific thesaurus] is a good tool to describe these concepts as well. The lists are very similar." (BR IN1)

Example 2:

"I didn't actually use any thesauri, because I don't use them for my current job, and I didn't have access to any that I thought would have been useful, the government category list, for example, wouldn't be useful to this. So I really picked terms out of the articles and then to some extend tried to use consistent terminology across the five. That's the sort of approach that I would normally take. ---

In some cases I thought there could be quite a bit of difference between doing it for sort of normal purposes and doing it within a European context, but in others I couldn't see that I would actually do it differently. --- They are mainly terms taken from the articles. They might not have been if I'd have been using a thesaurus." (BR IN3)

In phase two the perspective was thus very much similar to the first phase, and the British practices were considered almost identical to the European.

The number of descriptors (guideline 9-12) was considered to be reasonable by one informant (example 1), but more often considered as slightly more than that usually given (example 2 and 3).

Example 1:

"I managed to keep it between nine and twelve, but it's because I'm used to do this sort of thing." (BR IN1)

Example 2:

"9-12 was little more than would normally use. Depends partly on the length of the article and partly on the complexity. Eight is usually the average, has become a habit." (BR IN2)

Example 3:

"It is probably more than I would usually select for an article. I found it difficult to get beyond nine. I probably usually go for about five to eight, as a maximum. But it depends so much on the type of the material---" (BR IN3)

8.2.1.4.3 Updating the "gender contract"? Childcare reforms in the Nordic countries in the 1990s'

In phase 1 the British indexers provided 31 descriptors, out of which 26 were different. On average the article had 10.3 descriptors, and the terms used more than once were:

NO	Descriptors
3	childcare
2	families
2	gender roles
2	Scandinavia

Table 25: *Article 1, the most used descriptors by the British indexers in phase 1*

The descriptors used referred or were related to childcare and/or parenthood (childcare, children, fathers, maternity leave, mothers, parental leave, paternity leave, tot. 9), social systems and social benefits (family policy, social policy, social welfare, welfare reform, welfare state, tot. 5), working life (employment (of mothers), women's employment, work, working parents, work-life balance, tot. 5), family and family life (families, family roles, models (of families), tot. 4), genderism (gender roles, gender, gender contract, tot. 4), or to geographical place or time (Scandinavia, Sweden, 2002, tot. 4). There occurred a considerable amount of terminological and conceptual differences (cf. Iivonen & Kivimäki 1998), but the indexing was coherent at a general theme level (cf. Iivonen 1989).

Two of the indexers used descriptors referring to working mothers (*employment (of mothers), women's employment*), but otherwise the article was indexed in a equal way.

Phase 2 resulted in some changes. One indexer (BR IN1) provided one descriptor less (*work-life balance* removed) than in phase 1 and additionally replaced one with a more general one (*gender roles* with *gender*). The third respondent made several changes. *Scandinavia* was replaced with narrower terms *Denmark, Norway, Sweden* and *Finland*, whereas *maternity leave* and *paternity leave* were covered in the second phase with *parental leave* and *employment (of mothers)* with *employment*. *Models (of families)* was removed completely.

As a strategy the changes done (and undone) are similar to internationalisation (cf. 3.4.1.3 *Domestication or foreignisation – or internationalisation*; Venuti 1998; Lindfros 2001) which was also the case with respondent 2, who saw no differences between the phases. - Some abstract concepts were replaced with more general ones and concrete with narrower ones.

8.2.1.4.4 Trait judgments of stay-at-home and employed parents: A function of social role and/or shifting standards?

In phase 1 the depth (in numbers) was identical with phase 1 of the article 1, which is 31 descriptors, out of which 26 were different. The commonly used descriptors were:

NO	Descriptors
2	fathers
2	parenting
2	social roles
2	United States
2	working mothers

Table 26: *Article 2, most used descriptors by the British indexers in phase 1*

Most of the used descriptors referred to parents or parenting (fathers, parenting, working mothers, mothers, parent roles, parent-child relations, parents, working parents, tot. 11 occurrences) and/or roles or stereotypes (social roles, family roles, gender roles, gender stereotypes, "models, role based", parent roles, role perception, social role theory, stereotypes, tot. 10). Only two descriptors about gender issues were given (gender, gender role). Descriptors referring to research design were used by one indexer, and those were data analysis, experiments. Two descriptors referred to time or place (2002, United States). In all 5 descriptors were more general or rather broad (children, judgements, perception of others, trait ratings, "models, shifting standards").

The gendered descriptors provided referred to both men and women, but women were slightly more emphasised (*fathers* 2 occurrences versus *working mothers* one occurrence + *mother* 1 occurrence). Mothers were also often expressed more precisely with a narrower term (*working mothers*).

In phase 2 only one respondent (BR IN1) made changes. *Gender roles* was removed, and *psychology* and *gender* were added. The direction was into a more neutral and general direction. Two other respondents considered their indexing to already represent also European perspective. The strategy is thus considered as internationalisation.

8.2.1.4.5 The Marriages of Equally Dependent Spouses

In phase 1 the third article resulted in 30 descriptors, on average 10 per article. Out of these 25 were different, and only three descriptors were used by more than one indexer:

NO	Descriptors
3	divorce
3	marriage
2	Economic dependence

Table 27: *Article 3, most used descriptors by the British indexers in phase 1*

The used descriptors referred to or were related to economics (economic dependence, economic independence, economic status, employment, of women, household income, income, tot. 8 occurrences), to family members of family life (divorce, marriage, marital relationships, married persons, spouses, spouses' occupations, children, tot. 10 occurrences), to gender and equality issues (equality between men and women, gender relations, gender roles, working women, 4 occurrences), demography or research design (demographic change, demography, dependence relationships, sociological analysis, trait ratings, tot. 5), time or place (United states, 2001, tot. 2), or were in their nature more psychological or general (judgements, models, shifting standards, perception of others, tot. 3).

The indexing was done in a gender neutral way.

Phase 2 resulted in minor changes, with only one respondent (BR IN1) adding one general term (*families*). The indexing was thus seen as already international in phase 1.

8.2.1.4.6 Socioeconomic Differentials in Divorce Risk by Duration of Marriage

In phase 1 the fourth article resulted in similar number of descriptors as the previously discussed samples, which is 31 descriptors, out of which 26 were different. Three descriptors were used more than by one indexer:

NO	Descriptors
3	divorce
3	Finland
2	marriage

Table 28: *Article 4, most used descriptors by the British indexers in phase 1*

The descriptors given were about demography or research design (data analysis, demography, factor analysis, sociological analysis, statistics, tot. 5 occurrences), issues

related to socioeconomics (economic independence, income, social class, socioeconomic status, socioeconomics, unemployment, tot. 7), education (educational attainments, educational background, level of education, occupational class, spouses' occupation, tot. 4), marital relations (divorce, marriage, marital duration, spouses, tot. 8), place or time (Finland, United States, 2002, tot. 5) or was in its nature general (time, risk, 2 occurrences).

As in the case of the previously discussed article number 3, the indexing in this case was done in a gender neutral way.

In phase 2, two respondents made changes. One respondent (BR IN1) added one general term, *demography,* whereas the other (BR IN3) removed two specific ones which were given as near-synonymous terms in phase 1 (*marital duration*, *educational attainments*) and a general term (*socioeconomics*), and added one general term referring to a place name (*Western Europe*). The strategy used was thus closest to internationalisation, which do not favour national characteristics.

8.2.1.4.7 Men Don't Do This Sort of Thing. A Case Study of the Social Isolation of Househusbands.

In phase 1, the indexers provided 31 descriptors, out of which 21 were different. Seven descriptors were provided by more than one indexer, which is remarkably more than in the previous samples:

NO	Descriptors
3	child care
3	masculinity
2	alienation
2	gender roles
2	housework
2	men
2	social isolation

Table 29: *Article 5, most used descriptors by the British indexers in phase 1*

The sample was otherwise also more coherent, and there were no need to harmonise different spelling types between different indexers (e.g. *child care* was, in this sample, presented in only one writing style, cf. also inter-indexer inconsistency, Meriläinen 1985; Iivonen 1989).

The themes covered were men and/or gender issues (masculinity, men, gender, gender stereotypes, house husbands, husbands, working women, tot. 9), psychological aspects (*alienation, social isolation, self-perception, social identity, social interaction*, tot. 7), childcare and domestic responsibilities (*child care, housework, nurture, of*

children, tot. 6 occurrences), roles or stereotypes (*gender roles, family roles, men's role*, tot. 4), or referred to place or time (*1998, Australia*, tot. 2 occurrences).

The article was about men, masculinity and childcare, and the descriptors provided for men were clearly masculine related, and thus gender-neutral terms referring to persons, such as e.g. *homemakers*, were not used.

Phase 2, resulted in minor changes, with one indexer (BR IN1) adding one term referring to methods (*interviews*). The indexing strategy was thus already considered to have been international in phase 1.

8.2.1.5 Equivalence and translation strategies

The British indexers were asked directly what they though equivalence to be in a multilingual thesaurus construction – what is meant by equivalence and what it aims at. As with the other British groups studied they were not asked to translate anything.

The aspects they discussed are related to Koller's (1989) equivalence types such as denotative, pragmatic and formal equivalence (cf. ibid. and chapter 3.4.1.2 *The idea of equivalence*). Thus connotative and text-normative aspects of equivalence were not brought up.

When speaking about the **denotative level** of the words the respondents made a clear distinction between concept versus term equivalence, and stated the difference to occur also within one language representing different cultures (example 1 below). The denotative and conceptual aspects were constantly tied to culture (examples 1 and 2 below), and dictionaries were not considered as reliable sources (example 2 below).

Example 1:

"--- people understand different things through language and so it's, it would be important to know a lot more about it than I do. If you are constructing a multilingual thesaurus you'd have to have an understanding from the perspective of the languages and what is meant by that term and just translating the actual words might not give you that understanding. And that's even the case with legal terminology, where the same word can be used in Britain and in America, but mean slightly different things." (BR IN3)

Example 2:

"If you translate from one language to another, sometimes it's not quite the same. And I think you have to really bring a cultural aspect into it. I mean a concept, which is like a saying or common phrase in Britain, might mean something completely different in German or they might have - and again - a German common usage phrase might be very different in English. So I would say you have to yes do the basic translation and you have to be aware of the cultural differences. So really rather than just someone looking up it from the dictionary you really need somebody with the decent knowledge of the language." (BR IN1)

Pragmatic equivalence (cf. Koller 1989) was considered in the sense that users' expectations and the usability of translations were considered to be essential. As one (BR IN2) respondent stated: "You are aiming at to find a consistency, I suppose. So that people know that when they use a certain word everyone is sure what they are talking about. --- multilingual thesaurus could provide a kind off consistency and surety."

The respondents represented their views of **dynamic equivalence** (cf. Nida 1964/2000; Vehmas-Lehto 1999 and chapter 3.4.2.1 *Dynamic Equivalence Theory*), and accordingly to these views a "good" or proper translation should cause the same reaction in the target audience as the source text does in the original context, and thus instead of asking "is the translation correct", we should ask "for whom". They also represented closest domestication strategy (cf. Venuti 1998; Lindfors 2001 and chapter 3.4.1.3 *Domestication or foreignisation – or internationalisation*) when aiming at translations that reflected the target culture.

In the simulated indexing task the translation strategy (shift from domestic audience to European audience) was similar to internationalisation, in which different European national characteristics were not added, but the changes in abstract matters were moved to a more general level, and the concrete matters to a more specific level. One of the three organisations, which the indexers represented, also has a clearly international audience, but the two other are national in their primary function. The domestic practices were thus not considered as being biased when used by wider audiences (cf. Venuti 1995 and 1998). It is also noteworthy that the material indexed was rather international, and thus the terminology used in phase 1 did not necessarily reflect British practices. The lack of changes may also be due to the language used and the discourse type - English as a language is rather international (cf. Katan 2004) and the aim of documentary languages is to increase harmonisation.

- **Summary and Conclusions**

The **British indexers** considered knowledge of the subject domain to be prior to language knowledge. They recognised different discourses, and often together with different discourses (cf. Buckland 1999; Lehtonen 1994; Venuti 1998; Vehmas-Lehto 1999) the time aspect was raised (cf. Aitchison 1991; Suojanen 1993; Hirsjärvi and Hurme 2000). Different discourses and time aspects were seen as a construction problem as well as an update problem, especially in the context of gender-neutral language use (cf. UCL 2000; Brunel 1999; School of Psychology 2002; University of Southampton 1997; Engelberg 1993, 1998).

In all the cases there occurred a considerable number of terminological and conceptual differences (cf. Iivonen & Kivimäki 1998), and the indexing was remarkably more coherent at a general theme level (cf. Iivonen 1989). Coherence among the indexers also varied.

The respondents often made (in practice) a clear difference between **different types of meaning** and especially between semantic and pragmatic meaning (cf. Koller

1989; Vehmas-Lehto 1999) when discussing the family roles terminology. The studied concept, term, and the descriptor *family roles* were generally considered more broadly than in the thesauri studied (cf. chapter 9.1.1) and also included psychological roles. *Breadwinners* was considered to be out-dated and sexist as a common language concept and term, but in a more positive sense if evaluated as descriptor. *Heads of household* resulted in similar reactions to *breadwinners*, and it was considered to also need modern equivalents within a thesaurus context. *Homemakers* resulted in more positive reactions than *housewives*. The semantic and linguistic context of the two last mentioned was close to those of thesauri (cf. chapter 9.1.3), and in general they were understood rather similarly with the other discourses studied, but from a greater perspective (different time versions, modernisation, psychological thematic aspects).

In the **simulated indexing task** the indexing was considered rather international (or at least European) already in phase 1, when the indexers were asked to index certain documents for their normal user groups (cf. Adler 1997; Hall 1981). One informant considered the indexing as totally equivalent. The changes made by the other two in phase 2 were minor, and mostly towards internationalisation in which national characteristics are diminished and harmonised (cf. Koskinen 2000).

There was a gap between the theoretical and practical discussions. Discussion at the more principal level included aspects related to Koller's (1989) equivalence types such as denotative, pragmatic and formal equivalence (cf. ibid. and chapter 3.4.1.2 *The idea of equivalence*), and the connotative and text-normative aspects of equivalence were not brought up as when discussing with the terminological case. The idea of equivalence of the respondents was similar to **dynamic equivalence** (cf. Nida 1964/2000; Vehmas-Lehto 1999 and chapter 3.4.2.1 *Dynamic Equivalence Theory*). They also followed closest **domestication** strategy (cf. Venuti 1998; Lindfors 2001 and chapter 3.4.1.3 *Domestication or foreignisation – or internationalisation*) when aiming at translations to reflect the target culture, but in practice (cf. simulated indexing task situation) preferred **internationalisation** (cf. Koskinen 2000).

8.2.2 Finnish indexers

This chapter deals with how the Finnish indexers understood the terminological case (family roles terms) and translated them into Finnish, equivalence in multilingual and multicultural thesaurus construction, as well as Finnish indexing practices and results of the simulated indexing task.

The interview process is similar to the British indexer respondents except translations asked. First, the background of the indexer is discussed, then the indexing practices (material, languages, tools, principles), followed by equivalence, and the terminological case family roles (as concepts, sub-concepts, descriptors, word associations, translations into Finnish, definitions), and then finally the simulated indexing task (see appendix 3 *Interview guide for the Finnish indexers*).

8.2.2.1Background

Four Finnish indexers known to have experience in indexing of social science material were contacted by e-mail. One agreed immediately, two after a reminder and one did not want to participate in the study.

The three Finnish indexers interviewed are all female and they represent three different institutions located in two cities. Two informants represent university libraries and one informant a special information institution. All three are specialised in indexing social science material.

The indexed material was sent to the informants from one to two weeks before the interview took place. The interviews were conducted during November 2003 - May 2004. The interviews took place at the informants' work place and they lasted on average for 90 minutes (minimum 75 and maximum 105 minutes). As with the other groups studied, notes were written during the interview, the interviews were tape recorded, and later transcribed.

8.2.2.2 Indexing practices

In their everyday work the respondents needed to know (in addition to Finnish) the Swedish and English **languages** and one person also used German and Spanish. In their work the indexers indexed (i.e. gave keywords) in Finnish, and in one case also in English. The **material** indexed was most commonly Finnish or English, but often also Swedish. One person also indexed German and Spanish documents. The indexed material was mainly research reports and surveys, digital research data, articles and books. In two cases, articles and foreign language material other than English had decreased during the past few years, and this tendency was experienced as unwelcome.

University lectures, especially within sociology, were considered a good resource when keeping in touch with the current development in the field of the indexed material, and similar to the British indexers the Finnish indexers also emphasised the importance of good knowledge in the social sciences.

In the university libraries the **followed guidelines** were unified and "nationwide". The respondents, however, raised the fact that despite these congruent principles there is considerable variation in practice (e.g. use of linked descriptors, they "use already made linked descriptors, but do not create new ones") (see example 1 below). In one case the guidelines were organisation-specific, and mediated in in-house training (example 2).

Example 1:

"We use YSA but not any free words, and we made a decision two years ago that we don't pre-coordinate descriptors anymore, so it is the information seeker who then combines the single terms. And then… We don't actually have any written guidelines,

but these two [rules] we have: YSA and no pre-coordination. Well then we try if only possible to find geographical descriptors, and if possible to find, for example, research methods are mentioned in indexing, target persons and communities are recorded… (FI IN3)

Example 2:

--- our information specialist has organised broadly this kind of education that if the [indexed material] are reports so first and related to the material we do browsing, look at the titles and tables and underlying and… So first we do preliminary browsing, and then we go through the central parts, and then we refine the concepts." (FI IN2)

The Finnish indexers often aimed more robustly for flexibility at the principle level, rather than for more specific terminology. They considered it important to allow both pre-coordinated and post-coordinated indexing (example 1, cf. Gilchrist 1971; Miller 2003a; Hjørland 2006). YSA was considered sometimes to be too general, and therefore free terms were needed and should be allowed to be used in the indexing of social science material (example 2).

Example 1:

"--- quite well one finds in YSA and VESA [web-version of YSA] at the general level, but [the boss] has told that soon also pre-coordinated indexing will be again allowed, and that has been missing, the possibility to combine. For example, "uusintaminen" [reform], it may mean so many things and in such cases one misses the possibility for using phrases and compounds." (FI IN2)

Example 2:

"In the indexing of research literature we use merely YSA, so no any kinds of free terms are allowed. But then in ARTO [Reference Database of Finnish Articles] it is possible to use more freely also own words, and thus use also terms not included in YSA – and it is like a detective work to find them." (FI IN3)

The **tools** used depended on the database, but those most frequently used and superior in its vocabulary control is YSA. In addition, foreign library union catalogues were used, Google search, other controlled vocabularies (Finnish and foreign), dictionaries, writers of the articles and/or other social scientists, vocabulary lists provided by university departments on their web-pages etc. In the cases where YSA was considered the highest terminological authority, conceptual and terminological help was also sought from other sources, such as the British Union University catalogues. Foreign indexing thus also influenced how foreign material is indexed in the Finnish databases.

"If one thinks how I do the work, so if it is not found in any Finnish library, so then I take a look at COPAC, the British union catalogue, and then at LIBRIS [The Swedish university and research libraries search service] and then sometimes --- and try to find possible descriptors and then to get an idea how it could be expressed by using YSA and those Finnish [descriptors] if there happens to exists [suitable ones]. Or then one has to get the result by somehow going round it… So we use work done by others a lot, which is of course reasonable. Some vocabularies I have found from the Internet, like for example the vocabulary of department of environmental policy, and in Stanford the department of philosophy ---And then in ARTO there are of course totally

new journals so there are of course about that kind of matters which are not yet in any vocabularies ---

And then one can use the researcher himself/herself as a source, and ask what do you mean by this and then one is working by e-mail, which is pretty handy. --- And then I have used Google and filled in the search term and looked what comes up, and sometimes there are also proper sources, like the vocabulary of Stanford [department of philosophy] --- (FI IN3)

As already stated in chapter 1.1 *Background*, each information search in a database covers at least five different languages: the authors, the indexers, the synthetic structure, the users and the search strategy (Buckland 1999), which all represent a type of **discourse**. The indexers interviewed emphasised the information seekers' perspective and search practices, and stated it to be in their mind when indexing – and already present in the in-house education (example 1). However, the discourse of thesauri was considered in conflict situations to be superior (example 2).

Example 1:

"It has been emphasised in the education to take the clients perspective into account, how the material is retrieved later. It's truly an important matter, that one aims to highlight also those matters, which are then later used for finding the information." (FI IN2)

Example 2:

"--- But with social science theories one --- but one does not really find them [terms used in the indexed book] [in thesaurus] and then one is forced to use broader concepts or something and cannot index as precisely as one would like to." (FI IN3)

Indexing inconsistency was considered an unavoidable reality, but its negative results were aimed to be minimised by using thesauri and providing different levels of specificity. The same obstacles in the form of different discourses were thus considered to be present both at the indexing and at the information retrieval phase, but it is the indexers who are liable for successful information retrieval.

"It is very much depending on the indexing whether the user finds the right information or not. --- The indexer need to have a broad, all-round education and to know what is going on in the world. --- It is often thought [by seekers] that it is enough to know the title and the author of the document. But before that one [- seeker] has to know the title and the author." (FI IN1)

The indexers received very little **feedback** on their work**.** It was commonly considered that direct involvement with end-users at e.g. an information desk, is important for the indexers in order to be familiar with the different ways of expressing information needs (i.e. also discourses). One respondent (FI IN3) also mentioned a new type of collegial support in the form of regular meetings, in which difficult cases are discussed, and in the context of these meetings it had also been stressed that thesauri should be more informative concerning the meanings of descriptors.

"We have indexers' meeting ---- and they have been experienced as very useful and good way, because otherwise this is a very lonely work ---- And if one goes back to

these thesaurus correspondences so yes, they should provide more explanations about what is meant with this." (FI IN3)

The **usefulness of controlled vocabularies** was explained with indexing inconsistency, and even more so with different discourses (cf. Iivonen 1989; Lehtonen 1994; Venuti 1998; Vehmas-Lehto 1999) and time aspect (cf. Aitchison 1991; Suojanen 1993; Hirsjärvi and Hurme 2000). Although thesauri were considered to be useful, current Finnish practices were also criticised for not providing sufficient information and possibilities.

> "I think that a cathecism like this is absolutely needed - what does this term mean, and YSA is in that sense very very brief, but… When one knows that people may understand the same thing in so many ways and, for example, some scientific terms may have been understood in different times very differently, so the content of certain time periods would be very good to discover." (FI IN3)

The future of thesauri was seen in the sense that there is an increasing need but that they were also being asked to answer the current needs in connection to the challenges of **multilingualism** – already in the context of one geographical culture as Finland.

> "In libraries the common experience is that YSA is not enough. Already in the beginning [when YSA was being constructed] the idea was that it would be most convenient that there would be another one in English. For the users the Finnish YSA is not enough. When copying what others have done one gets also English descriptors. The more multilingual the better. But it may be very difficult to achieve equivalence between concepts, since they are so much bound to the culture. --- In the future human indexing becomes more and more important. Automatic indexing does not end the need for human indexing. The amount of information is increasing all the time and therefore the selection becomes harder and harder. ---" (FI IN1)

8.2.2.3 Family Roles terms by Finnish indexers

8.2.2.3.1 The term family roles as regarded by Finnish indexers

As with the case of the British indexers (cf. chapter 8.2.1.3.1) it is noteworthy that the first introduction to the terms had already occurred before the interview was conducted, in the simulated indexing task (the articles of the task were about family sociology and care-giving issues, cf. chapter 6.3.5.1 *Indexed documents*). The task was sometimes directly referred to when discussing the terminological matters.

> [What is meant with family roles?]

> "Well, as a first thought of it is roles between men and women in a family, who takes responsibility of what, and is the work division in a family, for example, more traditional, so called men's and women's duties, or… The articles [in the simulated

indexing task] were very interesting; there are so different national and international practices in different countries." (FI IN2)

The stimulus word *family roles* resulted in 18 co-occurrences and 13 unique associations. (They were all given in Finnish.) *Family roles* associated clearly with the idea of a nuclear family, consisting of parents and children. Similar to the other cases studied, consistency was greater at the concept level and especially at a broader level (e.g. *mother*, *father* versus *parenthood*, *spouses*).
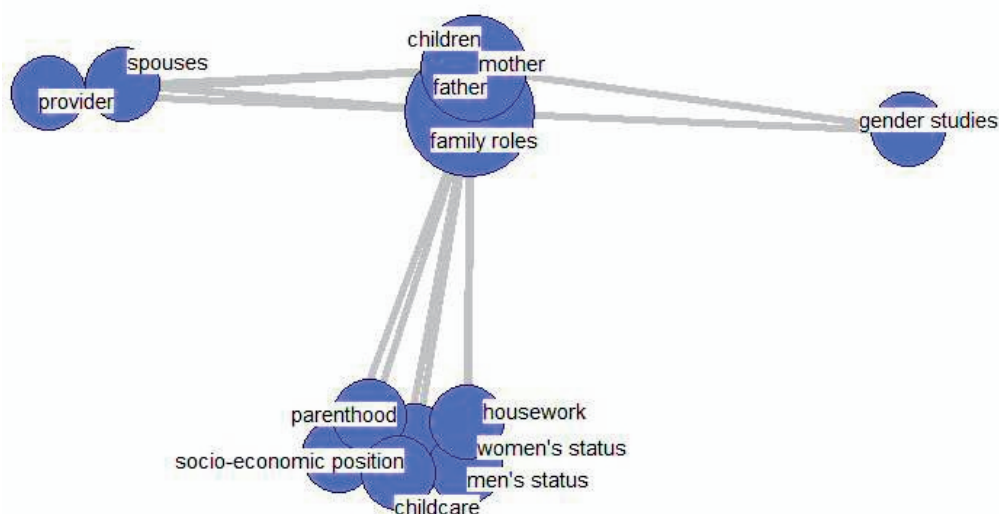
Figure 40: *Word associations for "family roles" by Finnish indexers*

The Finnish indexers more often approached *family roles* from the perspective of adults and in general considered the roles to be based on homemaking and on financial decision-making (example 1 below). (In the articles, the perspective was also from the adults, children's roles were not discussed.) When the discussion was more broad and included, for example, psychological aspects, were children also emphasised (example 2).

[What is meant with family roles?]

Example 1:

[About family roles as concept, respondent refers first to articles used in simulated indexing task and how they are understood in them.] How family members understand their roles within the family, but of course also how elsewhere. --- They are depended on how family members see it – what are their tasks within the family. --- In traditional thinking, well, they were clearer before. Today men's and women's images are closer to each other, due to changes in attitudes and practices. ---

Family roles are seen in who does the housework and who makes financial decisions, for example, about buying more costly acquisitions." (FI IN1)

Example 2:

"It is kind of division of work, that to whom in the family matters fall, or belong to. But it is as much also roles of interaction, so that somebody has the status of a child

and somebody has the status of a parent, and… Then there is naturally also that side, well I myself am from a family with several children, so who in the family is the family clown and who carries the weight of the world on one's shoulders and who is the oldest sister and who is the youngest child, and like these, and today when there are a lot of these new married families so who is the step-sister and who is the step-granny and so on. And then there [in the articles] was about those who bring home money, so this kind of division of responsibility there is also." (FI IN3)

Family roles was translated as *perheroolit* (direct literal translation). The translation was not considered to be a problem by the respondents.

As a topic it was considered important to be covered in thesauri. As a field of study it was discussed as cross-disciplinary, which within the social sciences and also in general was considered to be natural - "we live in a cross-disciplinary world" (BR IN1). Similar to the other groups studied, the roles were considered to have changed a lot during the past decades, and family roles were considered to depend more on the life situation rather than being gender-specific.

"Certainly everybody of our age [early middle age] thinks that the own family roles are pretty different from those our parents had, for example. At least I don't see in our family roles me being more of a care-taking person and my husband more kind of an actor. So at least it doesn't go like that in our family, we are equal partners, we both contribute accordingly to who happens to be around and available. But of course the tasks to do vary every now and then, so sometimes the other may be in a weaker role in something and the other is stronger, so it depends more on life situation than on gender." (FI IN2)

In Finnish, the sub-concepts would be named after mother, father, spouses, parents, grandparents, children and relatives. Naming in accordance with biological family roles is a practice very similar to thesaurus discourse (cf. later in chapter 9.1.1 *Family roles in thesauri*). When dividing into sub-concepts and sub-terms the respondents did not make a distinction between denotative (i.e. linguistic and semantic) versus pragmatic meanings (cf. Koller 1989; Vehmas-Lehto 1999), but the linguistic and semantic aspects were in practice prior to pragmatics.

The sub-concepts and terms were considered to also fit in with European practices, although the European context was, at the same time, not considered to be as gender-neutral as the Finnish practices (with reference to e.g. children's daycare system, see example below) and the family is more generally understood in broader terms than in Finland.

"[In an international thesaurus I found] one very good, very splendid equivalent. I looked for connecting father and man to family concepts, and very little was on offer. But of course it is emphasised in these English language versions, both in the America and in the Great Britain, that the woman has traditionally greater care-giving responsibility and the man is the family provider. For instance, I tried to look for a concept for these, dual earners, but didn't find in any way that kind of, actually there is not in Finnish (thesaurus discourse) an equivalent for them." (FI IN2)

As stated before (and also illustrated later) the time aspect, the cultural differences and the need for active updating of controlled vocabularies were discussed a great deal

by the respondents. In the case of *family roles* it was considered important to conform to modern practices, especially in marital status issues and in the common dual-earner model currently used (see example above).

8.2.2.3.2 Breadwinners and Heads of household by Finnish indexers

As a stimulus word **breadwinners** resulted in 23 co-occurrences, out of which 17 were unique.
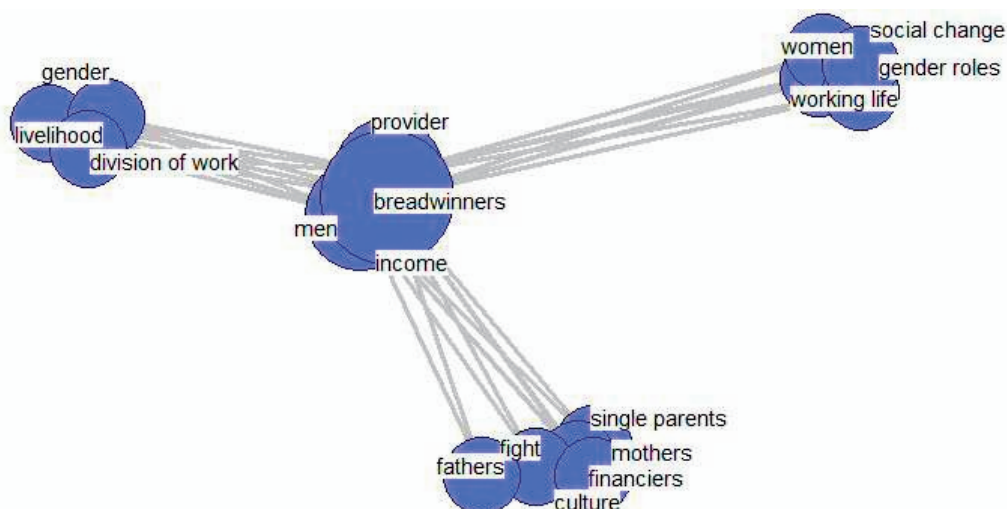
Figure 41: *Word associations for "breadwinners" by Finnish indexers*

It was translated into Finnish as *perheenelättäjä* (economic provider in the family) and as *ansiotulon saaja* (descriptive phrase, person with income).

> "Could it be "perheenelättäjä" or… Well, we don't truly have this kind of word. It is actually exactly in matters like this when indexing is terribly difficult. In spite of all one should express the reference [i.e. meaning], but there are no words to describe it…" (FI IN3)

It associated as a male (example 1 below), and was also in indexing considered to refer to men, and the gender-bias to be apparent always in the context used (example 2).

Example 1:

> "But it certainly does bring to mind that kind of culture where precisely men are considered as these. --- Well this is terribly male-gendered; one straight away thinks this is a man. I have had these kinds of, there is now literature about it that when both spouses are working, so there is this wonder that who is the actual "breadwinner", who is the main provider." (FI IN3)

Example 2:

"When I looked for equivalents [in an English thesaurus] so there was this talk about working women and working wives, but nothing about men, so they [men] have only the [descriptor] breadwinners, so yes, the gender is clearly a part of it." (FI IN2)

The stimulus word *heads of household* was discussed for the first time when it was introduced as a stimulus word for word associations. It resulted in 22 co-occurrences, out of which 20 were unique.
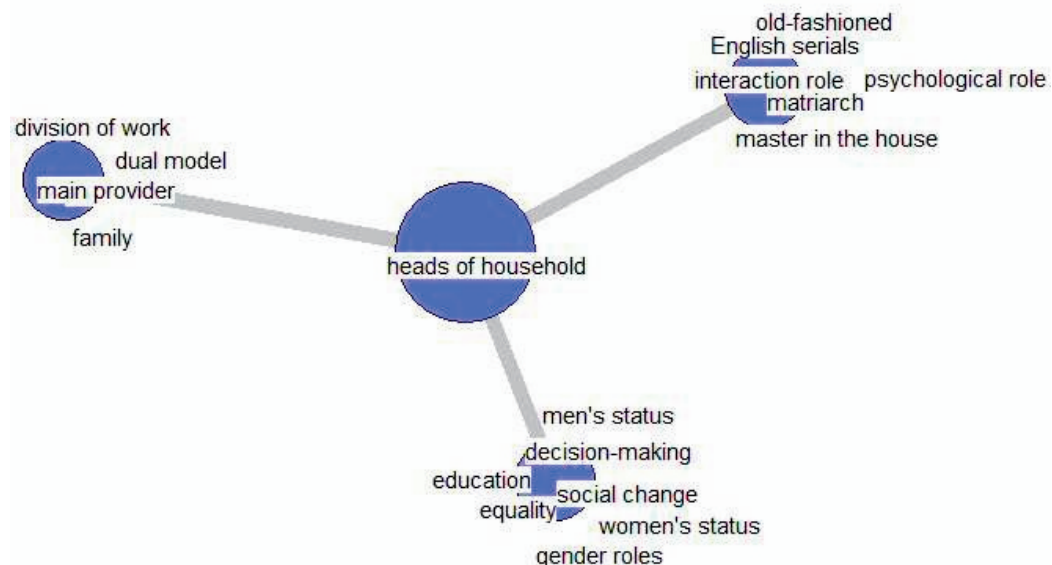
Figure 42: *Word associations for "heads of household" by Finnish indexers*

This was the most heterogeneous sample so far, although still not in conflict within the group or between the other groups studied. It associated together with words expressing gender status and roles within a family. It was translated naturally as *perheenpää* (~head of the family) - "whoever it then will be, because at least in our family we have two of them" (FI IN2)

It was considered to be a statistical term (example 1) and in the current common language usage to be more of a psychological family role (example 2).

Example 1:

"Yes, these are seen in use, at least in the international material, which we also have, is this being used just like this. But I think using these concepts [refers also to breadwinners] is difficult in Finland and in Scandinavian countries. Well, it is often defined --- who in the family is the main… But it is not so unambiguous at least here in the Nordic countries." (FI IN2)

Example 2:

"But then it can also be this kind of role in the family, this kind of psychological role, who wears the trousers ("määrää kaapinpaikan") or who turns over the neck at the end of which usually stands the man's head [laughing]- that kind of an interaction role, yes." (FI IN3)

Breadwinner was thus understood more as an economic role and *head of household* more as a decision-making role. There was seen to be a difference in acceptance and usability if the term was considered to be a common language word as opposed to a descriptor or a statistical term. In common language usage the terms were considered to be more like male-gendered euphemisms, whereas the statistical terms were to a certain extent accepted. The difference was considered to produce a detrimetral influence on the use of these terms in indexing because of their gender-specificity (cf. Ulrych 1992; Venuti 1998).

8.2.2.3.3 Homemakers and Housewives by Finnish indexers

As a stimulus word **homemakers** resulted in 21 co-occurrences, out of which 19 were unique (and the diversity was thus very much similar to the case of heads of households).
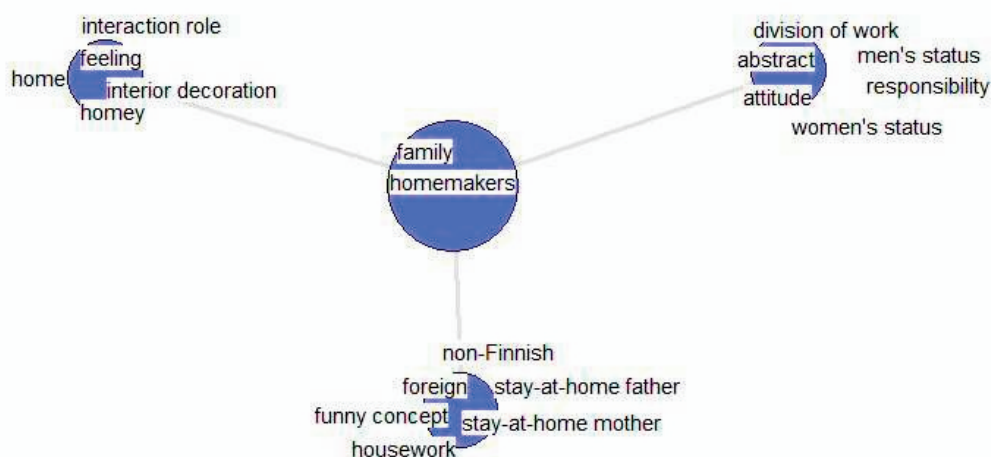
Figure 43: *Word associations for "homemakers" by Finnish indexers*

Although diverging at the term level, the themes were more unified (cf. Iivonen 1989). The stimulus word associated as a role within a family and together with words expressing cosiness, home-parenthood, status, housework and its usability as a term. The connotations raised, and later discussed, were both positive and ironic also by one respondent.

The term was considered to be a directly untranslatable euphemism and translated into Finnish as such, with a description and several alternative or overlapping phrases: *kodinhengetär* (~ lady or female spirit of the house), *kodintekijä* or *kodinluoja* (~ person making home), *kotivastuuhenkilö* (~person caring responsibility of a home – "or how to put it", (FI IN1)).

> "What does it actually mean, does not probably refer to anything concrete… [laughing]" (FI IN1)

> "It refers to a couple who is making their home, to a young couple. It may be an eternal interior decorator, who all the time brings something new to the home and organises there. But it can also be an interaction role, that someone is the one who makes the home feel like home." (FI IN3)

The way the term was expressed in the working version of ELSST was commonly criticised as too narrowly understood. At the same, the term was still acknowledged as being especially difficult to translate into Finnish.

> "*Kotitaloustyötä tekevät*" is too concrete. Homemaker refers to other aspects too; it can be the creator of the atmosphere, security and so on. "Kodinhenki" (~ spirit of the home), is not only connected to who washes and cleans. It is about who takes responsibility that the home is in every sense functional. --- And stay-at-home- fathers are missing in the English version." (FI IN1)

The stimulus word ***housewives*** resulted in 19 co-occurrences, out of which 16 were unique. All the respondents associated it together with stay-at-home mothers, which was also considered as the source term's modern equivalent in Finnish culture.
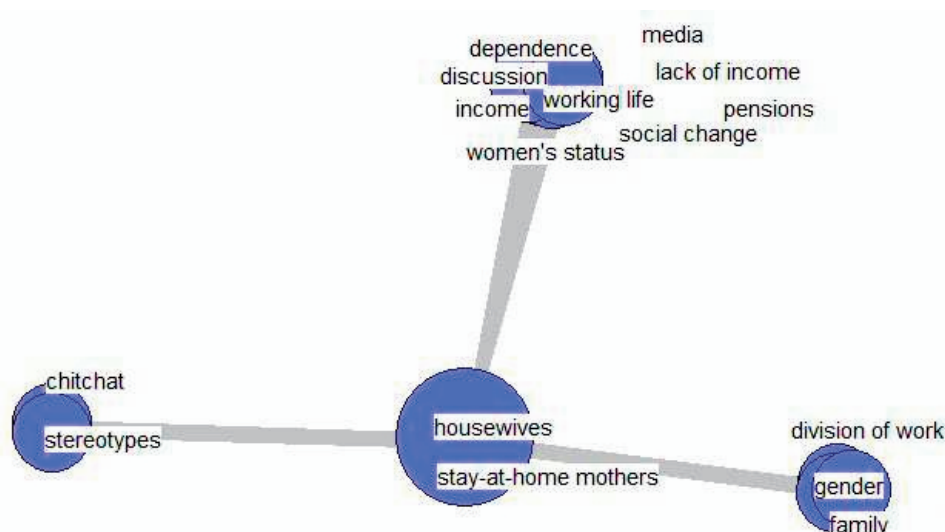
Figure 44: *Word associations for "housewives" by Finnish indexers*

The direct translation was thus *kotirouvat* (homewives), but as its modern and pragmatic equivalent the term *kotiäidit* (stay-at-home mothers) was given. It was also compared to *homemakers* and then seen as more serious work. As well as modernisation, the need was also discussed for a male equivalent, and *stay-at-home fathers* ("koti-isät") was considered necessary. Most probably due to one indexed article the term *househusband* was commonly discussed by the participants, especially in the context of *housewives*.

> [Evaluation of ELSST representation]

> "I see homemakers as one stage in the marriage and apparently not as a person taking care of home. So this housewives is much more like housekeeper type. ---

There should be included also stay-at-home fathers, yes. Although the English source term is about women. What was it here [in the indexed article]... Househusband in one article. But it is the same role, housewife and househusband, the other one is feminine and the other one masculine. --- One alternative could be "kotimies" [homeman] [laughing]. Someone stays as homeman. But it is so that he is outside all fun, well that's how it associates [laughing]" (FI IN3)

8.2.2.4 Simulated indexing task

8.2.2.4.1 Background

Similar to the British indexers interviewed (cf. chapter 8.2.1.5), the Finnish indexers were asked to index the five articles (see also chapter 6.3.5.1 *Indexed documents*).

Together with the articles, the respondents received brief information about the task similar to the British indexers, except that in phase 2 ("This time try to do it as if the end users were European social scientists and you were indexing for the European social science database.") the Finnish informants were guided additionally to provide descriptors in English.

In the interviews the task was briefly discussed – how the informants experienced the simulated indexing task, whether any of the articles were exceptionally difficult, challenging or otherwise exceptional, how the informants chose the descriptors and which tools were used.

8.2.2.4.2 Concerning the process

The attitude of the Finnish indexers towards the simulated indexing task was in general positive (example 1 and 2), and the task was described as interesting and even fun, although also as laborious, especially because of the English version (phase 2, example 2).

Example 1:

"Indexing of articles is very familiar to me --- All these had a very interesting content. But I didn't experience these very difficult in general. I think I got the matters pretty perfectly." (FI IN3)

Example 2:

"The task was interesting, it was fun to read the articles, and to do this indexing project, which corresponds with my own education and interests. But I could see that there are its own challenges. --- And the articles were partly rather long, but they were

indeed that kinds of, that were connected very well to this topic and provided new views to this topic and from different continents and in that way complemented each other very well. In addition, they were very explicit articles, so one well knew where one was going. --- The actual reading was fun, but the articles themselves, the raw job. --- But really, this was pretty laborious, because I don't usually index in English and it was pretty challenging." (FI IN2)

During the interview it was mentioned, surprisingly, that one informant had not completed the second phase, but considered it too demanding. She had started the task, but given up, because considered it was too difficult to take into account the European variations.

"But I really thought over how to index these European, because there might exist different user groups from many cultures, kind of, although Europe today is in some sense very coherent, but one would indeed need to know how these terms are used in certain countries and cultures. But I think it's too much asked for if one should understand their way of thinking, how they would seek for, one should then be a kind of renaissance genius herself. There are enough challenges already. ---

And in Europe there are also very different countries, there are these very well-developed and Northern welfare states and then there are these Mediterranean countries, in which for instance the attitude to some family matters is so different, and there are these poor Eastern-Europe countries, so there are these differences, yes." (FI IN3)

The articles were considered to also be culture-bound and to reflect different attitudes (cf. Venuti 1995, 1998).

"Within these was the article about stay-at-home fathers, and well, it is in Finland also so, that if a father stays at home in order to take care of the children, so it is very strange according to many, but also in Finland it is not strange that a mother stays at home. And actually now even more, it is in all media that when women have children they can stay at home. --- And yes, of course also today one says that when you are at home, you are only at home, it is not so valuable in our culture this homemotherhood. It brings to mind role conflicts and all these." (FI IN3)

The given guidelines were considered to be reasonable, and the amount of given descriptors as very familiar or rather familiar. When asked their opinions concerning the number of suggested descriptors, 9-12, the informants realised that they had not considered it very carefully, and the number had been fairly normal – one informant (FI IN1) stated that (s)he had most probably not even given the minimum and explained, that "instead of certain number of descriptors it is necessarily to see that the most essential is represented in the descriptors given so that one can retrieve them in information seeking", whereas the other indexers stated that they had provided more (example 1), and also used synonymous expressions (example 2). Thus, what is a natural choice for one is not necessarily natural for everyone.

Example 1:

"In guidelines we have at a maximum 15 descriptors, but often less, and then it is emphasised that in principle one cannot say the exact amount, because it is so dependent on, but the guideline is that one should not stretch very long. --- But now I notice that I have put 12 to each, and in some cases even too many. But I think that on

the level on which I normally do indexing, one gets a little deeper than surface when using several. So I think this is a good amount, yes." (FI IN2)

Example 2:

"It is forbidden to provide synonyms, but I do provide them in order to express different perspectives." (FI IN3)

Together with the task, the **indexing inconsistency** was also commonly discussed (cf. Iivonen 1989) and the indexer's perspective (cf. Venuti 1995 and 1998), and indexing was thus seen as a subjective action (and competence), and databases to include several types of indexer discourses.

"Exactly concerning this topic I would think that people with different views would index these differently, or people with different attitudes… I was, for example, very annoyed with the term spouses ["puolisot"], that why am I forced to put marital spouses ["aviopuolisot"] when YSA says so, it distort the thing then. And when you think of how content descriptor in a database is provided by different persons – and if another indexer would have done it – then there might be yet some other perspectives found. And when there has been discussion about indexing quality, every now and then you doubt your own competence ---" (FI IN3)

The vocabulary tools used in the simulated indexing task were identical in phase 1 - The General Finnish Thesaurus YSA and its web-version VESA. In phase 2 the sources varied: Sociological Abstracts Thesaurus (CSA), working version of ELSST, HASSET, dictionaries and Google searches. The vocabulary control tool most often referred was the Sociological Abstracts Thesaurus (CSA).

In the task, the Finnish phase was considered to be the one that dominated (example 1), although it could also be the other way round – one indexer tried to find equivalents from YSA in accordance with English terminology (example 2).

Example 1:

"The English thesaurus was not known beforehand and therefore indexing in English was done more according to the first phase and YSA." (FI IN1)

Example 2:

"Well just recently [--- refers to the indexing task] there was this case of spouses, and it is not allowed to use anymore [in YSA] but it should be either marriage partners or co-habiting partners, and there is no other alternative. I think spouses would be better, because there are these registered relationships and all these more modern. It is the two persons in the family who constitute spouses despite of whether they are married or not and what kind of marriage it is, so the role is there, the role of the spouse, and I was upset because there [in YSA] was not that possibility." (FI IN3)

8.2.2.4.3 Updating the "gender contract"? Childcare reforms in the Nordic countries in the 1990s'

In phase 1, the Finnish indexers provided a total of 43 descriptors, out of which 35 were unique. The provided descriptors were all from YSA (or from its web-version VESA, which are both referred to as YSA). On average, the indexers provided 14.3 descriptors per article. The descriptors used by more than one respondent were:

NO	Finnish (original)	English
3	hyvinvointivaltio	welfare state
3	lastenhoito	childcare
3	Pohjoismaat	the Nordic countries
2	sukupuoliroolit	gender roles
2	vanhempainvapaa	parental leave

Table 30: *Article 1, the most frequently used descriptors by the Finnish indexers in phase 1*

Similar to the British indexers, the descriptors used referred to or were related to childcare and/or parenthood (*childcare, parental leave, mothers, fathers, fatherhood*, tot. 8), social systems and social benefits (*welfare state, home care subsidy, social benefits, social rights, parental allowance*, tot. 8), genderism (*gender roles, gender, gender studies, mens status, womens status*, tot. 6), working life (*working life, reconciliation of work and family life, division of work, sociology of work, politics of work, participation in working life*, tot. 6), geographical place or time (*Nordic countries, Scandinavia, 1990's*, tot. 5), family and family life (*families, family life, family policy, family sociology*, tot. 4), or were more general terms or occasional themes (*attitudes, social change, equality, reforms, renewal*, tot. 6). Despite the shared controlled vocabulary, i.e. use of YSA in phase 1, there were many terminological and conceptual differences (cf. Iivonen & Kivimäki 1998). The indexing was still rather coherent at a more general theme level (cf. Iivonen 1989).

In the sample, the gendered descriptors represented three different perspectives. One respondent referred to men (*fatherhood*), one to both women and men (*mothers* and *fathers* or *mens status* and *womens status*). In addition, all referred to gender aspects by using gender-neutral terms (*gender roles, gender*).

In phase 2, one respondent made some slight changes in a more concrete direction (*isyys/fatherhood* replaced with *fathers, lastenhoito/child care* described with two descriptors *child care* and *child day care*). The tendency was the same with the other respondent, who removed some general terms in phase two (*family sociology, attitudes, Nordic countries*) and also replaced some descriptors with several more concrete and also neutral descriptors (e.g. *naisen asema/womens status* and *miehen asema/mens* status covered with *parenthood, equality/inequality, sex roles*). This strategy is similar to that of internationalisation and existential equivalence (cf. Koskinen 2000), since national characteristics are phased out, but no new ones are added.

8.2.2.4.4 Trait judgments of stay-at-home and employed parents: A function of social role and/or shifting standards?

In phase 1, 34 descriptors were given, out of which 26 were unique, and thus on average 11.3 descriptors per article. The descriptors used were all YSA terms. The descriptors that were provided more than by one indexer were:

NO	Finnish (original)	English
2	arviointi	estimation
2	koti-isät	stay-at-home fathers
2	kotiäidit	stay-t-home mothers
2	roolit	roles
2	stereotypiat	stereotypes
2	sukupuoliroolit	gender roles
2	vanhemmat	parents
2	vanhemmuus	parenthood

Table 31: *Article 2, the most frequently used descriptors by the Finnish indexers in phase 1*

The descriptors used referred or were related to parenthood (*stay-at-home fathers, stay-at-home mothers, parents, parenthood*, tot. 8), gender (*gender roles, men, women, gender, gender differences*, tot. 6), research design (methods or field of study) (*family sociology, gender studies, research methods*, tot. 3), or in their nature more general or as a theme occurred only occasionally (*evaluation, roles, stereotypes, characteristics. qualities, students, family life, transfer effect, social status, standards, efficiency*, tot. 14) or place (*United states*, tot.1).

The gendered descriptors were provided in an equal way referring to both men and women (*stay-at-home mothers* and *stay-at-home fathers*, or *men* and *women*).

In phase 2, one respondent replaced or complemented some general terms with more specific ones (e.g. *vanhemmat/parents* with *parent participation*, *roolit/roles* with *roles* and *social role theory*). One clear cultural change that was made, was when *koti-isät/stay-at-home fathers* and *kotiäidit/stay-at-home mothers* was replaced with *homemakers*. The other respondent also had *koti-isät/stay-at-home fathers* and *kotiäidit/stay-at-home mothers* in phase 1, but they were eliminated in phase two and only covered under *parenthood* (*vanhemmuus/parenthood* in phase 1). Some changes were thus due more to the vocabulary control used, which moved from a general vocabulary tool to specifically social science vocabulary tool, and some due to the more general cultural differences.

8.2.2.4.5 The Marriages of Equally Dependent Spouses

In phase 1, the Finnish indexers provided 46 descriptors, 15.3 per article. All except one (*kahdenpalkansaajan perheet/dual career families*) descriptors given were from YSA. Eleven were used by more than one indexer:

NO	Finnish (original)	English
3	avioero	divorce
3	avioliitto	marriage
3	Yhdysvallat	United States
2	aviopuolisot	spouses (marriage partners)
2	miehen asema	mens status
2	parisuhde	marital relationship
2	perhesosiologia	family sociology
2	riippuvuus	dependence
2	sitoutuminen	commitment
2	sosioekonominen asema	socioeconomic status
2	sukupuoli	gender

Table 32: *Article 3, the most frequently used descriptors used by the Finnish indexers in phase 1*

The descriptors used referred or were related to marital status or relationships (divorce, marriage, spouses, marital relationship, registered relationship, sexual identity, marital status, tot. 13), gender (mens status, gender, masculinity, womens status, gender roles, gender studies, tot. 8), family economics (socioeconomic status, income, earned income, dual career families, tot. 5), or were rather general or occasionally a theme (family sociology, family, dependence, commitment, benefits, advantage, human relationships, quality, equality, social change, equality, satisfaction, working hours, working life, tot. 16) or referred to a place (United States, tot. 3).

The gendered descriptors emphasised masculine aspects (*mens status* and *masculinity* tot. 3, *womens status* tot. 1). Gender-neutral terms referring to gender aspects were clearly more commonly used.

In phase 2, the proceeding was rather similar to the other two articles. Synonymous, near-synonymous, near-equivalent and antonymous expressions were eliminated (such as *avioliitto/marriage* and *avioero/divorce* replaced with just *marriage*, *parisuhde/marital relationship* and *rekisteröity parisuhde/registered partnership* with just *marital relations*), and gender-specific descriptors (*naisen asema/womens status* and *miehen asema/mens status*) were eliminated and encompassed by *sex differences*, the use of which is in the Sociological Abstracts Thesaurus scope note is explained as:

"Used in analyses in which a social problem or phenomenon is explained as a consequence or correlate of the differences between males and females. These differences may be biological or stem from sex-role socialization"

The respondents aimed to follow the terminology of the selected vocabulary, and thus to adapt to its views. The European context was also, in this case, understood as general and international, thus not a context with great variety, but with stricter harmony and internationalisation.

8.2.2.4.6 Socioeconomic Differentials in Divorce Risk by Duration of Marriage

In phase 1, the fourth article resulted in all 36 YSA-descriptors, thus on average 12 descriptors per article. 23 were different, and nine were used more than once. In the table below the commonly used descriptors are presented:

NO	Finnish (original)	English
3	avioero	divorce
3	avioliitto	marriage
3	riskit	risks
3	sosioekonominen asema	socioeconomic status
2	aviopuolisot	spouses (marriage partners)
2	kesto	duration
2	koulutus	education
2	perhesosiologia	family sociology
2	Suomi	Finland

Table 33: *Article 4, the most used descriptors by the Finnish indexers in phase 1*

The descriptors used referred to or were related to marriage (*divorce, marriage, spouses*, tot. 8), personal economics (*socioeconomic status, standard of living, property, income*, tot. 6), expressed place (*Finland*, tot. 2), research design (*family sociology, statistics, research, gender study*, tot. 5) or were used in a general sense or as an occassional theme (*risks, duration, education, housing, inequality, age, families, social status, social change, causes*, tot. 14). Only one descriptor was gender specific (*women´s status*).

In phase 2, the tendency found was slight, and moved in a more specific and economic direction (*perheet/families* eliminated and *economic activity* added, *naisen asema/womens* status replaced with *wives* and *socioeconomic status*, *omaisuus/property* with *family income*), and followed the practices of the changed vocabulary control tool. The strategy used was also in this case closest to internationalisation.

8.2.2.4.7 Men Don't Do This Sort of Thing. A Case Study of the Social Isolation of Househusbands.

In phase 1, 30 descriptors were provided, thus on average 10 per article. All except one ("*henkilöiden välinen vuorovaikutus*", *interpersonal interaction*) descriptor were from YSA. 21 descriptors were different, and the seven indexing terms that were used by more than one indexer were:

NO	Finnish (original)	English
3	lastenhoito	childcare
3	sukupuoliroolit	gender roles
3	vieraantuminen	alienation
2	isät	fathers
2	kotityö	housework
2	perhesosiologia	family sociology
3	lastenhoito	childcare

Table 34: *Article 5, the most used descriptors by the Finnish indexers in phase 1*

The descriptors used referred to or were related to domestic responsibilities and family life (*childcare, housework, family life, division of work*, tot. 7), men (*fathers, stay-at-home fathers, men, mens status*, tot. 5), emotions (alienation, isolation, self-concept, outsiders, tot. 6), roles (*gender roles, role change*, tot. 4), research design (*family sociology, interview survey, gender study*, tot. 4), place (*Australia*, tot. 1), or were in their nature general or used as a theme which only occurred occasionally (*interpersonal interaction, social status, social constructionism*, tot. 3).

The article was about the social isolation of fathers, who stay at home to take care of their children instead of earning money from employment, and the father-aspect was expressed with various terms – *fathers, stay-at-home fathers, men* and *mens status*. One respondent used one descriptor to refer to the aspect, one respondent used two, and one used three. (In the articles the title term *househusband* was used.)

In phase 2, no radical changes were made. One respondent provided identical indexing (in the form of direct translations) with phase 1, and used the term househusbands for *koti-isät/stay-at-home fathers*. The other respondent had used in phase 1 the descriptors *isät/fathers* and *miehet/men*, and in phase 2 covered the aspect of 'male parent' with *man/men*. In this case the strategies used were thus domestication versus internationalisation. The original keywords of the article were *fatherhood, househusbands, masculinity, motherhood, domestic work, housework, child rearing, children, gender*, and thus the male aspect was originally much stronger. (Taking care of home aspect was covered in similar depth.)

8.2.2.5 Equivalence and translation strategies by Finnish indexers

As with the other Finnish groups studied equivalence and translation strategies were studied via several contexts – by asking directly what is aimed at in multilingual thesaurus equivalence, how the respondents would translate certain family role terms, and in more general discussions such as those concerning the usability of multilingual thesauri.

The indexers tended to follow the traditional view on equivalence in multilingual thesaurus and indexing in the sense that it is normal to aim for equivalence between different languages. In the determination of equivalence, similar to thesaurus constructors, the indexers considered the documents to be indexed and retrieved in important and essential.

> [Aim of multilingual thesaurus equivalence?]

> "Well, as unambiguous description as possible is probably aimed at with all used languages, so that it, of course, also would match with the documents described." (FI IN3)

In addition, one respondent mentioned the aims of multilingual indexing as a challenging question to consider and extend. The respondent (FI IN1) asked if the indexing in different languages should be equivalent with each other or could several language versions be used multilingually in order to complement each other. This would allow Finnish indexer to additionally use the English culture-bound terms when indexing an English book, for example.

> "How it benefits the information seeker depends on the information seeker. – How information seeker is able to operate in other languages and use foreign language material. If the seeker has good knowledge of language, the different language versions can be used to complement each other, multilingually." (FI IN1)

The way the translation corpus of the study (family roles terms) were represented in the working version of ELSST was considered narrow and outdated from the modern Finnish perspective. The three Finnish indexers interviewed would have liked a more gender-neutral and psychological representation (cf. Venuti 1998).

The respondents did not pay much attention to the style as a significant factor as did the other groups studied (cf. chapters 7.2.2.5 and 9.3.2), and there was in practice great variation both within and between the respondents. The source term and what kind of problems it presented provided the greatest influence.

Family roles (FI: *perheroolit*) was translated formally (literal, direct translation), but it was noted, that the pragmatic meanings differ to some extent. *Breadwinners* (FI: *perheenelättäjät*, *ansiotulon saajat*) was translated more dynamically (cf. Nida 1964/2000; Vehmas-Lehto 1999) and was also used as a descriptive phrase. *Heads of household* (FI: *perheenpäät*) was considered as a concept to be difficult, but the term itself was easily translatable. The conceptual problems were considered to exist also in the source language and culture, but to become greater in the Finnish context. *Homemakers* resulted in several alternative translations, most often euphemisms

referring to the lady of the house. *Housewives* had two equivalents depending on time. Its traditional equivalent would be a well-established but nowadays out-dated term *kotirouvat* (homewives), whereas its modern equivalent would be – and especially in the Finnish context - *kotiäidit* (stay-at-home mothers). Today, and especially in the Finnish context it would require also a masculine equivalent and addition *koti-isät* (stay-at-home fathers). The *housewive*'s direct masculine form *househusbands* was considered to differ in style and partly also meaning in English.

Translation strategies were also varied. – *Breadwinners*, *heads of household* and *family roles* were translated in accordance with *internationalisation*, whereas *homemakers* in accordance with foreignisation and *housewives* with domestication (cf. chapter 3.4.1.3 *Domestication or foreignisation – or internationalisation*).

Internationalisation strategy was discussed both at the level of selected concepts and terms and for the planned thesaurus corpus. The common idea was that it is not reasonable to aim at exact equivalence, but to understand dissimilarity as an unavoidable reality (cf. Wierzbicka 1991), and thus aim at consensual equivalence (cf. Koskinen 2000).

> "It is challenging that the concepts are different, and there does not exist similar concepts in the other country, for example, as in Finland. But it is just precisely that, that they do not go word-for-word right, but that there exist as great correspondences as possible, that they are understood as similarly as possible in different countries." (FI IN2)

> "Well, there might of course be the relatives – if you think about for example some Italian family, so there the concept of relatives is much broader, for us it is pretty much like nuclear family. But well… I suppose same words would be used." (FI IN1)

Thus together with internationalisation strategy, existential equivalence, or consensual equivalence (cf. Koskinen 2000) was also present. Similar to thesaurus constructors, the indexers also considered compromising as essential in multilingual thesaurus construction.

> "One has to know how the things are in different countries, [since] terms are very much bound to the cultures. Compromises are needed in multilingual thesaurus construction." (FI IN1)

> "European thesaurus reflects many [countries/cultures/languages]. One must also adapt, since it cannot reflect Finland." (FI IN2)

The differences and thus problems in the European context were commonly considered to appear at several cultural and at sub-cultural level. Europe, in general, was considered to be rather uniform. As discussed recently, the tools for multilingual indexing were considered necessary, and being familiar with all the variations within the European indexing and information seeking context (example 1 below) as being too great a demand. Within Europe Finland and England were considered to be close to each other (example below).

The English are more European, there is clearly a separated Nordic line. And well, certainly Britain is not so strongly that kind of stronghold of Catholic and stay-at-home motherhood, but it is closer to European than Nordic. And in Britain women have been

working a long time, but still the reconciliation of work and family and all these social benefits are well… Sure they do have something there, but for example the childcare system is different in Nordic countries." (FI IN2)

The studied case was considered difficult, especially due to the time aspect (cf. Aitchison 1991; Suojanen 1993; Hirsjärvi and Hurme 2000). – Which time period to reflect in the translation, if to reflect only one?

> "[These are] difficult to understand and translate. These matters have been in a flux and the change has been extensive." (FI IN1)

Discourses were not discussed very much, but different ways of seeking information (example 1), and cultural differences were emphasised. Linguistic problems at the level of language (Finnish versus English) were also considered to cause problems. Sometimes there was a common concept in both cultures, but it was not simple to express it in the other language (Finnish or English). Whatever the reason for differences, concept harmonisation was commonly considered to be required and thus the function of thesauri was raised in relation to ideas of internationalisation, and in one case it was considered to also have had influence in other academic writing by harmonising concepts (example 2).

Example 1:

> "Social policy is very country depended and the British have already a very different system, not to speak of Americans, and not to even speak of the rest of the planet, in all the countries they don't necessarily even know the thing. [About gendered financial responsibility in relationships, men having in Britain more responsibility whereas in Finland the economical roles in relationships are more gender-equal] --- so in these very similar matters the view can be so different, so yes, it must certainly influence somehow also how people seek for information, even more to it then." (FI IN3)

Example 2:

> "I think that in this increasingly globalising world we, however, do need… Well there are so many kinds of national characteristics and concepts, and to translate them directly is terribly difficult, so it is very good that with these kinds of tools which do not aim at word-for-word correspondence, but to match interpretation at the content level as much as possible. And then, of course, there is this fact that thesauri create something completely new too, a new way to produce scientific text, when these concepts will become sort of standardised for the users. So, often we speak about same matters we use different words for it, so it would be good for them too to have a certain kind of conformity when it is possible. (FI IN2)

As discussed in chapter 3.4.1.2 *The idea of equivalence*, Koller (1989, see also Koller 1995; Zethsen 2004) separates five factors, which lead to equivalence types and then further into five translation types. If two words are true cross-linguistic equivalents, they agree in all the five levels (factors). In the specification of equivalent types the five factors are: extralinguistic content, connotations, text and language norms, receiver and certain formal-aesthetic features of the SL texts. (Koller 1989, 100) How did the Finnish indexers consider equivalence in relation to the typology of Koller?

In their aims and translations the Finnish indexers tended to favour the denotative level, although the connotative level was still in some cases considered to be superior. There was no clear consensus with regard to problematic terms in which these two aspects conflicted.

Text-normative and formal aspects were not considered as much as pragmatic aspects of translation. It was considered important that the translation is accepted and usable in the target context (in the way it was in source context) and that the effect is the same. Translation should thus produce similar results in an information seeking situation as the source version does.

In the simulated indexing task only two respondents also provided European indexing terms (in English), whereas one respondent considered it to require more cultural knowledge than is realistic to suppose for an indexer to have without a specific (European wide) controlled vocabulary tool. The two first mentioned respondents considered European indexing to be international, and used American, European as well as British vocabulary sources. The strategy (which the used terminology represents) was closest to internationalisation, in which national characteristics are not emphasised (cf. Koskinen 2000). There was thus great variation for domestication (one respondent) versus internationalisation (two respondents) strategies.

- **Summary and Conclusions**

Similar to British indexers (cf. chapter 8.1) the Finnish indexers considered knowledge of subject domain to precede knowledge of foreign languages in multilingual indexing. They also recognised different discourses and considered concepts and terms as time-dependent (cf. Buckland 1999; Lehtonen 1994; Venuti 1998; Vehmas-Lehto 1999 and cf. Aitchison 1991; Suojanen 1993; Hirsjärvi and Hurme 2000). Aiming at gender-neutral language use made the terminological challenges even harder (cf. UCL 2000; Brunel 1999; School of Psychology 2002; University of Southampton 1997; Engelberg 1993, 1998).

The respondents tended to make difference between different types of meaning and styles (cf. Koller 1989; Lehtonen 1994; Buckland 1999; Vehmas-Lehto 1999) when defining and translating the family roles terminology studied. *Family roles* was understood in a broader sense and especially as having a more psychological meaning than represented in the working version of ELSST. It was translated fluently as *perheroolit*, but the respondents emphasised, that the thesaurus context should also reflect modern practices (e.g. different kinds of marital status and dual-earner model). The other English terms studied were considered to be foreign concepts resulting in problems in translation and/or use. *Breadwinners* was translated as *perheenelättäjä* or *ansiotulon saaja* and understood rather similar to the other discourses studied, as a family member providing more income to the family. It was considered to be gender-biased, and in practice not in use for women breadwinners (cf. Koller 1989, Engelberg 1993 and 1998). *Head of household* was considered as easy to translate into Finnish (as

perheenpää), but the term's usability was considered to be very much context-dependent. As a common language word it was considered to have negative connotations, as a statistical term positive or neutral, and as a documentary language term slightly negative. *Homemakers* was the hardest one to translate providing several descriptive phrases as Finnish equivalents. *Housewives* was considered as an out-dated concept, and the modern equivalent to be stay-at-home mothers and stay-at-home fathers (*kotiäidit + koti-isät*). The Finnish ELSST versions were accepted expect in the case of *homemakers* (FI: *kotitaloustyötä tekevät*)

In the Finnish context, the impact of the nature of the vocabulary tool was clearly seen, and in second phase of the simulated indexing task the social science discourse became more apparent than in phase 1, where the respondents used almost without exception descriptors from The General Finnish Thesaurus YSA. In the family terminology the changes were more clearly cultural, and English tools missed specific words common to the Finnish context, such as *stay-at-home fathers/househusbands*. In general the two Finnish indexers, who also provided descriptors in phase 2, were aiming at rather gender neutral indexing even when the article was gender-specific.

Similar to the British samples studied, in the Finnish indexer's discourse there existed greater consistency at the broader theme level than at the more specific term level (cf. Iivonen 1989; Iivonen & Kivimäki 1998). As with the database indexing studied (of ARTO, LINDA and COPAC) there were variations, especially as to whether to express the indexed study's time, place and methods of research.

As with the other Finnish groups studied, equivalence and translation strategies were studied via several contexts – by asking directly what is aimed at in multilingual thesaurus equivalence, how the respondents would translate certain family role terms, and in more general discussions about usability of multilingual thesauri.

The Finnish indexers tended to follow the traditional view on equivalence in multilingual thesaurus and indexing in the sense that it normally aims at equivalence between different languages. Similar to the thesaurus constructors they also reflected and emphasised the role of the documents to be indexed and retrieved in.

The respondents did not pay as much attention to the style as a significant factor as the other groups studied did and there was, in practice, great variation both within and between the respondents. Translation strategies were also varied, but internationalisation was the most commonly used strategy. Internationalisation strategy was discussed both at the level of selected concepts and terms and at the planned thesaurus corpus. The common idea was that it is not reasonable to aim at exact equivalence, but to understand dissimilarity as an unavoidable reality (cf. Wierzbicka 1991), and thus aim at consensual equivalence (cf. Koskinen 2000) and together with internationalisation strategy existential equivalence or consensual equivalence was also presented (cf. Koskinen 2000).

The differences and thus problems in the European context were commonly considered to appear at several cultural and sub-cultural levels. Europe was considered, in general, to be rather uniform. The studied case was considered to be difficult, especially as regards the time aspect (cf. Aitchison 1991; Suojanen 1993; Hirsjärvi and

Hurme 2000). Discourses were raised usually in the context of different ways of seeking information, and in this matter cultural differences were emphasised. Linguistic problems at the level of language (Finnish versus English) were also considered to cause problems. Whatever the reason for differences concept harmonisation was commonly considered as necessary and thus this function of thesauri was brought up in relation to ideas of internationalisation.

The practical considerations were partly in conflict with the theretical considerations. In their practical aims and translations the Finnish indexers tended to favour the denotative level, although the connotative level was still in some cases considered superior. Text-normative and formal aspects were not considered as much as the pragmatic aspect of translation. It was considered important that the translation is accepted and usable in the target context (in the way it was in source context) and that the effect is same. Translation should thus produce similar results in an information seeking situation as the source version.

In the simulated indexing task, only two of the Finnish respondents also provided European indexing terms (in English), whereas one respondent considered it to require too much cultural knowledge. The two first mentioned respondents considered European indexing international, and used American, European as well as British vocabulary sources. The strategy (which the used terminology represents) was closest to internationalisation, which does not favour national characteristics (cf. Koskinen 2000). There was thus great variation for domestication (one respondent) versus internationalisation (two respondents) strategies.

One Finnish respondent brought up that multilingual indexing may not necessarily need to aim at giving equivalents (in the sense of translations of descriptors) in many languages, but it may be used to aim at complementing monolingual indexing.

8.3 Similarities and differences in indexing

The emphasis in the analysis of the Finnish ARTO and LINDA and the British COPAC database was on the time period between 1990-2003/2004 and on descriptors versus title words. In the analysis, the limitations were clearly seen concerning the systems, the various practices, and also the lack of a common vocabulary tool for indexing in COPAC (cf. von Ungern-Sternberg 1994).

The way family roles terminology is understood in the working version of ELSST and in the other thesauri studied (cf. 9.1) was, in general, rather foreign and unusable in the Finnish ARTO database. *Family roles* (FI: *perheroolit*) resulted in some relevant records, in which family was commonly understood to consist of two generations (parent(s) and children), but the types of families varied (traditional nuclear families, remarried families etc.). The Finnish aspect clearly had a strong emphasis on parenthood. In LINDA the *family roles* was generally understood as it was in ARTO, but in a more general sense and the emphasis to have a more agricultural context and also in the division of labour. As a concept it clearly existed in LINDA. In COPAC

family roles was hardly used by indexers. Its lexical context was mainly an issue of the role of grandparents.

Breadwinners (FI: *perheenelättäjät*) and *heads of households* (FI: *perheenpäät*) were foreign concepts in ARTO and LINDA databases and although seldom met, were understood in a euphemistic sense and therefore differently from the other discourses studied. In COPAC, *breadwinners* represented natural language and was used by the authors as a title word in the sense commonly understood in the other discourses studied. The term was often used to refer to women and female-related terms were used as qualifiers (male understood as a standard, cf. Engelberg 1993; Venuti 1995 and 1998). *Heads of household* as a COPAC term was used most frequently in the form of its female equivalent *women heads of household*. It represented more documentary language than natural language, and it was commonly understood as a statistical and thesaurus term (cf. chapters 7 and 9), although its common context in COPAC was developing countries and single-mothers. Similar to the case of *breadwinner*, the masculine was understood as a standard, and thus when *heads of household* was used to refer to women it was used together with words expressing the female aspect.

In ARTO *Homemakers* (FI: *kotitaloustyötä tekevät*) was also a foreign concept as well as the English version *housewives*, the Finnish (direct) translation *kotirouvat* was seldom used except in an aggravated sense. The Finnish translation *kotiäidit* and *koti-isät* was familiar to Finnish practices, but the topics related to the terms had a stronger emphasis on caregiving and well-being issues than in the English versions, where labour force issues are respectively emphasised more than in ARTO. In LINDA *homemakers* was little used and usually understood differently than in the thesaurus discourse (cf. chapter 9) and as a home aid (occupational title). The Finnish ELSST equivalent was not used.

In LINDA *housewives* was used more often as a title word than as a descriptor, and the Finnish reference to the concept was often expressed with the descriptor women (*naiset*). The direct Finnish translation *kotirouvat* did not represent academic or controlled language. The equivalents *kotiäidit* (stay-at-home mothers) and *koti-isät* (stay-at-home fathers) commonly represented as modern (cf. chapters 7 and 9) had become more popular during the 1990's. The two last mentioned represent equally documentary and natural language terms in LINDA.

In COPAC, *homemakers* was used rather ambiguously, but mostly in the sense of stay-at-home mothers and housewives (cf. 7.1.1.4). It was clearly more common as a descriptor than as a title word. It was often used in the context of family and women's roles. It was clearly a feminine term and not used for men. *Housewives* was a frequently used term in COPAC, and most commonly used in the context of English language material. Although very general in nature, the term resulted in mostly social science material. It was clearly used more often as a descriptor than as a title word. The topics related to the term mostly concerned women's employment, home economics, social conditions and sex roles.

The number of terms representing transliteration problems rather than translation problems was less than one tenth in ARTO and LINDA, and in COPAC about one fifth.

The indexing style was more gender-neutral in the Finnish databases than in the British. Issues such as domesticity, modernity and male gender were commonly "hidden" in the indexing of the samples studied and consequently in bibliometric maps unobservable or biased matters (cf. Chen 2003; Venuti 1995; Yule 1996).

The British and Finnish indexers considered knowledge of subject domain to be prior to knowledge of the language.

In the **simulated indexing task** the British respondents considered their indexing to be rather international or European in phase 1, where the indexers were asked to index certain documents for their normal user groups (cf. Adler 1997; Hall 1981). In the simulated indexing task the Finnish and British indexers had different reasons for not complementing phase 2 (completely or partly). Whereas one British indexer considered phase 1 to also cover phase 2, one Finnish indexer considered phase 2 as too demanding due to the differences in culture. In general, the changes made were mostly towards internationalisation, in which national characteristics are reduced and harmonised (cf. Koskinen 2000).

In the indexing the aim was to describe both the general and the specific level of the topic, and it was also considered as a clear guideline (cf. Rowley 1988). Different **discourses** were also considered (cf. Buckland 1999; Lehtonen 1994; Venuti 1998; Vehmas-Lehto 1999). Together with the different discourses the **time aspect** was also often discussed (cf. Aitchison 1991; Suojanen 1993; Hirsjärvi and Hurme 2000) and it was seen as a construction problem as well as an update problem. The problematic becomes more challenging especially in the context of gender-neutral language use (cf. UCL 2000; Brunel 1999; School of Psychology 2002; University of Southampton 1997; Engelberg 1993, 1998). In all the cases there occurred a considerable number of terminological and conceptual differences (cf. Iivonen & Kivimäki 1998), and the indexing was remarkably more coherent on a general theme level (cf. Iivonen 1989). As with the database indexing studied (of ARTO, LINDA and COPAC) there were variations especially as to whether to express the time, place and methods of the indexed research.

The Finnish and British indexers frequently made (in practice) a clear difference between **different types of meaning** and especially between semantic and pragmatic meaning (cf. Koller 1989; Lehtonen 1994; Buckland 1999; Vehmas-Lehto 1999) when discussing the family roles terminology. The studied concept, term and descriptor *family roles* was in general considered more broadly than in the thesauri studied (cf. chapter 9.1.1) where it also included psychological roles. The Finnish indexers translated it fluently as *perheroolit* (direct word-for-word translation), and similar to the British respondents also emphasised that the thesaurus context should also reflect modern practices. British and Finnish indexers considered *breadwinners* to be a common language concept and as a term as out-dated and sexist (cf. Koller 1989, Engelberg 1993 and 1998), but in a more positive sense if it was evaluated as a descriptor. In Finnish, it was additionally clearly a foreign concept, and this resulted in a variation of translations. *Heads of household* resulted in similar reactions as *breadwinners*, and it was also considered to need modern equivalents in a thesaurus context. Although problematic as a concept, it was easy to translate into Finnish (as *perheenpää*).

Homemakers resulted in more positive reactions than *housewives*. The semantic and linguistic context of the two last mentioned was close to those of thesauri, and in all they were understood rather similarly to the other discourses studied, but from a broader perspective. *Homemakers* was the most difficult one to translate into Finnish providing several descriptive phrases as Finnish equivalents. In the Finnish context *housewives* was considered to be an out-dated concept, and the modern equivalent to be stay-at-home mothers and stay-at-home fathers (*kotiäidit + koti-isät*). The Finnish ELSST versions were accepted except in the case of *homemakers* (FI: *kotitaloustyötä tekevät*).

There was a gap between theoretical and practical discussions. Discussing on the more principal level, the aspects discussed by the respondents were related to Koller's (1989) equivalence types such as denotative, pragmatic and formal equivalence (cf. ibid. and chapter 3.4.1.2 *The idea of equivalence*), and the connotative and text-normative aspects of equivalence were not raised when discussing the terminological case. The respondents presented their views on **dynamic equivalence** (cf. Nida 1964/2000; Vehmas-Lehto 1999 and chapter 3.4.2.1 *Dynamic Equivalence Theory*). They also presented views closest to the **domestication** strategy (cf. Venuti 1998; Lindfors 2001 and chapter 3.4.1.3 *Domestication or foreignisation – or internationalisation*) when aiming at translations to reflect the target culture, but in practice (cf. simulated indexing task situation) preferred **internationalisation** (cf. Koskinen 2000). In the Finnish context the impact of the nature of the vocabulary tool was clearly seen, and in the simulated indexing task's second phase the social science discourse became more apparent than in phase 1, where the respondents used almost without except descriptors from The General Finnish Thesaurus YSA. In the family terminology the change was more clearly cultural, and the English tools failed to find specific words common in the Finnish context, such as *stay-at-home fathers/househusbands*. In general, the two Finnish indexers, who also provided descriptors in phase 2, were aiming at a rather gender neutral indexing even though the article was gender-specific. The differences in the simulated indexing task were thus in their nature both organisation and culture specific.

In most cases there was a clear difference between the semantic and pragmatic meanings especially on gender-relatedness (cf. Lehtonen 1992; Engelberg 1993; Yule 1996; Venuti 1998; Vehmas-Lehto 1999).

It became obvious that not only the aims of multilingual thesauri, but also the aims of multilingual indexing should be discussed properly. One Finnish respondent raised the point that multilingual indexing may not necessarily need to aim at giving equivalents in many languages, but to be used to aim at complementing monolingual indexing.

9 Equivalence and family roles in thesaurus construction

This chapter firstly discusses how the case, family roles terms, are represented in different thesauri, and secondly, how thesaurus constructors understand and translate the five terms studied. It thirdly discusses how thesaurus constructors understand equivalence and aims of translation in multicultural thesaurus construction. Construction practices and possible changes during the research process are also reported. Accordingly to the starting point for the empirical case, the emphasis is on the Finnish aspects (see esp. chapters 1 *Introduction* and 6 *The Study*). As with the other groups studied, it is noteworthy, that due to the qualitative material and small numbers there is no statistical significance in the analysis and in the bibliometric analysis maps.

As dicussed in chapter 6 *Methods and material*, the different types of materials and participants – e.g. thesauri versus British thesaurus constructors versus Finnish constructors – do not represent actual comparison pairs, but are used to complement each other and illuminate equivalence matters in multilingual and multicultural thesaurus construction from different angles. The number of participants does not aim at being even and as such comparable – for example, there are more thesaurus constructors, since the final focus is mostly on them. Similar to thesaurus constructors, indexers are also considered to be information specialists, whereas social scientists are considered to be experts on research terminology and potential thesaurus users more from the perspective of information seeking.

9.1 Semantic lexical networks in thesauri

Pragmatic indexing term equivalence reflects several discourses, since – as already discussed in chapter 1.1 *Background*, each information search in a database covers at least five different languages: the authors, the indexers, the synthetic structure, the users and the search strategy (Buckland 1999, see also Bates 1986; Iivonen 1995), which all represent a type of discourse. In order to gain knowledge about pragmatic indexing term equivalence one of the discourses studied is the discourse of existing thesauri and it is done within the selected theme, family roles.

The nine **thesauri** used to examine and compare the representation and understanding of the studied terms (case family roles) and the idea of multilingual equivalence are:

- multilingual UNESCO Thesaurus, Eurovoc and ELSST

- monolingual HASSET (British-English), SOSIG (British-English), CSA Thesaurus of Sociological Indexing Terms (American-English/International), ICPSR Subject Thesaurus (English-International), ERIC (American-English)

and YSA (Finnish-Finnish). (For more about the studied thesauri see Chapter 6.3.1 *Thesauri*.)

Thesauri were studied from 2001-2003 and the material was updated in January 2005 and in January-February 2009. The situations in 2003 and 2009 were analysed. It transpired that ELSST was still not published, so no comparisons were made with the working version of 2001. SOSIG was no longer an independent thesaurus (with its roots in HASSET), but in the General Social Science catalogue the thesaurus used is nowadays (2009) HASSET, and because of this the changes are more due to the change in the thesaurus used.

Thesauri matches are reported in detail in the tables in Appendix 8 *Semantic lexical networks in thesauri*. (In the tables additions are marked with parenthesis (x) and removals with brackets [x]). In the analysis and in the bibliometric maps different spelling types are represented according to the most common one. Bibliometric maps represent the situation before interviews were made, i.e. in 2003. – New maps were not produced, since the changes in contents were rather minor (as later discussed in more detail).

9.1.1 Family roles in thesauri

The studied term *family roles* was found in six of the studied thesauri: UNESCO, ELSST, HASSET, SOSIG, ERIC and CSA Sociological Thesaurus. (The representations found in them are represented in table 1 in Appendix 8.)

The lexical network of term *family roles* in the matched six thesauri can be illustrated with the help of Bibexcel as:
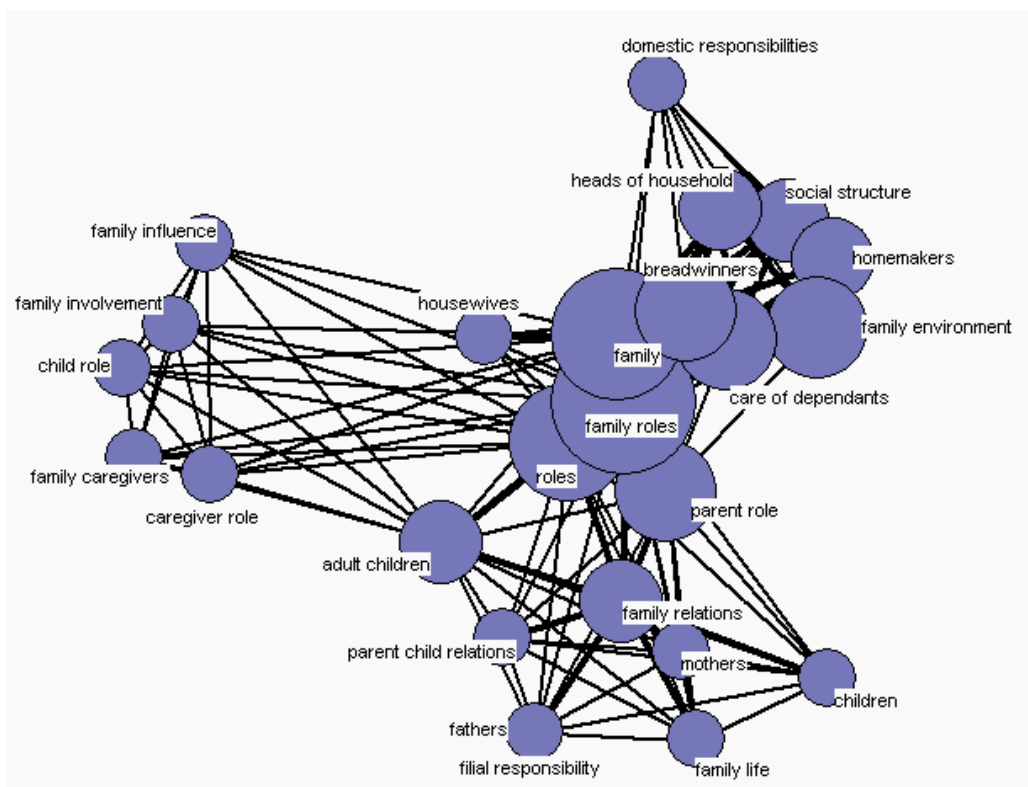
Figure 45: *Family Roles –lexical network in thesauri*

According to the studied thesauri, *family role* is understood as a social role and in a strong connection to caretaking roles and also in connection to supporting and occupational roles. It is usually not directly about family members (mother, father, child, grandparent), although the examples given in the Sociological Thesaurus' scope note also suggest this aspect (see Appendix 8: *Semantic lexical networks in thesauri*).

The studied term *family roles* did not occur in three of the studied thesauri: Eurovoc, ICPSR and YSA. These thesauri were studied further in order to find near-equivalents.

In Eurovoc "*social role*" has as a Finnish equivalent "*sosiaalinen merkitys*"[84], the meaning of which differs even more from the original idea of the role, however, *family/perhe* has as a narrower term *head of household/perheenpää*. Together *social role* and *family* may represent inexact partial equivalence to the term *family roles*.

ICPSR includes several terms related to family roles. It includes terms such as *families*[85], *family relationships*[86], and *family relations*[87]. It does not include roles or tasks

---

[84] [~ social significance/status]

[85] NT dual career families, extended families, farm families, single parent families, stepfamilies.

or duties, but more specific terms such as *social roles*, *gender roles*, *family work relationship*[88] and *domestic responsibilities*[89]. Together they may represent partial equivalence or a narrower or broader equivalence to the term *family roles*. On the basis of the thesaurus context of the terms, the especially relevant ones seem to be *family relationships*, *family work relationship* and *domestic responsibilities*.

In YSA *perhe/family*[90], *perheet/families* and *roolit/roles* can be found, which together may represent inexact partial equivalence to the term *family roles*. As a narrower term for the plural form *perheet/families* different types of families are listed; *lapsiperheet/childfamilies*, *monikkoperheet/multiple birth families*, *nuoret perheet/young families*, *sateenkaariperheet/rainbow families* (*/LGBT families*), *suurperheet/~extended families*, *tukiperheet/"support families"*, *uusperheet/remarried families*, *vauvaperheet/baby families*, *yksinhuoltajaperheet/single-paret families*. As RT *family/perhe* is given and as UF *ydinperheet/nuclear families* is given. It belongs to thematic group *Väestötiede. Demografia / Demography.* (Note 22.1.2009) The singular form *perhe/family* refers to the concept as an institution. As related terms *avioliitto/marriage*, *avoliitto/ cohabitation without marriage* (*common-law relationship / common-law marriage, domestic partnership*) and *perheet/families* are given. The term is also used for *ydinperhe/nuclear family*. This term belongs to a different group than its plural form, namely *[60] Sosiologia. Sosiaalipsykologia / Sociology. Social psychology.*

An even clearer difference with the singular versus the plural form is with the case of household, in which the singular form represents a much broader concept than its plural case, but is used in the YSA thesaurus as having one meaning. The singular form *kotitalous* (household/domestic science/housekeeping/household management) has as broader terms *oppiaineet/subjects* and *taitoaineet/practical subjects*, and it belongs to the thematic groups *[24] Ravitsemus. Ravitsemisala. Elintarvikeala. Majoitusala. Kotitalous / Nutrition. Nutrition field. Food industry. Accommodation.* and *[57]*

RT adoption, childlessness, children, family background, family conflict, family counseling, family courts, family history, family leave, family life, family planning, family relations, family relationships, family size, family violence, family work relationship, genealogy, grandparents, household composition, households, living arrangements, marriage, older parents, parents, siblings, single parents, spouses.

[86] NT adult children, children, daughters, grandchildren, grandparents, inlaws, parents, relatives, siblings, sons, spouses, stepchildren, stepparents.

RT child rearing, families, family structure, genealogy, intergenerational conflict, stepfamilies.

[87] RT families, family conflict, family life, love.

[88] SN: Refers to the effects of employment on family life as well as changes in work organization resulting from the increase in number of dual career families.

RT childcare, dual career families, families, family life, job stress, occupations, single parent families, work, working hours, working mothers, working women.

[89] SN Applies to maintenance of the home environment. NT housework, RT working mothers, working women. Housework in its turn has as NT cleaning and cooking. (Note 22.1.2009)

[90] RT avioliitto/marriage, avoliitto/common-law marriage, perheet/families. UF ydinperhe/nuclear family. Belongs to thematic group [60] Sosiologia. Sosiaalipsykologia / Sociology. Social Psychology. (Note 22.1.2009)

Kasvatus. Opetus. Koulutus. / Rearing. Schooling. Education. Its plural form *kotitaloudet/households* has as a related term *ruokakunnat/households*, and it belongs to the same thematic group [24] as its singular form, but in addition (and instead of [57]) also into *[61] Väestötiede. Demografia/Demography.*

Ruokakunnat is also a Finnish equivalent for the term household. In YSA it has a scope note ("Huomautus"): Ruokakunnan muodostavat henkilöt, joilla on yhteinen ruokatalous / Household consists of persons, who have a common food management. It belongs to a thematic group Demography. As a related term asuntokunnat / households is given. In its scope note the meaning is clarified: "Asuntokunnan muodostavat kaikki samassa asuinhuoneistossa vakinaisesti asuvat henkilöt" / A household consists of all the individuals permanently sharing a flat. The term also belongs in Demography. In other words, family's near-equivalent household is covered in YSA with several near-synonyms, which in a thesaurus context is otherwise narrow.

In YSA roolit/roles has as NT sukupuoliroolit/gender roles. The term belongs to several different kinds of thematic groups :[55] Psykologia/Psychology, [60] Sosiologia. Sosiaalipsykologia, / Sociology. Social Psychology. [95] Teatteri. Tanssi. / Theatre. Dance.

In ERIC *role* had been added in 1994 and it has a scope note: "*Functions or tasks expected of or performed by individuals, groups, or things (Note: Use a more specific term if possible)*". As narrower terms ("Suggested additional or alternate search terms that are more specific than the original term searched"), the terms given are: *Administrator Role; Board of Education Role; Caregiver Role; Child Role; Citizen Role; Community Role; Counselor Role; Family Role; Government Role; Institutional Role; Language Role; Mass Media Role; Parent Role; Role of Education; Role of Religion; Sex Role; Staff Role; Student Role;* and *Teacher Role.* It is noteworthy, that *caregiver role* and *parent role* are not under *family role.* Related terms ("Suggested additional or alternate search terms that are conceptually linked to the original term searched") given are: *Influences; Relationship; Responsibility; Role Conflict; Role Models; Role Perception; Role Playing; Role Theory* and *Status.* Role is used for (UF) *Functions (Sociology)* and *Social Role.*

During the studied time period 2002-2009 no major changes occurred for the term *family roles.* The new narrower terms *householders* and *parental role* had been added in HASSET (checked in 22.1.2009, the page last updated 19 September 2008).

Family is covered with different kinds of structural depth and conceptual scope in different thesauri. It is not easy to cover the term *family roles* with one or even a few other terms in the thesauri, which did not include the studied term. In addition, the concept and term *family* has been in active flux. New, more precise terms referring to household, and scope notes had been added, and some were removed. – The trend in the studied thesauri is not coherent, unlike, for example, in YSA; in ERIC many family related terms are not preferred anymore [91].

---

[91] ERIC Jan 2009: extended family 1977-2004, households (added 1966), nuclear family 1977-2004, rural family 1966-2004, stepfamily 1982-2004, traditional family unit 1986-2004. Family has been added in 1966

9.1.2 Breadwinners and Heads of Household in thesauri

Five of the nine thesauri studied included the studied term *breadwinner*, and they are represented in table 2 in Appendix 8: *Semantic lexical networks in thesauri*. In Eurovoc *breadwinner* was understood as part of a family, and in ELSST, HASSET and SOSIG consistently as a family role. EUROVOC includes both Finnish and English equivalents and the Finnish equivalent given is *perheen huoltaja* [~family provider, family supporter]. ERIC guides the user to use the form *heads of household* (22.1.09 in all 348 "postings") instead of *breadwinners* and *family breadwinners* (last mentioned added to the thesaurus 11/11/1969).

The semantic lexical netword is thus brief. Below the lexical network of *breadwinners* in the five thesauri where the term occurred, is illustrated. It consists of nine different terms with altogether 15 co-occurrences. The most popular co-occurring terms with *breadwinners* were heads of household (with 4 hits) and family roles (with 3 hits).
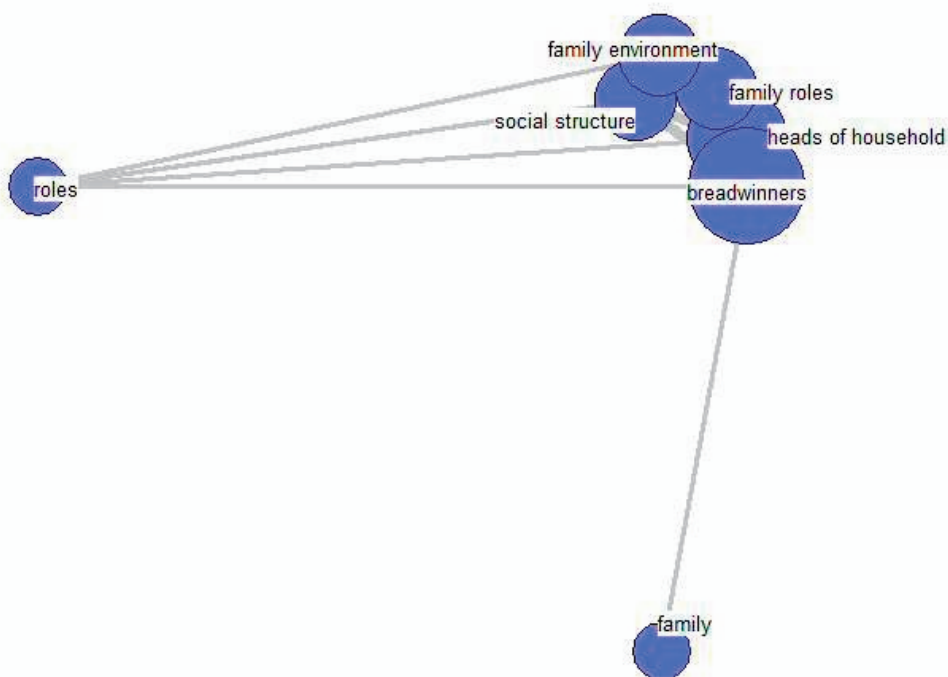
Figure 46: *Breadwinners - lexical network in thesauri*

(in 29.1.2009 with 4406 postings) and in the scope note the term's meaning and use is clarified as: "Group of individuals related by blood, marriage, adoption, or cohabitation (Note: Use a more specific term if possible)"

What about the other thesauri studied, which did not include the keyword *breadwinners* incorporated the thesaurus? UNESCO did not include the studied thesaurus term *breadwinner(s)*. The nearest equivalent found was the broader term *family role* or *parent role*. In the ICPSR the nearest equivalent found was *wage earners*[92] and in the YSA the near-equivalent found was *palkansaajat*[93]/*wage-earners*. Other relevant terms (together with *perheet/families* or *perhe/family*) were *perheenjäsenet/ family members*, *toimeentulo/living*, *subsistence*, *varallisuus/capital, poverty, wealth*, *huoltajat/supporters, providers*. The Thesaurus of Sociological Indexing Terms also did not include the keyword *breadwinner*: the nearest equivalent found was *Heads of Households*.

Was the case similar with *breadwinner*'s near-equivalent **heads of household**? The keyword *Heads of household* was found in five of the studied thesauri: the ELSST, Eurovoc, ERIC, HASSET and (instead of SOSIG) CSA Sociological Thesaurus (see table 3 in Appendix 8).

Below is illustrated the lexical network of *heads of household* in the five thesauri where the term occurred. It consists of 24 different terms and altogether 39 co-occurrences, thus clearly providing more information than *breadwinners*. The most popular co-occurring terms were family (with 3 hits) and households (3 hits).
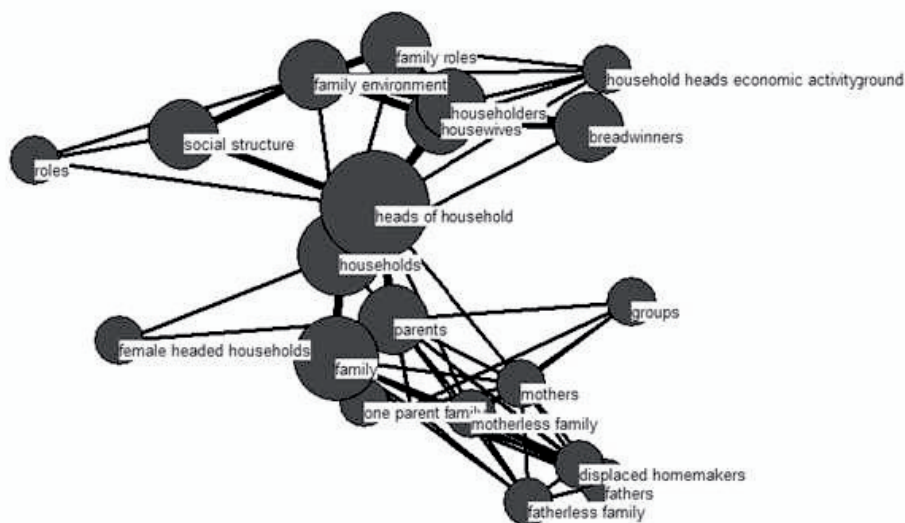
Figure 47: *Heads of household – lexical network in thesauri*

---

[92] RT employment, household income, labor force, wages and salaries. ICPSR includes also terms like *domestic partners, spouses* etc.

[93] Belongs to the thematic group [78] Työelämä. Työsuojelu. Työterveys / Working life. Industrial safety. Occupational health.

The term is represented commonly with statistical terms expressing the structure of the family (e.g. *household heads economic activity*, *lone parent family*). Five of the terms were clearly feminine related (e.g. *female headed households*, *housewives*), whereas only two clearly masculine related (*fathers*, *fatherless family*).

The Finnish Eurovoc equivalent is perheenpää/head of household (MT/BT perhe/family, RT kotitalous/household).

In UNESCO, the nearest equivalent found was (as in the case of *breadwinner*) the broader term *family role* or *parent role*. UNESCO also includes relevant terms such as *parent child relationship* and *households*, but no terms referring to an (economic) provider. In ICPSR, the nearest equivalent found was *wage earners*. In YSA, the case is similar to *breadwinner*: the near-equivalent found was *palkansaajat /wage-earners*, and other relevant terms (together with *perheet/families* or *perhe/family*) were *perheenjäsenet/family members*, *toimeentulo/living*, *subsistence*, *varallisuus/capital*, *poverty*, *wealth*, and *huoltajat /supporters*, *providers*.

Heads of household and *breadwinners* has stayed the same as a thesaurus term within the studied sample during the entire period studied. Only in HASSET has a scope note been added, in 2006, in order to clarify the meaning of the term ("Those whose earnings are the primary source of support for themselves and their dependants", 2006), but otherwise the changes in the studied thesauri were in their nature more structural than conceptual.

9.1.3 Homemakers and Housewives in thesauri

Six of the nine studied thesauri included the term *homemaker*: UNESCO, ELSST, HASSET, CSA Sociological Thesaurus, SOSIG and ERIC. These, except ERIC, are in their origin related to each other. In ELSST and in ERIC the meaning had been clarified in a similar way:

> "Men or women who carry major responsibilities for household or family management" (ERIC)

> "A person, either male or female, whose role is to look after the home and family instead of earning money from employment." (ELSST)

In all the six thesauri *homemaker* is still strongly associated to the female: it either had *women* as a broader term (UNESCO) or *housewives* as a narrower term but no masculine related equivalent (ELSST, HASSET, SOSIG) or was used for *housewives* and had *mothers*, *married women* or *wives* as related terms (UNESCO and Sociological Thesaurus). In the case of *homemaker*, ERIC has already become a more gender-neutral by 1980, when a term *househusbands* was added as a non-preferred term for a preferred term *homemaker*. Similarly, *housewives* (1968-1980) was also changed for a non-preferred term. Other related changes were: *Displaced Homemakers* (added 1980) USE *Females AND Heads of Households*; and *Visiting Homemaker* (added 1966) USE *Housework* (2004). In the thesauri *homemaker* thus even more clearly today represents a family role concept.

Below, the lexical network of *homemakers* is illustrated in the six thesauri where the term occurred. It consists of 34 different terms and the most popular co-occurring terms were housewives (with 5 hits), economic activity (3), family roles (3) and domestic responsibilities (3).
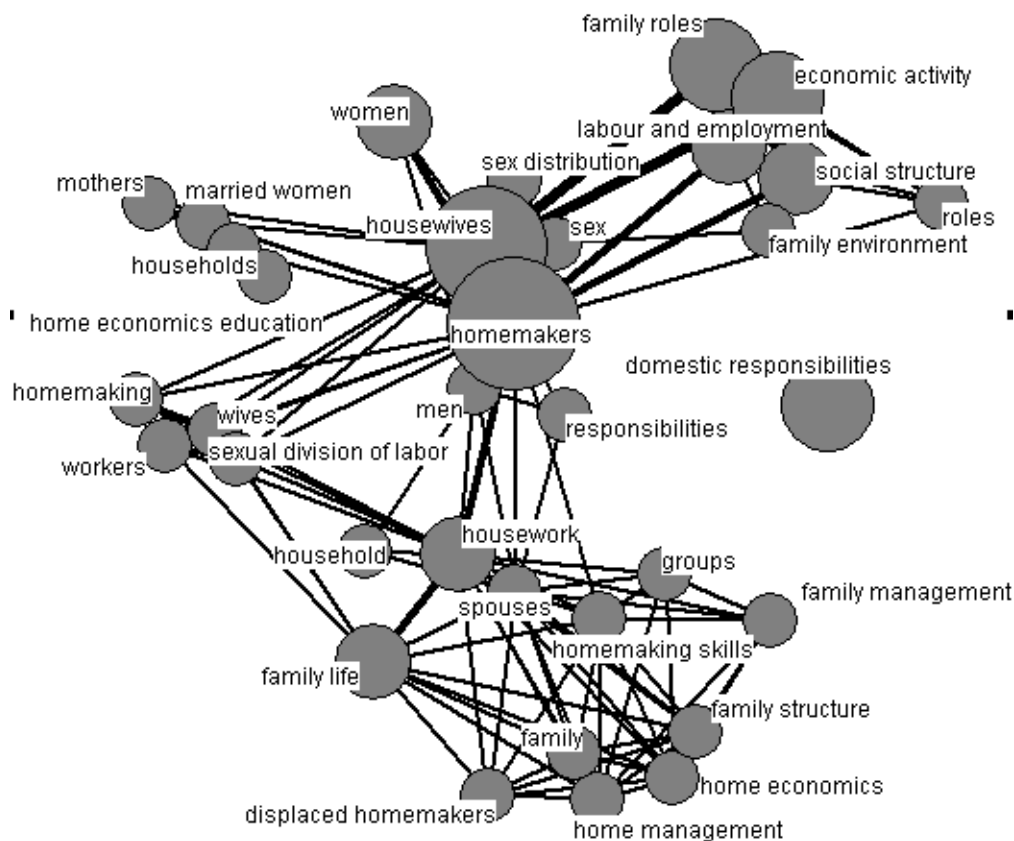
Figure 48: *Homemakers – lexical network in thesauri*

In the studied thesauri *homemakers* is represented unambiguously as a family role. Three significant clusters were found: issues related to 1) housewives, women, men, sex and responsibilities, to 2) economic activity (incl. labour and employment issues) and to 3) housework and family management. It is thus strongly connected to economic activity and to housework. It is represented as a near-equivalent to *housewives* and it refers to women.

The other thesauri were also further studied in order to discover the near-equivalents (within the context of the studied theme, i.e. family role). In Eurovoc, a near-equivalent *housewife* was found (Finnish equivalent *kotirouva*). In ICPSR, the nearest equivalents found were *husbands* and *wives* (BT *spouses* -> BT *family relationships*). YSA does not have equivalents in English. The nearest equivalents found were *koti-isät* (stay-at-home fathers = a father, who stays at home to take care of children and home instead of earning money from employment) and *kotiäidit* (stay-at-home mothers = a mother, who stays at home to take care of children and home instead

of earning money from employment). It also includes relevant terms referring to homemaking (*kotityö/housework*, *domestic work* and *kodinhoito/homecraft*, *household management*, *homemaking*), but terms referring to *lastenhoito/childcare* and *hoivatyö/caregiving*, *nurturing work* were not represented within the context of the family.

Although *homemaker* can be considered in principle as a gender neutral term, in practice it is associated with a female related term in the thesauri studied. From the gender neutral discourse perspective the most modern thesauri studied were, in this case, ICPRS and YSA, both of which did not have an exact equivalent for *homemaker*. – It is noteworthy, that since *homemaker* is here considered as a family role, terms referring to occupational titles etc. (like *housekeepers/kodinhoitajat*[94]) were not considered as relevant and equivalent.

Were there any important changes during the research process? - HASSET has nowadays (02/2009) as a narrower term for *homemakers* also *house husbands* and in addition a similar scope note as in ERIC and in ELSST:

> "A person, either male or female, whose role is to look after the home and family instead of earning money from employment. Term created Sept. 2000. Prior to this the term "domestic responsibilities" might have been used." (HASSET 2009)

Although the term *homemaker* did not have exact equivalents in the Finnish thesaurus studied, it is not alien - although still foreign. In YSA especially, the subject field of a caregiver parent is covered, but from another perspective and thus with other kinds of terms. The emphasis is more on parenthood and not on economic roles as in the English thesauri studied.

Was the case similar with the studied term **housewives**? - There exists great variety among the studied thesauri, and in fact, *housewives,* was the most heterogeneous term in the sample with regard to the existence vs. non-existence and preferred vs. non-preferred term in the studied thesauri.

Four of the nine included the term as a preferred term. In HASSET, ELLST and SOSIG *housewives* has as a broader term *homemakers*. HASSET also has as broader terms *family environment*, *labour and employment*, *roles* and *social structure*. Eurovoc has as a broader term *household*.[95] Within *housewives* these four thesauri have no terms

---

[94] (YSA 29.1.2009): kodinhoitajat / housekeepers: RT kodinhoitoapu / domestic aid, kotiavustajat / home help, kotisisaret / home nannies. Belongs to thematic groups [24] Ravitsemus. Ravitsemisala. Elintarvikeala. Majoitusala. Kotitalous / Nutrition. Nutrition field. Food industry. Accommodation, [73] Sosiaalipolitiikka. Yhteiskuntapolitiikka. Sosiaaliturva. Terveydenhuolto. Asuminen /Social policy. Social security. Health care. Inhabitation, housing.

[95] HASSET: housewives, BT family environment, homemakers, labour and employment, roles, social, RT economic activity, heads of household, mothers.

Incl. also *house husbands*: BT family environment, homemakers labour and employment, roles and social structure , RT economic activity, fathers, heads of household.

ELSST BT homemakers

SOSIG Housewives BT Homemakers

referring to a male equivalent such as *househusbands*, except HASSET (in 2/2009) which names it under *homemakers* as a narrower term.

Three of the nine thesauri studied represent *housewives* as a non-preferred term for *homemakers*. UNESCO and the CSA Sociological thesaurus do not provide any male equivalent, but ERIC also lists *househusbands* as a non-preferred term for *housewives*.[96] In addition, ERIC also includes *Houseparents* as a non-preferred term for *Resident Advisers*[97]. *Houseparents* taken purely linguistically could be used as a broader concept for *housewives* and *househusbands*.

In the ICPSR thesaurus the case is similar to *homemaker* – it does not include terms referring to "stay-at-home –aspect" such as *homemaker*, *housewives*, *househusbands*, *homemother*, *homefather*, *stay-at-home-mothers* etc, but it does include relevant terms such as *caregivers* and *spouses* and as a narrower term it provides *wives* and *husbands* (represented equally in depth and scope). It also includes terms such as *working women*, *working mothers* and *unwed mothers* (without possible male equivalents like *unwed fathers*). It also covers the aspect with relevant concepts like *domestic responsibilities*[98], which do not refer to people, but to tasks. In YSA the case is similar to *homemakers* – the near-equivalents are *kotiäidit/stay-at-home mothers* and *koti-isät/ stay-at-home fathers*.

Similar to *breadwinners* and *heads of household*, *homemakers* and *housewives* were also foreign concepts in the Finnish thesaurus context. The concepts are covered in the Finnish thesaurus context with terms referring to parents and to domestic responsibilities (*kotityö/housework, domestic work* and *kodinhoito/homecraft, household management, homemaking*).

EUROVOC housewife   MT 2816 demography and population , BT1 household, BT2 composition of the population, RT housekeeping economy (1621)

[96] UNESCO USE homemakers

CSA USE homemakers. (Includes also an occupation title *housekeepers*).

ERIC Housewives Use Term Homemakers; Add Date: 03/05/1980

ERIC Househusbands Use Term Homemakers; Add Date: 03/05/1980

[97] ERIC Resident Advisers

Scope Note: Personnel who live with and coordinate the activities of residents of boarding schools, dormitories, college fraternity or sorority houses, care and treatment institutions, children's homes, group homes, or similar establishments. Category: Students, Teachers, School Personnel

Broader Terms: Employees Related Terms: Boarding Schools; College Housing; Dormitories; Group Homes; Pupil Personnel Workers; Residential Care; Residential Institutions; Residential Programs; Residential Schools; Student Personnel Workers;

Used For Cottage Parents; Houseparents; Resident Assistants (2004); Resident Supervisors;

Add Date:.06/21/1983

Postings: 67 (situation in 02/2009)

[98] SCOPE NOTE(S): Applies to maintenance of the home environment. NT housework

RT childcare, working mothers, working women

9.1.4 Equivalence in multilingual thesauri

Most of the studied thesauri were monolingual (see more details in chapter 6.3.2 *Studied thesauri*). The multilingual thesauri studied were the UNESCO Thesaurus, EUROVOC and ELSST, which are all designed to be used in a multicultural environment.[99] Did they provide information about equivalence between different language versions on their web-pages?

As stated earlier, at the time of the analysis, ELSST was not yet published and thus its instructions could not be studied. The UNESCO Thesaurus did not include any kind of information about its equivalence between different language versions. – The only information about different language versions provided was:

> "The UNESCO Thesaurus includes French and Spanish equivalents of English preferred terms. Indexes of French terms (with Spanish and English equivalents) and Spanish terms (with French and English equivalents) have also been provided." (UNESCO Thesaurus 2009)

EUROVOC gives very brief information about its equivalence between different language versions, and also states the way in which it follows standards:

> "All these languages have equal status: each descriptor in one language necessarily matches a descriptor in each of the other languages.
>
> However, there is no equivalence between the non-descriptors in the various languages, as the richness of the vocabulary in each language varies from field to field.
>
> The Eurovoc thesaurus has been compiled in accordance with the standards of the International Standards Organization:
>
>   * ISO 2788-1986 - Guidelines for the establishment and development of monolingual thesauri;
>
>   * ISO 5964-1985 - Guidelines for the establishment and development of multilingual thesauri." (EUROVOC 2009)

In 2009 (22.4.09), all language versions of the Eurovoc thesaurus were stated to comprise 21 fields, 127 microthesauri, 6,645 descriptors (of which 519 are top terms), 6,669 reciprocal hierarchical relationships (BT/NT), and 3,636 reciprocal associative relationships. The fields, microthesauri, descriptors, hierarchical relationships and associative relationships were described to be "strictly equivalent in all languages", and the numbers of non-descriptors and scope notes, on the other hand, to vary from language to language.

---

[99] In addition, YSA also has equivalents in Swedish, and Allärs is stated to be a Swedish translation of YSA. In its nature YSA and Allärs are monocultural, and information about equivalence between different language versions is not provided.

For instance, in the English version there were 759 scope notes and 6,769 "non-descriptors", and in the Finnish version 859 scope notes and 5,445 "non-descriptors". (Ibid.)

The multilingual thesauri studied were in their nature to a certain extent international – the underlying basic assumption about equivalence was that each descriptor has an equivalent in other languages, but the non-descriptors may vary and are dependent on current language version. The dynamic nature of language was thus acknowledged, and formal equivalence was aimed at to a certain degree, but still relatively strictly (cf. Meriläinen 1997, 100; Nida 1964/2000, 129-130).

- **Summary and Conclusions**

In accordance with the construction principles of thesauri (cf. chapter 3.3 *Thesaurus construction*) the studied thesauri represented the terms unambiguously. For example, *homemaker* is also an occupational title and a euphemism (cf. chapter 7 *Family roles in common and scientific language usage*), but its context in the thesaurus was always coherent (cf. chapter 9.1 *Semantic lexical networks in thesauri*).

Radical changes during the studied period 2003-2009 were not found in the way the term were represented in the thesauri studied, except in HASSET, where the term *homemaker* had become more gender-neutral due to (in all likelihood) feedback received within the ELSST project. The largest changes were that some of the studied thesauri had been replaced with other thesauri, and thus the sample became more homogeneous, and it was not anymore possible to see if and how the possible changes were spread in terms of time and style.

The multilingual thesauri studied were in their nature rather international, and the underlying basic assumption about equivalence was that each descriptor has an equivalent in other languages, but the non-descriptors may vary and were dependent on the current language version. Formal equivalence was thus aimed at, but only to a certain degree. The result conforms to the ideas commonly presented in thesaurus construction guides and norms. As discussed in chapters 3.3.3.2 *Equivalence in multilingual thesaurus construction* and 3.4.3 *Equivalence in thesaurus construction* when compared to the main translation strategies such as domestication, foreignisation and internationalisation, the ideology behind the ISO 5964-1985 standards is closest to the idea of internationalisation. In practice this means that the content of the thesauri should be designed so that it is possible to aim at equivalence between different language versions.

9.2 British thesaurus constructors

In order to study how British thesaurus constructors consider the studied case and concept of equivalence in thesaurus construction, six participants experienced in

thesaurus construction were met with, and four interviews were made in England. - In two cases two of the participants were present at one of the interview situations and they represented the same kind of experience. No significant differences between the interviews made with one informant versus the two informants were found, except that the number of words provided in word associations and the reasonings provided and discussed were greater. In these two cases the informants spoke mainly from shared experience gained in specific thesaurus project(s). They thus strengthened what has been previously said by the other participant or complemented the answers given. The minor differences reflected more the attitudes or nuances than actual construction practices.

> "I think you're [refers to the other informant] right. One just tends to be sceptical and say it [exact equivalence] is impossible." (BR TS3a)

The informants interviewed as thesaurus specialists had, in general, higher positions (such as managers) than the indexers interviewed. They were often in charge of content or knowledge management, and had also previously worked as indexers. They represented three different cities and four different institutions, which were or had been involved in thesaurus construction project(s), and are well-known and highly respected in their field. All of the informants were specialised into social sciences.

Two informants (= one interview) were found on the basis of suggestion, made by another informant. Others were found on the basis of their thesaurus experience, known by the author due to previous work tasks in the Finnish Social Science Data Archive and from thesaurus construction literature. Four of the informants were male and two female, but also here gender did not seem to influence the results. In addition, one information specialist, specialised in the technical solutions of thesaurus construction, was met with, but this meeting is not included in the research material. This meeting and guidance concerning the technical problems and solutions was still considered useful.

The interviews took on average ninety minutes (minimum one hour, maximum one hour fifty minutes). All the interviews took place at the informants' work place. Notes were written during the interviews, but interviews were also tape-recorded and transcribed for further analysis. The interviews were made during November-December 2003.

The interview guide is presented in Appendix 6: Interview guide for the British thesaurus constructors.

The construction practices, the studied family roles terminology and equivalence and translation strategies are discussed next, from the perspective of British thesaurus construction.

9.2.1 Construction practices

One of the general research questions is concerned with the possibility to identify different discourses and to see what kind of discourse differences can be found between cultures, subcultures and disciplines; what does this mean for multicultural thesaurus

construction; and what is the pragmatic indexing term equivalence. As stated before (cf. chapters 1.3 *Research questions* and 6 *The Study*) the general research questions are operationalised into several sub-questions:

- How are the studied concepts understood? To what extent are the differences due to institutional versus cultural differences? What is the semantic invariant?

- How is equivalence understood? What do thesaurus constructors, indexers and social scientists aim at in their translations?

Thus, in order to discover British thesaurus construction practices ("are the differences due to institutional versus cultural differences") as a contextual information for further equivalence matters, the informants were asked about their working methods, the (possible) guidelines followed, the terminological sources and the thesaurus role models. In addition, the future of thesauri was discussed in the interviews.

The respondents named or referred to in all nine different thesaurus or controlled vocabulary projects, three of them were shared by more than one participant or (in case of double interviews) institution. Two of the respondents had been directly involved in the ELSST project. Those individuals involved in the ELSST project were asked additional questions regarding that project, and a slightly different interview guide was used (see appendix 6). It is noteworthy, that due to the terminological starting point of the study, experience in ELSST project was considered to be an important factor in the selection of respondents (theoretical sampling, see chapter 6 *Methods and material*), but the informants could have been asked to participate to the study already on the basis of their previous experience with thesauri.

The informants did not represent translating as an ideal way of constructing a multilingual and multicultural thesaurus. The gained experience also tended to favour, in practice, more truly multilingual and multicultural working methods. (Cf. previous chapters and Hudón 1997 & 2004.)

What kinds of working methods were used? There existed a comprehensive consensus regarding thesaurus construction that is always team work, but there existed great differences as regards the level of the decision-making democracy. A centralised structure, in which all decisions are taken by a central agency (cf. ISO 5964 -1985, 32 and chapter 3.3.3.1 General principles and aims of multilingual thesauri), was most commonly referred to. Theaurus projects were considered as long, and often as never-ending, and the working method was also often in flux. Here gained experience also had a great impact on the selected working method, and the more experience, the more common a decentralised or a semi-centralised structure became. In one case referred to the changes actually occurred in these three stages – i.e. first the move from a centralised structure into a decentralised structure, and second a move from a decentralised structure into a semi-centralised structure. In the ISO 5964 -1985 standard (p. 32) it is stated that a decentralised structure is generally regarded as least effective, and a semi-centralised structure as the most effective, and the experiences discussed in the interviews were in accordance with this.

What kinds of comments were received by informants? When there was a question concerning a translation project, the non-British translators were usually "true experts on thesaurus construction, and made a lot of questions concerning definition of the term and asked for scope notes, and it truly made to think about what is the meaning of the term and is it useful in the European context" (BR TS1). Translation processes tended to influence the corpus and a lot of changes were thus also made to the British version. However, many differences existed in the commenting activity, and generally it was considered that the constructors do not receive enough feedback.

What other kinds of problems were mentioned? The other problems mentioned considered the lack of time and resources. The lack of a common understanding was also considered to be a problem. Together with thesaurus construction problems the problems of indexing were also discussed – e.g. how to achieve a common understanding of indexing specificity. Poor appreciation of indexing was also considered to impede thesaurus construction.

What kinds of sources and authorities were mentioned? ISO-standards of thesaurus construction were frequently mentioned, as well as instructions written in the leader organisation. In most of cases the material provided by the institution was considered to be the most influencing factor.

A crucial question in thesaurus construction is how to select a preferred term. Style was considered to be a major factor in selecting a preferred term. Another term might have been more descriptive, but if it was not understood as part of a documentary language it was not generally accepted. In multilingual thesaurus construction they were not necessarily concerned with exact equivalence (example 1 below) (cf. ISO 5964-1985), in which case they wanted the term to be clear for the user of the thesaurus. Otherwise, successful thesaurus was considered to be transparent and minimalistic from the information seekers point of view (see example 2 below).

Example 1:

"And also basically when it deviates from exact match, we have recommended to use what we call a translation scope note, so we are saying this term deviates from the concept of… in certain way. In this language you cannot really find an exact match for this concept, but this is the… ---" (BR TS1)

Example 2:

"So we had two versions, one is for the cataloguers, so to the people who do the indexing, and that's the full version of the thesaurus, and then we have another version that we make available on the public site, and that is all of the terms that appear in the records that have been indexed by [us], so it's a cut-down version. So we run the index through all the records that we got and all of the terms, the related terms appearing in the public version, which means that somebody will never find a term that hasn't have anything related to it, which is adequate ---." (BR TS4)

One of the themes in the interviews was the usability and future of thesauri. Generally, the British thesaurus specialists interviewed believed in the future of thesauri – they will also be needed in the future and the more complex the using environment becomes, the more necessary the thesauri become. It was further stated, that

paradoxically, the more necessary thesauri will become, the more difficult the issue of their existence becomes. Some of the problems were due to differences in cultures and in languages and the problems were also considered to be related to language competence.

"If you want precise results, a thesaurus is absolutely essential. --- I think it's useful for standardisation across languages, but you have to be very careful to that you're not putting meanings into the boxes where they shouldn't be. I mean it's a standardisation problem. ---

The thing is that if you weren't very well versed in the sort of target language, you would not feel comfortable about to say I would index in English and therefore I have this thesaurus and therefore I can also index in Finnish --- You wouldn't feel confident, even if it was a good thesaurus. --- I mean it must be a good thing but not for abstracting, because there must be this huge error - possibility of errors [laughing] --- It's a good thing if it comes, if it's been constructed in consensus of different language speakers and each language speaker uses the version that's been read-up on in their language. But then I don't know about the question of translatability, it's a double-edge source, isn't it…" (BR TS3)

In the context of multilingual thesaurus construction some of the problems were also related to the role and function of thesauri in general (example 1 below), such as to the crucial question of transparency and different aspects of user-friendliness (example 1 and 2 below).

Example 1:

"I think they are, but I think that people need perhaps to know that they are using them or that they are being used. --- I think we still got a problem of understanding what a thesaurus does, so I think ideally they would been used behind the scenes, so if somebody typed there a term it would automatically go off and look for thesauri and say that term also means this, this, this and this, and do the search for those terms, but then you'd got the problem of when you return the records the user is seeing my term isn't in there, why is this record been returned to me, so it's quite tricky, yeah." (BR TS4)

Example 2:

"[Thesauri are] needed also in the future, people are becoming more aware of them. The thing is that they want them hidden, people want them to be there, they would prefer not to know they are using it, well they know there is something behind the scenes, something matching up the concepts…" (BR TS1)

More guidance as to the idea of thesauri for all the user groups was requested, and often for the information seekers, because although very useful, they were also considered as demanding tools – "using a thesaurus is an advanced skill [in information seeking]" (BR TS3).

The respondents gave several suggestions of advice to the thesaurus constructors, based on their experience. The need for scope notes was emphasised the most.

"I would definitely put the scope notes. I mean, I did realise that, I think that even though you think you know exactly what a term means and everyone else understands it, they could easily be – themselves thinking they are understanding it and translate it

the way they think and that comes back after a lot of times… --- Lots of scope notes to always make it very clear." (BR TS1)

"It was also a joint emphasise on the need for close co-operation between indexers – "indexers are needed to keep the thesaurus up-to-date" (BR TS4), and because of indexing not to include too similar descriptors (in terms of meaning) into a thesaurus, because "then the indexing process becomes too labour-intensive and interindexer inconsistency greater" (BR TS4).

9.2.2 Family Roles terms as used by British thesaurus constructors

In order to study pragmatic indexing term equivalence, the understanding and translating of terms related to family roles were also studied among the British thesaurus specialists interviewed.

Component analysis was used when relevant. The aim was to gain more knowledge on equivalence matters by a triangulation of method, data and theory (cf. chapter 6 *Methods and Material*). – In the previous subchapter the participants talked freely about translation matters and the concrete cases discussed were raised by the informants, however, in this section the participants have been asked directly how they would translate certain family role terms.

The first introduction to the studied terms was word associations. After the word associations (for family role terms as stimuli words) definitions and aims for the translations were asked for, and then a discussion followed as to how the terms were represented in the working version of ELSST.

9.2.2.1 Family roles by British thesaurus constructors

The stimulus word *family roles* resulted in 37 associations, out of which 33 were different. The consistency on the theme level was greater than on the word level (cf. Iivonen 1989), and e.g. children were covered with different kinds of words (such as *son, daughter, siblings, child* and also *childcare, sibling rivalry, parent-child relations*).
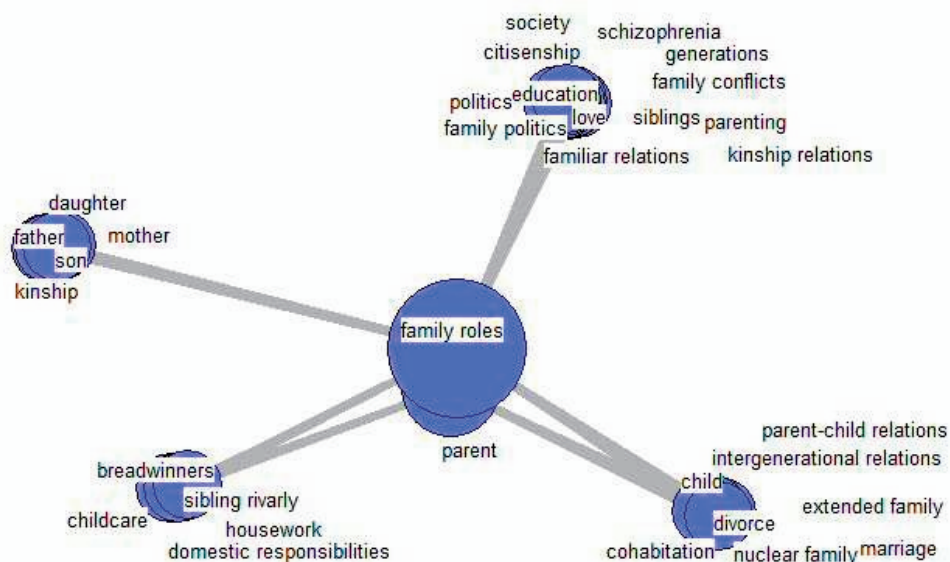
Figure 49: *Word associations for "family roles" by British thesaurus specialists*

Family roles was normally associated with the idea of a nuclear family i.e. having parents and children as family members. Extended family and kinship was also commonly present as a subsidiary. Issues related to adults (such as *breadwinners* and *marriage*) and children (such as *daughter*, *son*, *siblings*) were emphasised most. Family roles also associated clearly on societal level (response words such as *education, citisenship*). The connotations were mostly neutral, but negative connotations (such as *schizophrenia, divorce, sibling rivalry*) were more common than positive connotations (*love*). The connotations provided more information about the semantic meaning of the word, than about the tone or style of the requested word.

It is noteworthy that word associations provided by thesaurus specialists were not – as with the other groups studied - given as thesaurus terms. As one informat stated, they were "things that are in my head but which I wouldn't necessarily use in a thesaurus" (BR TS 4).

As a term and descriptor *family roles* was commonly understood to refer to roles within a family, where parents have different responsibilities. It was commonly considered to reflect the past, because today the different roles are more frequently shared by both sexes. It was not considered to be difficult to translate as a thesaurus term, but it was usually stated, that in different cultures and subcultures it may carry different kinds of connotations and contents – e.g. the role of kinship may be greater.

9.2.2.2 Breadwinners and Heads of household by British thesaurus constructors

The stimulus word ***breadwinners*** resulted in 37 associations, out of which 33 were different. – As in the previous cases of *family roles*, the associations provided by the British thesaurus specialist, the consistency was greater on a theme level than on the word level (cf. Iivonen 1989), and e.g. wage-earning was covered with different kinds of words (such as *primary wage-earner, financial responsibility, greatest income, primary earner, wage-earners, providers*).
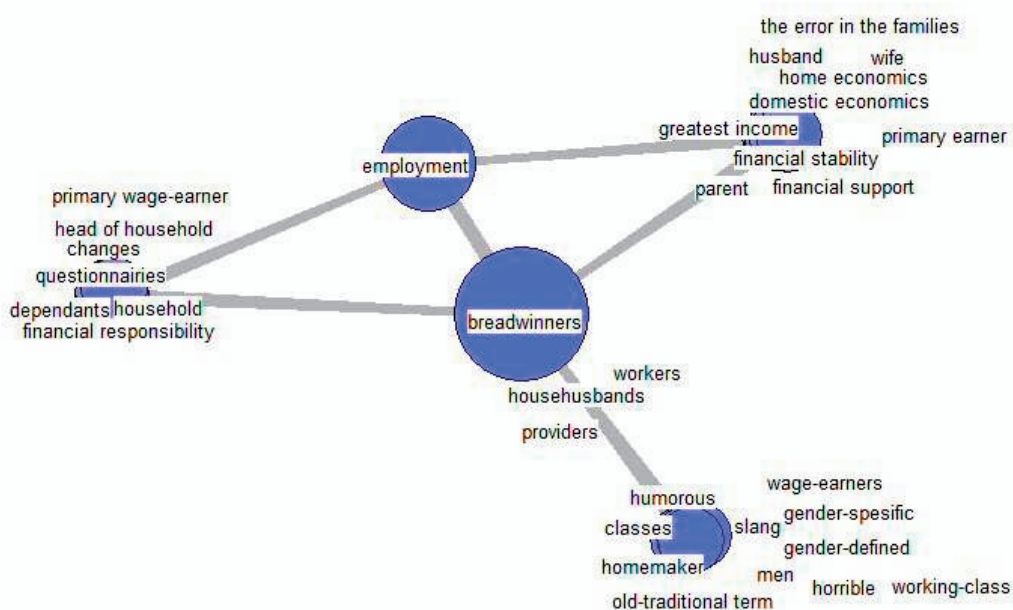
Figure 50: *Word associations for "breadwinners" by British thesaurus specialists*

Breadwinners resulted in stronger connotations than *family roles*, and interestingly some of the associations (such as *humorous, horrible*) directly concerned the connotative meaning of the stimulus word. Its usability as a descriptor was also discussed more actively. As a descriptor it was generally considered to be out-dated, but also useful in seeking older material.

> "This terminology changes over the years. When I take a look at our questionnaires we don't get breadwinners anymore the way we used to. Head of household is also that kind of variable which meaning has changed quite a bit over the years." (BR TS1)

According to the informants *breadwinners* refers to wage-earners with the greatest income within a family. Here also the associations reflect the idea of a nuclear family as a starting point. *Breadwinner* was also considered to be an opposite for *homemaker*

("everything that isn't homemaker, the opposite" BR TS2) and *househusband* ("as an opposite" BR TS4).

How close to *breadwinners* is **heads of household**? The stimulus word *heads of household* resulted in 25 different word associations with 30 co-occurrences, and here the consistency was also greater on a theme level:
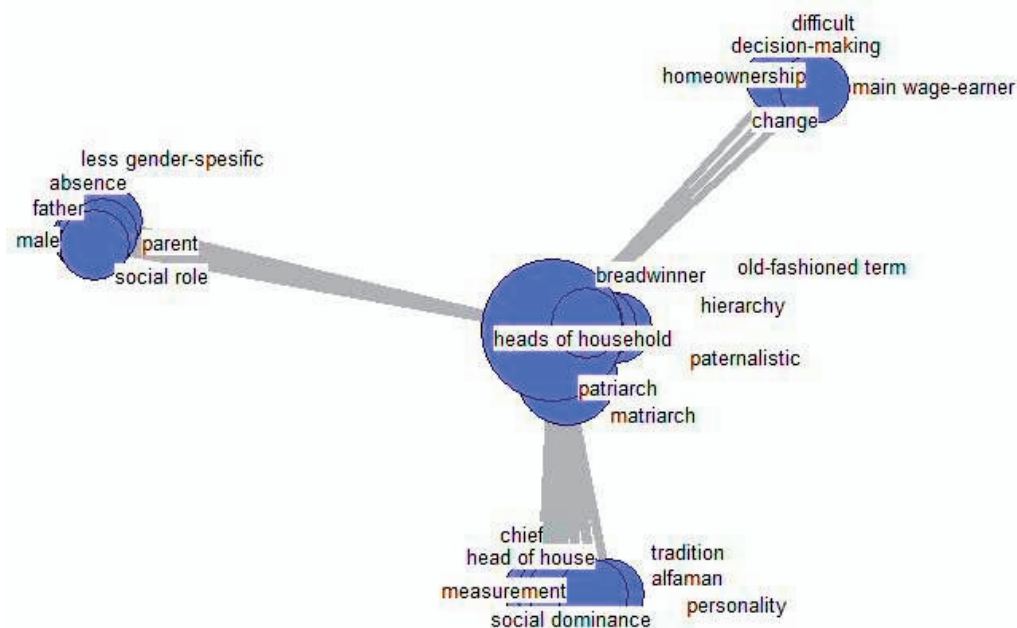
Figure 51: *Word associations for "heads of household" by British thesaurus specialists*

The stimulus word *heads of household* often associated with strong connotations and was understood as old-fashioned (- "it's not even me that comes to my mind, it's my father, so from my experience in my childhood" BR TS2) and a sexist term (such as *alfa-man*, *male parent*). As a descriptor *heads of household* was usually considered to be more useful and neutral (*less gender specific*) also today than *breadwinners* - "they still have it in most of the big surveys" (BR TS1). One informant suggested that the proper equivalent and preferred term for it would be *breadwinners*.

In this case the distinction between different kinds of words was especially essential. – The word was understood as negative if considered as a common language word and as a concept reflecting practices within a family and society, but as a descriptor it was considered as necessary and more neutral. Here also the function (cf. chapter 3.4.1.1 Function of translation) of the word played a crucial role in the acceptance of the studied term.

9.2.2.3 Homemakers and Housewives by British thesaurus constructors

As a stimulus word *homemaker* resulted in 37 different word associations, out of which 32 had different co-occurrences. As in the previous cases, the consistency was greater on a theme level, and the aspect of stay-at-home mother was covered with several different kinds of word, which were synonymous or near-synonymous (*mother, housewives, home-mother*):
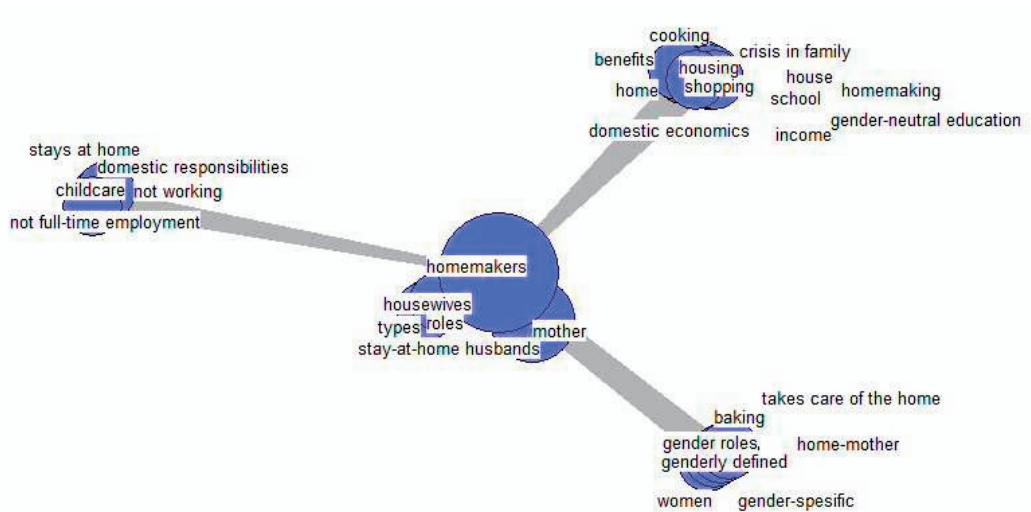
Figure 52: *Word associations for "homemakers" by British thesaurus specialists*

As a word this term connotated in a rather neutral way. Pragmatically it was still considered to be artificial, because it was considered very exceptional for men to stay at home in order to take care of the home and children instead of earning money from employment. It was thus considered to represent an idealistic term, not a term reflecting reality, and thus its style was operative instead of informative (cf. Vehmas-Lehto 1999). As stated before, different types of meaning are related to different types of text functions (ibid, 73), and here the selection criteria for the preferred-term is to promote gender-neutrality.

In the case of *homemakers* the British informants (also in the other groups studied) tended to also emphasise the nature of associations – they are general and not necessarily personal ("Wife. This is not my personal... I'm not suggesting that my wife would do nothing but take care of the home" BR TS2).

How was the stimulus word **housewives** understood? As a stimulus word it resulted in 28 different word associations with 37 co-occurrences:
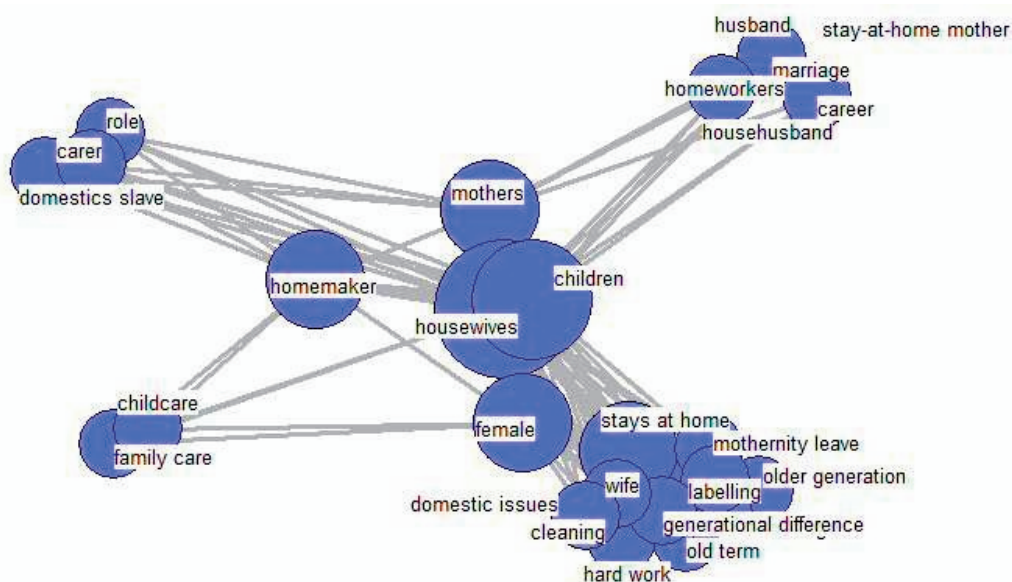
Figure 53: *Word associations for "housewives" by British thesaurus specialists*

The term *housewives* produced the greatest consensus among the British thesaurus specialists. A closer look (with the help of component analysis) at the semantic meaning provided in definitions by the respondents also showed the result as rather homogenous. It was generally considered to be a modern term to refer to stay-at-home mothers with children (see table below).

housewives	BR-T1 (double)	BR- T2	BR-T3 (double)	BR-T4
married	++	++	++	++
non-working	++	++	++	++
female	++	++	++	++
domestic responsibilities	++	++	++	++
childcare	+	+	+	+
mother	+	+	+	+

Table 35: *Semantic chracateristics of housewife according to British thesaurus specialists*

Although in practice (cf. semantic meaning) *housewives* was considered to be very close to *stay-at-home mothers* and as a term to be old-fashioned, stay-at-home mother was still not considered to be a sustainable alternative. - As a term *stay-at-home mother* was not considered to represent a documentary language style ("Media uses it, it is not a proper term, but it can be probably politically correct term, yes". (BR TS3), cf. Aitchison 1984; TSK 1989; Varantola 1990; Wierzbicka 1997; Lehtonen 2000; Katan 2004; Forsman 2005 and chapter 3.1 *On Language, Culture and Meaning*). One suggestion for a modern near-synonym for *housewives* and *homemakers* is *homeworkers*.

"I'm not necessarily directly following but, media is increasingly talking about homeworkers – who work from home, they take some of the tasks of househusband or housewife and combine it with a career." (BR TS3)

Semantically and pragmatically *housewives* was generally considered more of a total role than *homemaker*, and also a more serious role:

"The moment the wife does go to work, even if it is a part time, I wouldn't classify her as a housewife. A housewife has to be a hundred percent. ---

There is something... In comparison to homemakers a housewife sounds like a hard work, sounds like there is an awfully lot of cleaning and... really running a house" (BR TS3)

9.2.3 Equivalence and translation strategies by British thesaurus constructors

One of the research questions and themes is equivalence, which lead into such questions as how is equivalence understood, and what do British thesaurus constructors aim at in their translations. This theme was covered by directly asking how equivalence has been understood and possibly defined in thesaurus projects, and what the aim of different translations has been or how they would define equivalence in multilingual thesaurus contexts. This topic has also been mentioned in other contexts throughout the interviews, such as why multilingual thesauri are needed.

It is again noteworthy that the British thesaurus specialist presented another view to that presented by the Finnish thesaurus specialists interviewed. In all the cases, English had had a major role – as the only language, the exchange language or often as the source language in the thesaurus project, and at least as a common working language. Most of the participants had been working monolingually and/or been in charge of a monolingual corpus, and thus the material represents a more corpus-based view or a general (monolingual) thesaurus construction. Thus the British and Finnish respondents do not represent true comparison pairs, but the aim of interviewing different relevant groups and conducting a triangulation of methods and material, is to gain a holistic view (cf. chapters 5 and 6).

What was aimed at in translations? How did the informants understand equivalence in a multilingual thesaurus construction context?

Each participant stated that the translation unit is a concept, and not a term – "a term that actually is the same concept in other languages" (BR TS4). The participants did not usually refer directly to ISO standards, but they tended to talk in terms common to the ISO 5964-1985 standard, and also to follow the guidelines for preference of equivalence degrees. This means that if an exact match is not possible, then they should select a broader equivalent, and then a "possible combination of two narrower terms and so on" (BR TS1). Non-equivalence was seldom met. What differed from the standards was that it was considered natural that in multilingual thesauri one language version could also include foreign language terms.The dynamic nature of languages and translation was clearly acknowledged (cf. Nida 1964/2000; Vehmas-Lehto 1999).

"In our guidelines for use by the other countries, we are trying to interpret the concepts. We have set out some rules for them to follow, so that their first aim is to try and get an exact match, obviously, and if that is impossible then maybe a broader term, and then possibly a combination of two narrower terms, etc. Going down to a paraphrase or… And then actually, I think they are actually using the English term… Even not necessarily an English term, but using another… So it could be for example a French term…" (BR TS1)

Together with the dynamic nature of languages and translations the text type was also frequently discussed – thesauri represent a special kind of discourse (cf. Reiss 1981/2000; Buckland 1999; Lehtonen 1994; Potter & Wetherell 1990; Pälli 2003), and the lack of a common understanding of the nature of a thesaurus causes problems.

"Depending on where the thesaurus is going to be used for, but I think that also… some of the comments that I have had from some people, I think is, in my opinion they don't clearly understand much about thesauri, and what's so --- thesaurus is a retrieval tool, it doesn't matter if the synonym isn't exact, this is not a dictionary, it's not an encyclopedia, it's a tool for searching. And I think people don't understand that. And I think that can be a problem. I mean everyone working with it must understand what a thesaurus is, what it is for." (BR TS1)

Even though there was a shared understanding as to what a thesaurus is, the informants also considered different discourses to be a problem, and social sciences were considered as an especially difficult thesaurus field because of its complexity and extent. The field also demands a lot of conceptual expertise, which cannot be fulfilled by a single person. (The aim was to choose terms which are in accordance with the practice.)

"There are so many associative meanings, not only... The social sciences don't hang together at all, we find differences between the sociologists, the anthropologist, the psychologists, they can all use the same term in a very different way, so equivalence is quite difficult, and quite dangerous as well, because if you think of concept equivalence, but it can actually mean something completely different... I mean, we lie on our subject specialists' knowledge, when choosing a term --- that they use in community". (BR TS4)

As in the Skopos-theory (cf. Reiss & Vermeer 1989 and chapter 3.4.2.2 SKOPOS-theory) translating was considered to be a decision-making process. During this process the aims also become more realistic and influence the content and structure of thesauri.

"So when a deviation occurs, they are recorded. I think, again, maybe that is another change from… during the project, that we came in that we naively thought that we would be able to have exact matches --- [laughing] that came soon disagreed I think." (BT TS1)

Perhaps due to the previous experience of the respondents, which was mainly on constructing a (potential) translation corpus, **existential equivalence** (cf. Koskinen 2000, 83) was never discussed. Together with ideas about translation as a decision-making process the idea of **consensual equivalence** (cf. 9.3.2) was often accepted, which also means selecting a translation strategy.

"It is going to be certain decision point, where you say, ok, there are different connotations between one language and one understanding of the other, if you want

this equivalence relationship you better say at a certain point this is ok. --- " (BR TS3a) "I suppose it depends how precise you want to be. If you are looking at an exact equivalence with all the social-cultural import, maybe, which is behind the word or... --- But obviously the more precise you want to be the more difficult is going to be the thesaurus construction." (BR TS3b)

Denotative meanings were referred to, and together with these the pragmatic and connotative aspects of equivalence were often discussed (cf. semantic versus communicative translations in Newmark 1988 and chapter 3.4.2.2 *SKOPOS-theory*), and the fact that the different aspects were usually not clearly prioritised - "Terms of equal meaning and equal importance with... Well, I don't think equal meaning, it's not that, it's... similar connotations. ---" (BR TS2).

The respondents did not directly favour any translation strategy (cf. chapter 3.4.1.3 *Domestication or foreignisation – or internationalisation*), but in the examples given they tended to favour internationalisation and foreignisation. Nevertheless, it was customary to accept different styles and motives – as long as the different meanings were reported in e.g. translation scope notes.

> "We had a… short-time working, which for us meant that, we did have a scope note, well for us it meant that people, when their hours of working is reduced because of economic time, because of business conditions, there is not enough work, they have to go down to narrow the time or three day time week --- And I think when, I think it was [a Scandinavian partner], to them short time working meant, you know, a woman's choice to reduce the hours when she had children, and that is something completely different --- So the same term, if we hadn't discussed about it, it would have translated in a completely different way." (BR TS1) [100]

Were the difficulties due to linguistic or cultural reasons? The British thesaurus constructors acknowledged both cultural and linguistic reasons, but they spoke most often about different discources. Time aspect was often considered to be a problem (cf. Forsman 2005).

> "Thesauri don't reflect current media, there are these terms that the media use, and by the time they get into thesauri they have probably moved into another buzzterm." (BR TS4)

- **Summary and Conclusions**

The Bristish thesaurus specialists were rather unified when discussing the style and meaning of the studied concepts and terms. When opinions or views differed the variation occurred mostly on the level of style and not of meaning.

---

[100] "Short-time working occurs when employees are laid off for a number of contractual days each week, or for a number of hours during a working day."

Acas 16.6.2009, ACAS, Advisory, Conciliation and Arbitration Service: "Lay-offs and short-time working", Retrieved from the World Wide Web, URL <http://www.acas.org.uk/index.aspx?articleid=1639>

The British thesaurus constructors acknowledged both cultural and linguistic reasons for making the construction of multilingual and multicultural thesauri difficult. The terms which were considered most difficult to translate into Finnish by the Finnish informants interviewed (cf. chapters 7.2.2, 8.4.2 and 9.3.3) were also considered to be complicated by the British thesaurus specialists. The difficulty was often due to different discourses, as well as due to time, which changes the meaning and/or also the style of the words (cf. Aitchison 1984; TSK 1989; Varantola 1990; Wierzbicka 1997; Buckland 1999; Lehtonen 2000; Katan 2004; Forsman 2005).

Together with the dynamic nature of languages and translations the British thesaurus specialists also frequently discussed the text type – thesauri represent a special kind of discourse (cf. Reiss 1981/2000; Buckland 1999; Lehtonen 1994; Potter & Wetherell 1990; Pälli 2003), which causes problems in thesaurus construction. Furthermore, eventhough the users of a thesaurus would share an understanding of what a thesaurus is, the respondents considered different discourses to also be a problem. Additionally, social sciences were considered to be an especially difficult thesaurus field, because of their extent and complexity, and to demand a considerable amount of conceptual expertise, which cannot be supplied by one person or indeed even several people.

Each informant stated that the translation unit is a concept, and not a term. Different levels of equivalence were considered as an unavoidable reality. Multilingual thesauri were seen as being possibly multilingual also within one language - one language version could also include foreign language terms.The dynamic nature of languages and translation was thus clearly acknowledged (cf. Nida 1964/2000; Vehmas-Lehto 1999).

As in the Skopos-theory (cf. Reiss & Vermeer 1989 and chapter 3.4.2.2 SKOPOS-theory) translating was considered to be decision-making process, and during the process the aims were considered to become more realistic and also influenced the content and structure of thesauri.

Perhaps due to the previous experience of the respondents, which was mainly on constructing a (potential) translation corpus, **existential equivalence** (cf. Koskinen 2000, 83) was never discussed. Together with ideas about translation as a decision-making process the idea of **consensual equivalence** (cf. 9.3.2) was often accepted, which also means selecting a translation strategy.

Denotative meanings were referred to, but often in the context of pragmatic and connotative aspects of equivalence (cf. semantic versus communicative translations in Newmark 1988 and chapter 3.4.2.2 *SKOPOS-theory*). The usability issues (which represent pragmatic aspect) were prioritised, but otherwise the different aspects were usually not clearly put in any order of importance. The respondents clearly did not favour any translation strategy (cf. chapter 3.4.1.3 *Domestication or foreignisation – or internationalisation*), but in the examples given they tended to favour internationalisation and foreignisation. Nonetheless, it was considered normal to accept different styles and motives – as long as the different meanings were reported in e.g. translation scope notes.

The British thesaurus specialists generally prioritised usability in indexing and in information retrieval. They were not aiming at equivalence with dictionaries, but emphasised aspects of pragmatic equivalence – predictability, user-friendliness, coherence with the indexed and retrieved in material. The greatest challenge lay in the understanding of the text type and function of thesauri – and whether information seekers interpreted the thesaurus in similar way to the indexers.

In word associations the differences were not great. At the broader level most of the aspects expressed by the British thesaurus specialists were shared, but expressed with different kinds of words (cf. Iivonen 1989). The major difference was that in pair-interviews (two cases in which there were two participants interviewed simultaneously) the number of word-associations and views tended to be greater than in the other two interviews with two informants.

9.3 Finnish thesaurus constructors

In order to study how Finnish thesaurus constructors consider the studied case and concept of equivalence in thesaurus construction, five interviews were made. The participants represented five separate and well-known organisations in three different cities. The interviews took, on average, two hours (the minimum was one hour fifty minutes, maximum two hours fifteen minutes). The interviews took place at the participants' place of work. Notes were written during the interviews, but the interviews were also tape-recorded and transcribed for further analysis.

Accordingly to the aims of the theoretical sampling, the Finnish informants were experienced information specialists or indexers, who also had experience with thesaurus construction. All the participants had been involved in multilingual thesaurus construction, although it was more usual for them to work monolingually. Four of the five participants were directly specialised in social sciences and also to the Finnish-English language pair in their other work tasks.

It was relatively easy to locate Finnish information specialists with experience of thesaurus construction, based on earlier experience gained through the author's previous work tasks at the Finnish Social Science Data Archive (FSD) and at the thesaurus seminar held in 2001 by FSD. The response to the interview requests was positive. One person, who was considered very relevant to the study, was unfortunately not avalailable because of his/her leave of absence. (There were also other relevant and possible respondents, who were not contacted because of the long distances to their place of work.) The interviews were conducted during November2003 - June 2004.

In the following is a discussion of how the Finnish thesaurus constructors considered the studied terminological case family roles, what kind of practices they have followed in their thesaurus construction project, how equivalence has been understood and what translation strategies were followed, how they see the future of thesauri and what kind of advice they would provide for future multilingual thesaurus work. The

interview guides are represented in appendix 5: *Interview guide for the Finnish thesaurus constructors*.

9.3.1 Construction practices

When studying the construction practices (as a contextual information for further equivalence matters), the themes in the material concerned working method and especially team work, translation as a construction method, comments needed and received during the process, the authorities referred to and the sources used. In addition, the future of thesauri was briefly discussed in the interviews.

In order to find out the practices of Finnish thesaurus construction, the participants were asked about their working methods, the (possible) guidelines they followed, terminological sources and thesaurus role models. Eleven different thesauri or controlled vocabulary projects were named and referred to, three of them shared by more than one informant. Each of the informants had been involved in some way in the construction of YSA, i.e. as a constructor, member of a think tank, commentator or at least as a proposer of a new term to be added in YSA. Three of the informants had been directly involved in the ELSST project, and in addition one had been involved indirectly. The three directly involved in the ELSST project were asked additional questions regarding that particular project, and a slightly different interview guide was used (see appendix 5 theme 7). It is noteworthy, that due to the terminological starting point of the study, experience in the ELSST project was considered to be an important factor in the selection of participants (theoretical sampling, see chapter 6 *Methods and material*), but two of the three participants would have been asked to participate to the study purely on the basis of their previous experience with thesauri.

What were the working methods used? A comprehensive consensus existed regarding the fact that thesaurus construction must always be team work:

> "No single person can see all the nuances. But it's often terribly difficult to choose what would be the wise thing to do." (FI TS1)

> "--- No single country can construct a multilingual vocabulary without leaving there culture- or system-bound terms, because they cannot see them." (FI TS5)

> "When working alone, the perspective always becomes distorted, there are no others who would put one's foot down, discuss what about this and that, does it cover also this and does it truly mean that." (FI TS3)

According to the informants not only collegial support is needed, but also co-operation with users (i.e. information seekers and indexers) and terminological specialists are important. If the thesaurus is also meant to be used multiculturally, native speakers of all the language versions are needed too. This principle is in line with the guidelines presented in the standards and guidebooks, which indicate that close cooperation is essential for the successful construction of thesauri (cf. chapter. 3.3 *Thesaurus Construction*). In practice, the multilingual projects were of two kinds –

translation projects aiming at multicultural or international content, and multilingual projects, reflecting one culture – i.e. the content was monoculturally Finnish.

The common working method in multilingual projects was similar to that of a centralised structure, in which all the decisions are taken by a central agency (cf. ISO 5964-1985, 32 and chapter 3.3.3.1 *General principles and aims of thesauri*). The smaller or culturally homogeneous the project the more flexible the method used. In monolingual and/or monocultural projects the favoured method was similar to a semi-centralised structure, in which the work and responsibility is divided carefully, but the final responsibility remains in the hands of one authority. It was rare, but it sometimes occurred that the organisational structure changed from a centralised or de-centralised structure to a semi-centralised structure during the project due to increased experience. The centralised structure was, in practise, the result of the selected construction method, which was translating a previously constructed corpus. In this case, the final authority was often the constructor of the material to be translated (see example 1 below). In practice the most difficult challenge was that all the material was not necessarily received at the beginning of the project, which made the process (especially selection of term meanings) even more difficult (see examples 2 and 3 below).

Example 1:

"The situation was even worse in the EU project, because there we had ready and badly collected material – because it guides the perspective in a specific way, there is no unbiased look when one starts… So in the EU project there were at the bottom a project of four languages --- and the starting point was a ready English-language thesaurus. Then the four groups had been able to do it that way that they had truly been sitting around the same table and discussed the thesaurus term by term. Then --- two more groups of four languages were established, and we were bound to the English-language culture, when we were engaged in the project, because the English version had guided the whole. One could see it also in the [other language] versions, that there were some parts translated willy-nilly. But we had a good [sub]project, North-European, there it was easy to discuss also culturally." (FI TS2)

Example 2:

"(S)he had organised the process so that we would have proper sections before every meeting [in our subproject]. But the only problem was that we did not receive fast enough translation material from England, we could have progressed faster --- the slow delivery from England was the most difficult part." (FI TS3)

Example 3:

"[When the translation corpus was received in pieces there were problems] in outlining the relationships, you can see it when you have the whole hierarchy, but one can make truly misleading conclusions on a basis of a single word." (FI TS5)

As discussed previously in chapter 3.3.3 *Multilingual thesauri*, the ISO 5964 - *Guidelines for the Establishment and Development of Multilingual Thesauri* recognises three approaches to the construction of multilingual thesauri – the establishment of a new multilingual vocabulary without direct reference to the terms or structure of an existing thesaurus, the translation of an existing monolingual thesaurus and the

reconciliation and merging of existing thesauri in two or more working languages. How are these forms used in the Finnish projects studied?

When asked about their experience of constructing multilingual thesauri (possibly also with the aim of having a multicultural content), the interviewed Finnish thesaurus constructors realted that they had mostly been involved in translation projects. Furthermore, despite all the disadvantages of this method, it was also considered to be the most common way to construct a multicultural thesaurus. The reasons mentioned for this were economical and organisational – it saves time and money, and is usually already selected as a working form, when the Finnish specialists step into the project. However, when constructing mono- or multilingual thesauri to be used in Finland, the working form was more aligned with that of the establishment of new thesaurus or vocabulary, although existing thesauri (especially YSA) also played a significant role as terminological sources.

There was still a clear gap between the ideal and the real situation, which was discussed by each of the participants either as generalisation (example 1 below), or a specific experience (example 2 below) and/or literature (example 3 below).

Example 1:

"Should prioritise, get proper organisations, should be invested so that people representing different expertises could be hired. At the moment one feels herself terribly deficient. One can only touch upon the whole. The work itself feels important, so the deficiency comes with the amount of work. There should be a different network, but that requires money, so that people could be engaged full-time." (FI TS1)

Example 2:

"[Our collection-specific controlled vocabulary] is aimed to be three-lingual. The Finnish basic vocabulary is constructed first. But it would have been good, if all the three vocabularies [language versions] would have been developed side by side, but it is a question of resources, we just absolutely don't have enough gang for that." (FI TS2)

Example 3:

"If you think of these thesaurus construction guidelines and the principles they are based on, so in them the real multilingual thesaurus is not based on any specific language, but it is constructed so that all the languages are equal, and there is one working language, but no language is actually the starting point. This would be a starting point, in which you don't have these problems that are bound to culture and to just one kind of system --- (FI TS4)

Cooperation was also frequently regarded as a problem – there is not enough co-operation, the network is too small and/or there is too little time for commenting, - and most often, the thesaurus constructors do not receive enough feedback.

"I sent [the working version of thesaurus] for commentary, but received quite few comments. But in addition I used the in-house specialists, with whom we discussed the most difficult terms." (FI TS2)

What kinds of comments were received by informants? As stated before, each of the five informants emphasised the importance of wide and deep co-operation and also mentioned that comments are needed in order to construct a usable thesaurus.

The comments received most concerned style or more precise terms proposed by the social scientists (used as terminological experts). In the discussions and given examples the style changes were mostly related to terms expressing people (see example 1 below). The desire for more precise terms were often in those cases where the thesaurus constructor(s) had chosen an equivalent, which would be in harmony with YSA, and thus was considered by them to be an already established indexing term. Since the subject specialists were most often concerned with concepts or terms, the participants seldom received comments or suggestions regarding the structure of the constructed thesaurus, the place of a term in a hierarchy or thematic section etc. – The term's place in thesaurus was raised in two kinds of contexts by the participants: a) the translation corpus was received in sections and therefore it was difficult to judge the term's meaning and relationship, which often changed when the rest of the material was received, b) changes in time also affects the content of terms (example 2 below).

Example 1:

"Well, I was thinking that something of these… that [the comments concerned] usually more correct language use. That our strategy was similar to newspaper style. Of course we did not suggest anything like criminal control ["rikollista torjuntaa"], but, anyway, that it must be crime prevention ["rikosten torjuntaa"]. And there must not be terms that diminish or label people. - They were the most important points. We did not receive feedback concerning the use… That the structure of the vocabulary would have shown to the researchers that we use a wrong term, that it does not mean that, so these kinds of comments I don't remember us receiving. So it was nearest about language use that won't stamp people. But then, in our fields it is the sensitive point. ---

So yes, the [potential descriptors with their relationships] were given to our lectures – lecturers of sociology and social policy – they read them and yes, they commented, and especially, well not maybe so much the hierarchies, but the language use." (FI TS4)

Example 2:

"When the meanings of the words change, we must also make changes to its relationships and thematic groups [in the thesaurus]" (FI TS1)

What kinds of other problems were mentioned? It was common to mention technical solutions, which did not provide proper tools for thesaurus constructors. In one case only a text-edition-software was in use. The lack of time was constantly mentioned as a problem for all the participants in the project. All the respondents expressed the thought that the outcomes would be better with deeper examination, and one suggested that both the commentators and thesaurus constructors should have played a more active role in their co-operation. A common thought was, that every thesaurus project they had participated in had been unique, and also a learning process concerning working methods.

> "We gave a month [for commenting], but, however, it is quite a demanding task… So we did receive comments, some terms got even very exact comments, but then there was a lot without any feedback. So if you think of this kind of terminological work, maybe it could have been organised as some kind of happening, maybe go to meet the researches and discuss with them their comments and go through them more in detail than just on the basis of their written comments. It might have helped. Because especially with our own specialists… We received feedback like they were terribly busy, and some [comments] came very late. And some gave comments about the whole thematic section and some only about small sections… But we did receive such comments that were taken into consideration and was useful, but maybe we could have got more out of it." (FI TS4)

Although the informants spoke a great deal about the problems in thesaurus construction, in several contexts they also mentioned the nature of lexicological work – how emotive it can be. - "The discussions may be very explosive, but it is good to have different perspectives" (FI TS1). Although this work was considered to be very emotive and the members of management team to have strong opinions, the atmosphere was always discussed in a positive sense, and sometimes even with an air of amusement.

> "[My colleague] has made a very amusing picture how we with [other colleague] redouble about a word, and [the colleague] gets warm and takes off her pullover. And when we have received harmony and when no more words cause disagreements so [the colleague] starts to get cold and puts the pullover on, and there becomes a new conflict --- Yes, it is very passionate work." (FI TS3)

However, the feedback from the users of published thesauri has not always been constructive. One informant even mentioned a threat of legal act if a certain term in a thesaurus was not changed into a different one. (- It is noteworthy, that the term in the thesaurus was in line with the Basic Dictionary of the Finnish Language.)

What kinds of sources and authorities were mentioned? The authorities named by the Finnish thesaurus constructors were the Research Institute for the Languages of Finland (Kotimaisten kielten tutkimuskeskus[101]), the Basic Dictionary of the Finnish Language (Suomen kielen perussanakirja, published by Research Institute for the Languages of Finland) and ISO standards for mono- and multilingual thesaurus construction (ISO standards number 2788 and 5964), thesaurus books written by Aitchison and Soergel which were concidered "classics, basic books that are still in use" (FI TS2). In addition, special lexicons related to the area of the constructed thesaurus were named, such as, e.g. in the case of legal terms the Encyclopædia iuridica Fennica. The sources referred to were not used without criticism, and the dictionaries were especially criticised, and it was also pointed out that how dictionaries may be misleading

---

[101] "The Research Institute for the Languages of Finland is a national research centre and expert institution for linguistic studies. It studies the Finnish, Swedish, and Saami languages, Romany and Finnish Sign Language.

Their main tasks include language planning, compilation of dictionaries and various research projects. (Kotimaisten kielten tutkimuskeskus 2007)

sources by representing other discourses than these used in thesauri. Google search was brought up as a useful tool in the evaluation of dictionary terms.

> (*In the context of heads of household –term*)

> "Dictionaries are the last thing you should trust with these kinds of terms, or you should take them as very... These are very much based on special terminology and the normal common language dictionaries do not get along with them. If you take a look at NetMOT dictionary or similar, so there are often translations that you know already at the first glance that no way. And there are translations that you feel that are very strange, you have never jumped into that kind of expression, and then you go and google and seek for it as a phrase and put even site:uk, so that only pages published in Britain would come out, and like with only three hits as result. So then you know for sure that this is not used anywhere, it is a made up, an artificial expression. This is namely one of the things I work with every day when I translate --- so common language dictionaries are very dangerous if you trust them." (FI TS5)

The authoraties and sources mentioned were thus both terminological and methodological, but in the interviews the terminological sources were discussed more regularly by the informants and in more detail. Furthermore, the authorities were the sources that gave direction, but the uppermost criterion was usability, and as in the dynamic equivalence theory (cf. Nida 1964/2000; Vehmas-Lehto 1999 and chapter 3.5.2.3 *Dynamic equivalence theory*), one question was always for whom is the translation meant to be correct:

> "Already from the beginning standardisation has been present in this work, and I think it's good, but in details usability goes ahead. Basic lines have to be in order, so that when needed, you can make exceptions. The better you know your authorities, the better you can consider whether you can deviate from them or keep them in the same line." (FI TS2)

However, the special thesauri (in foreign languages) used as terminological sources was presented more as a source for term ideas or comparisons, than a clear role model, because they were often considered culturally biased from the Finnish perspective and thesaurus needs.

> "The content of the Sociological Abstracts is a little bit American, and also the structure of thinking. For example, we do not have the sociology of poverty as a big issue, not in that way, it is different here. And our [i.e. in Finland] writing about ethnical groups is the status on Swedish-speaking Finns, which is not in that way miserable and unhappy that it would require description of those who are worse off." (FI TS3)

Nevertheless, when constructing special thesauri, the most important thesaurus used was the YSA, and otherwise the most important source for terms and evaluation is the material that is going to be indexed and sought with the thesaurus-under-construction.

The specialists used were mainly of two kinds – experts on thesaurus matters as information specialist and experts on terminological matters as researchers representing certain discipline(s). Similar to the British thesaurus specialists (cf. chapter 9.2 *British thesaurus constructors*) the Finnish thesaurus specialists also mentioned the fact that in

some cases the role of a thesaurus was not clear for the researches, and it was considered as one factor resulting in conflicting comments (see examples 1 and 2 below). The level of specificity was also frequently considered to be a common problem (example 2 below).

Example1:

"Well, there were two kinds of experts. There were researches and then there were people of our field, so… In the covering letter it was, indeed, explained what it is about, but it [thesaurus] is supposedly not a familiar thing that way…" (FI TS4)

Example 2:

"One could say that the result is the same as when there are different kinds of apples at the store, but they are all apples. So it depends what the user needs the material for, when (s)he selects the apple, does (s)he take Granny Smith or something sweeter? If one is making an apple pie Granny Smith is very good, but if one is making jam one should select something else. So this is the principle behind in these projects. --- We also try to get the researchers to realise that we cannot begin with the concepts tightly defined in this thesaurus or controlled vocabulary, because it is too restrictive from the perspective of our seekers. In practice it would mean that there would be a need for a huge amount of those concepts and their names and then the [seeking] result would be just one or two documents which are necessarily not at all relevant for the user. So we aim at loosening the hard-and-fast terminology. This means emphasising the starting point that client is not going to receive any kind of ready and smart result when seeking with this controlled vocabulary system, but (s)he self selects which perspective of the topic is relevant for her/him." (FI TS2)

A crucial question in thesaurus construction is how to select a preferred term. A common problem discussed by each of the informants was the several alternatives for a preferred term, which was considered to require considerable work and co-operation. This was especially the case where it was a question of a new phenomenon, and in these cases the project team (representing the same language and culture area) was needed – and if there was the opportunity they were also used a great deal.

"--- But we did always receive a common ground. --- First we had discussed it with each other [with colleagues in the project], either of us had first translated --- and we had an auction, and then we send the common suggestion to the [third colleague], who then commented and then we went through the comments and swallowed without biting what we swallowed and in some cases continued disagreeing and then…. More or less, in fact, it was the compromised version we sent to the specialists." (FI TS3)

Despite discussions with project co-members, when evaluating the usability of candidate thesaurus terms, Google-search on the Internet was frequently named as today's most practical method. - As discussed before, dictionaries were also named, but they were often also considered as possibly misleading sources. The researchers in the particular field were considered as the most important ones, and, at the same time, also – due to diverging perspectives - as the most complex ones (example 1 below). (For thesauri that had been published update suggestions and new terms were most often received from indexers.) At the term selection level, harmonising with YSA was

naturally aimed for, but it was also often considered to be a complicated thing to do. – In conflict situations the tendency was to favour the practices and suggestions received from the specific group of potential users (example 2 below).

Example 1:

"--- We discussed the terms in expert communities, where the experts were researches. So we used them as sources to check the validity and correctness of the terms. But it seemed that if three researches were asked, then three different answers were received. And then we were there, thinking which one then to choose… (FI TS2)

Example 2:

"Well, we have here in Finland of course the special situation that we have YSA, and it has a standard position. And then we have of course the indexing term standards, but they are formal with criteria coming from outside. But then we have these certain principles coming from the discipline or from the users of our own organisation --- which must be taken into consideration. So reconciliation of all these is many times pretty difficult. The needs we face are sometimes little bit conflicting... --- According to the experts' opinion we stayed at the beginning too much in coherence with YSA. Especially when the YSA term was on a general level, the experts commented that we ought to choose more specific terms, and then we made these decisions." (FI TS4)

One of the themes in the interviews was the usability and future of thesauri. – At the end of the interviews the informants were asked to reflect on the practices and usability of thesauri today and in the future.

All the five Finnish participants thought that thesauri will also be needed in the future. Most of them stated that they are also useful in a full-text environment and some said that thesauri are becoming increasingly important as our information channels become more complex, broad and varied, especially when considering the Internet environment. Although multilingual and/or multicultural thesauri were considered to be very challenging to construct, they were considered essential, especially in indexing work (example 1 below), but also the needs of information seekers were often referred to (example 2 below).

Example 1:

"I have indexed also before in English, before the [multilingual version] --- and it was very difficult for me to try to make guesses about the term and look for them with the help of hierarchies. But as soon as I got the [new version] the seeking of possible descriptors truly rapidly fastened. When you can start the process with your own language… Because the results are sometimes pretty surprising, especially if it is a question about a case like homemaker, terms that are somehow artificial or terms that are created later, so they are not that kind of terms that would ever come to ones mind. So the benefit was noticed truly concretely." (FI TS5)

Example 2:

"This kind of conceptual analysing at this stage is needed also in the future, although it does not go to the terminological level… One could say that a word so simple as family roles may mean different things in different disciplines, and it is important to know them when seeking information, so that you can limit your search on specific meanings instead of being forced to go through a huge amount of…" (FI TS2)

Some of the Finnish thesaurus specialists challenged the technical solutions, and saw the traditional relationships as too foreign in actual information seeking situation. One suggestion (FI TS2) was to make thesauri more visual and more like a mind-map, so that it was more associative and reflected the material better and the relationships in databases.

The challenges for multilingual thesauri mainly concerned the linguistic and cultural differences. The most crucial question was to decide where the border lies between being sufficiently general and being too specific if a thesaurus is designed to have multilingual and multicultural or international content equivalent in all language versions. The informants reproached the demand for equivalents in all language versions, and instead suggested that irrelevant and/or too foreign descriptors should be left without Finnish equivalents, and similarly the Finnish language version could include terms that have no equivalents in (all the) other language versions.

> "And in a way the starting point should be going from up to downwards, begin with top terms and not with these kinds of specific terms. And one should move on to the narrower terms until a stage is achieved where one cannot anymore… Systems change or cultures are so different that one cannot speak with same concepts. I don't know… Maybe it would be a starting point. And, of course, there is this limitation that one cannot operate with very specific concepts, but one has to stay on a pretty general level. And in a way there is this possibility to make national additions, but it should be clearly transparent to the users that when they go to that level with all these differences when the national and cultural differences have stepped into the game. And like we have [here in Finland], that the same indexing tool in libraries, archives and museums might be in use, but at the same time everyone can also have their own special concepts…" (FI TS4)

As previously mentioned, the participants pointed out the need for co-operation and team work – when working alone one is always biased. In addition to this the other needs mentioned and advices given were the need for assertive organisations, concrete working plans, good division of work and realistic scheduling, the possibility of discussing term by term with several people, intense cooperation with the help of e.g. e-mail and an efficient technical (computer) solution. In the answers, two things were emphasised most - broad experience of indexing and information seeking was emphasised more than language skills (– maybe language skills were considered as obvious) (see example 1 below) and in the case of translation projects a carefully planned corpus (see example 2 below).

Example 1:

> "Experience as an indexer and information seeker is even more important. If the only thing you have done is constructing thesauri, they will become peculiar. When someone asks you [for information] you learn to know how differently things are approached. So the indexing is just imposing things from your own perspective, if there weren't the other [information seeking] side." (FI TS3)

Example 2:

> "The most important working stage is the thesaurus construction, I mean the design of translation corpus, which is going to be translated, to decide what to translate, so that

the terminology is possible to have in multilingual thesauri… And then, of course, the words related to different systems are probably that kinds of, that one can never achieve proper correspondencies, because it is not by any chance possible. They have to be left on a pretty general level. – Unless our dear EU will come and standardise everything!" [laughing] (FI TS5)

9.3.2 Family roles terms by Finnish thesaurus constructors

In order to study pragmatic indexing, term equivalence and the understanding and translating of terms related to family roles were studied.

In this section Koller's equivalence factors, Newmark's summary of various translation methods, and different (main) translation strategies are also used (cf. chapter 3.4 *Translatability and equivalence*) when evaluating the empirical terminological case, which is supposed to represent a problematic case for the informants. Component analysis is also used when relevant. The aim is to gain more knowledge concerning equivalence matters by a triangulation of method, data and theory (cf. chapter 6 *Methods and Material*). – In the previous subchapter the participants spoked freely about translation matters and the concrete cases discussed were raised by the participants, however, in this section, they have been asked directly how they would translate certain family roles terms.

The first introduction to the studied terms was word associations. After the word associations (for family role terms as stimuli words) they were asked for definitions and translations, and how the terms were represented in the working version of ELSST was discussed.

9.3.2.1 Family roles by Finnish thesaurus constructors

The stimulus word *family roles* resulted in 25 different word associations, with 33 occurrences.
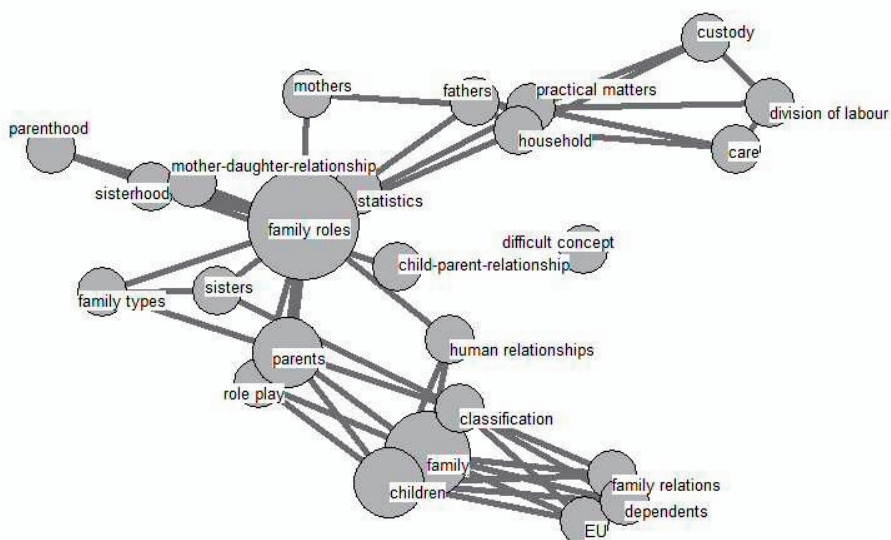
Figure 54: *Word associations for "family roles" by Finnish thesaurus specialists*

At the first glance the associations, provided by the informants for the stimulus word *family role*, were very much related to the biological roles of family members, and from the perspective of a nuclear family, i.e. of a family consisting of parents and their biological children. – However, as one of the informants pointed out, the roles are also social and/or psychological. In the definitions the biological roles and family members were still those most emphasised.

> (*when asked what is meant by family roles*)

> "Well, I think that family roles means, for example, first-born child, middle child, youngest child, or one can think also other kinds of, like father and mother are also family roles, and they represent also facts, because they are biological matters. But they can also be roles, a mother can be a child, if the mother has, for example, an alcohol problem. Then the child may suddenly take the role of a mother or father. And there may also be other kinds of, for example --- from the situation of alcoholism, how children may try to compensate or try to balance, try to keep the family situation as normal as possible, to take the role of a clown, or role of a troublemaker, and then children who bow out… These kinds of matters are meant by family roles." (FI TS2)

As family role terms, the names of (biological) family members were considered to be the same all over Europe, but the content of the terms and possible search results in actual information seeking situations were considered to differ, possibly even in a significant way due to differences between cultures, subcultures and time differences.

> ("*Could this also be the same at a concept level in England and in Europe?*")

> "Yes, it could, yes. I think this is that type of concept that isn't cultural. Of course they can be very different in different cultures, but, however, usually there is a mother, a father, children and other relatives. So here this is based on the European idea of a nuclear family, so it's different if we go to such cultures where a nuclear family

doesn't exist in a way, but there the role of the mother or the role of the father can also be divided albeit to five persons. But this is based on biology and chronology. --- But the material that comes up with these terms might easily be very different, so how the terminology works in practice.... --- But for our sociologist and social policy and social work researchers this would be usable. --- And it is the same also in Europe. In our cultural context." (FI TS2)

The studied term *family roles* was considerd easy to translate into Finnish at the term level, and all the five Finnish thesaurus constructors interviewed translated it as "*perheroolit*" (direct translation)[102]. However, there were differences on the concept level, and the term was considered to be an ambiguous descriptor in a similar way to the way it is seen in other languages, as is shown later together with examples.

The studied term was considered to be represented in ELSST in a very narrow sense, referring to economic roles within a family. Furthermore, it was not considered relevant from the broader social science perspective, which the participants tended to prefer.

9.3.2.2 Breadwinners and Heads of Household by Finnish thesaurus constructors

The stimulus word ***breadwinners*** resulted in 36 (conceptual) responses (25 of which were different).
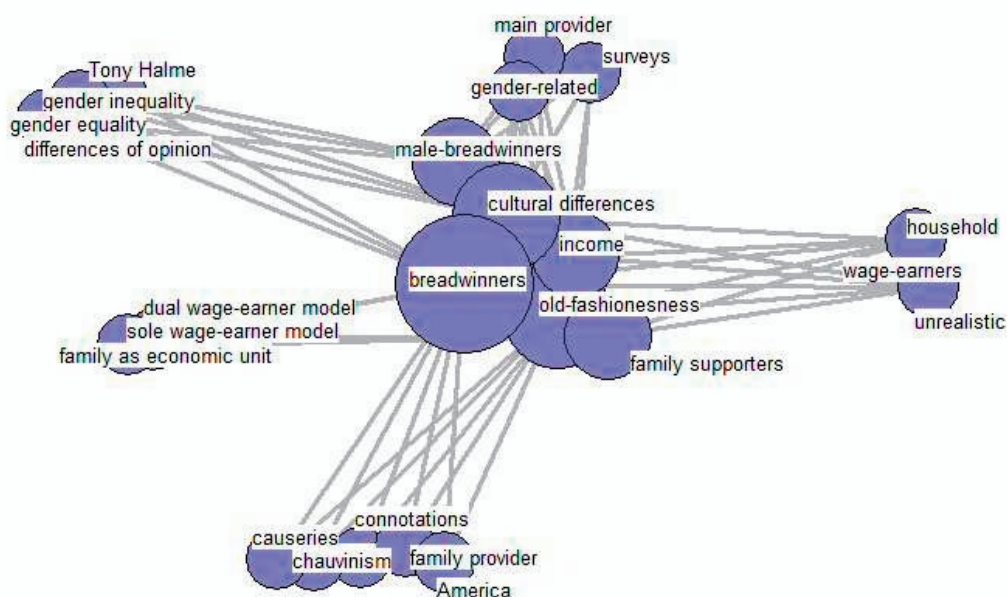
Figure 55: *Word associations for "breadwinners" by Finnish thesaurus specialists*

---

[102] The translation is similarly transparent as the source term, when in the coumbound the basic component is *rooli/role* and the adjectival compound *perhe/family*.

The Finnish equivalents provided for the term *breadwinners* were "*leivänhankkija*" (descriptive phrase, person who gets the bread) and "*perheenelättäjä*" (term, person who takes care of the family in economic sense). There were a considerable number of terminological differences, but on the more general and conceptual level the consensus was great. The respondents also commented on the connotations. The stimulus word was commonly understood as a non-Finnish concept, and strongly as an old-fashioned, gendered, and irrelevant word in the Finnish context.

> "Main provider, for example, in Finnish surveys they may ask if you are the person with the greatest income in the family, but in other cultures it may mean something else. The word itself is out-dated, because it clearly refers to the old model in which women were at home and men were working, so it is very difficult to imagine a woman as a breadwinner. So it is a gender related term and because of that it does not cover the concept very well." (FI TS5)

As a problematic term, the descriptor was considered to demand extensive work. It is also an illustrative sample of the situation, when a constructor has to consider different strategies and equivalence types, which in this case vary from domestication to existential equivalence and from partial equivalence to non-equivalence.

> "Breadwinners? I cannot so off-hand… I know it is a concept, which in English is perhaps not so loaded, but I think it is clearly an American word. And if I should translate it I should first study it, make a search in data where it has been used and look at the context, what is the conceptual environment, it clarifies, and, of course, if I would have to translate it right now I would consult our translator, who is American – and I would ask opinions also from others. This is a typical word in terminology which demands more work, although in parenthesis, I can say that even certainties often appear to be something else. --- I assume that this word breadwinner would demand a lot of work. And what should be done is to my mind to make the decision that whether to translate at all or to use the word breadwinners and put to it with a scope note saying that this word is used for this and this matter and that it is not known as a concept in Finnish, not in the Finnish language and not in the Finnish culture… It is pretty close to a person having custody or family provider, but we don't have it. Well in taxation we have the person with greatest income, so (s)he is well the one in the family who is more breadwinner than the other. So this word opens up quite conceptual and societal matters." (FI TS2)

The stimulus word ***heads of household*** resulted in 30 different word associations with 40 co-occurrences.
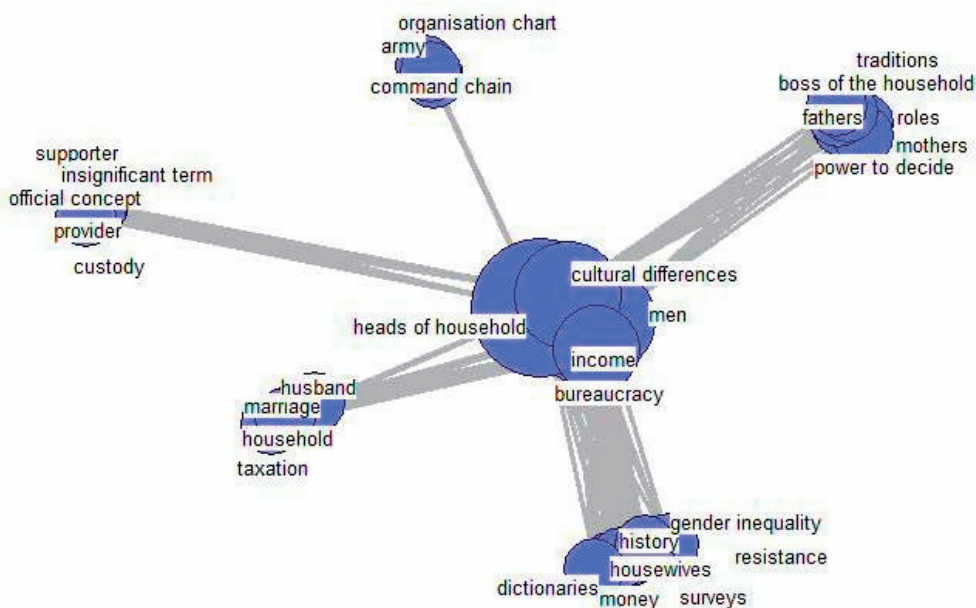
Figure 56: *Word associations for "heads of household" by Finnish thesaurus specialists*

The reactions to the studied term *heads of household* were very similar to the term *breadwinners*. It was strongly considered to be a gendered and old-fashioned concept.

> "In the Finnish society the problem with "head of household" is the same as with "breadwinner", so it's already like a concept… One has to take into consideration that when Kela [The Social Insurance Institution of Finland] or any other institution sends a paper to a woman with her husband's name, Finnish woman will blow a fuse. So this is based on culture and in the Finnish culture, in the Finnish agrarian culture, the wife had a strong position. So I guess Finnish women have never been – well I haven't studied this but I have that kind of conception that [Finnish women] have never been in this kind of oppressed position." (FI TS5)

Its meaning was still considered to differ from *breadwinners*, and *heads of household* was considered to be more related to custody, and as an official concept referring to a person having custody, to a provider. It was not considered a general language word – "certainly we don't use very much this kind of concept as head of household – it is a person having custody" (FI TS4).

9.3.2.3 Homemakers and Housewives by Finnish thesaurus constructors

The stimulus word ***homemakers*** resulted in 44 co-occurrences, out of which 31 were different. This word caused the greatest variety of word associations, and they were in their nature both positive and negative, even from the same participant. It was

also the most gender-neutral of the studied narrower terms - although only on its surface (as later discussed).
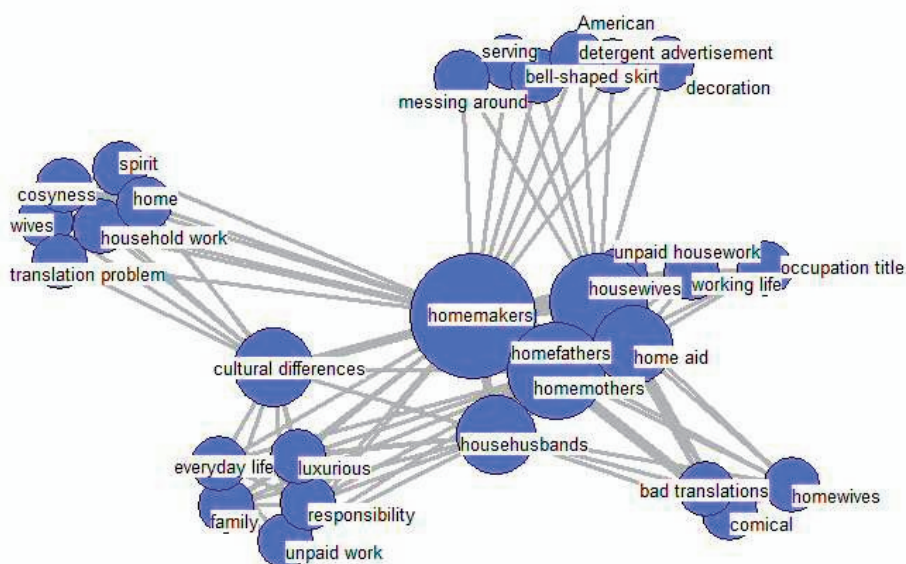
Figure 57: *Word associations for "homemakers" by Finnish thesaurus specialists*

The translations into Finnish represented three different concepts. One version was to consider the term to mean in Finnish "homemothers + homefathers" (cf. single-to-multiple term equivalence in ISO 5964-1985), the other "homemothers + housewives" (cf. also single-to-multiple term equivalence) and the third a "person who takes care of household instead of earning money from employment" (in ISO 5964 –terms semantically more exact than the two previous samples, but not a true keyword). The distinctive components between different versions were thus parenthood and gender.

Its semantic components were also discussed significantly more than the other terms. It was generally considered to be a translation problem and this was mostly due to its superficial gender-neutrality. The fact that, in practice (in Finland), the term refers to parents with small children was also considered as a translation problem. Therefore a clear conflict existed between the conceptual meaning and the pragmatic meaning.

> "Well home aids, but home aid is an occupational title for us… Housewife or mother of the family, but mother of the family can also go to work… Stay-at-home mother, but it can be also stay-at-home father… So homemaker is understood as either a person doing unpaid housework or as the partner in a marriage or cohabiting relationship who takes care of the children. But childless homemaker in this sense, well, why wouldn't it be possible too, but if the husband is retired and the wife is still working, the husband cannot be considered as homemaker, although he would do the housework. And then of course there is this occupational title, home aids, and especially communal home aids, so I don't know how much it is an occupational title, but anyway not this kind of matron or housekeeper or maybe this housekeeper… Well, why not, but we don't… --- " (FI TS3)

Similarly to the previously discussed terms, the term ***housewives*** was also considered to be a foreign concept and word within the Finnish context - "all these asked terms are different in Finnish" (FI TS1).

Housewives also resulted in various word associations, but clearly more negative ones than the word *homemakers*. In all it resulted in 31 associations, out of which 25 were unique.
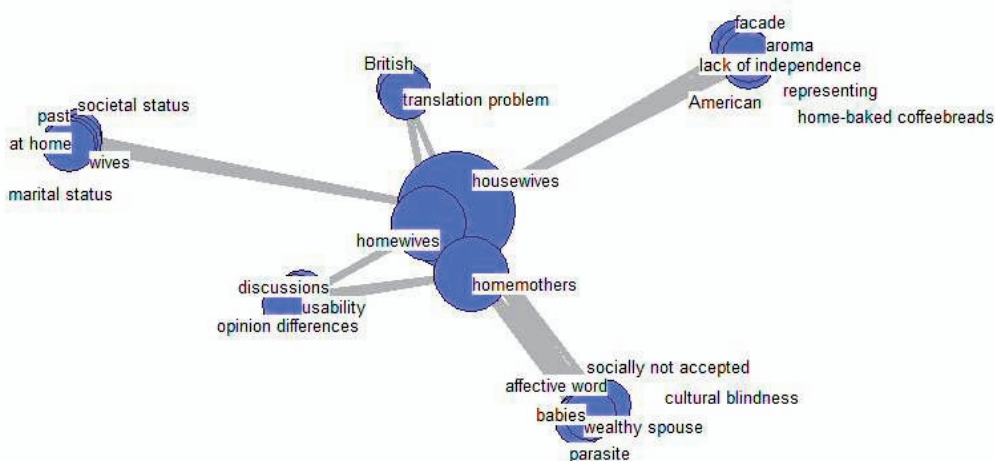
Figure 58: *Word associations for "housewives" by Finnish thesaurus specialists*

The translation strategies varied, and it was not a simple matter to select any particular one. The most common alternatives were literal and a source language emphasised translation "*kotirouvat*" or a target language emphasised communicative translation "*kotiäidit*" (*stay-at-home mothers*). In this case even neologism was discussed, which was surprising in the light of the previous more theoretical and methodological discussions, and also when considering the norms and rules given in the studied standards and guidelines. The neologism would have been a direct translation of *housewives* as "*kotivaimot*" (*home+wives*).

> "Kotirouvat [homewives]. And [(s)he] suggested that "kotivaimot" [homewives], but I don't know actually… wife is at home, whether officially married or not to the mister, that there is not actually… But as an occupational title it is not accepted, and before one had to be like a wife of bricklayer or somehow to express that the livelihood is organised in some way. And evidently also the status, that whether you are a wife of a professor or of a bricklayer. And it also lasts in the widowhood, that one was the driver's widow. But you are still a housewife even if one has never worked… Well, a widow is a housewife. So the existence of the husband does not change it to anything else." (FI TS3)

Although the term was considered to be linguistically difficult, the main difficulty lies in the differences between cultures and their values, and in this case the connotative aspects were discussed extensively.

"In Finnish we have housewife, which associates very clearly with a wealthy marital partner, and then we have stay-at-home mother, which associates that you have babes, so we don't have any kind of neutral expression… It's like the Finnish society would despise for being at home without children [laughing]. --- Surely the Finnish society, if one would be at home with older children, would call you a parasite. So we have our values or society's cultural tags in the background. --- Certainly it is socially more accepted to say that you are unemployed than to say that you are at home." (FI TS5)

9.3.3 *Equivalence and translation strategies by Finnish thesaurus constructors*

As stated previously (cf. chapter 9.2.2) one of the research questions and themes is equivalence, which raises questions such as: how is equivalence understood; what do Finnish thesaurus constructors aim at in their translations? The theme with the Finnish informants was also covered by asking directly how equivalence had been understood and possibly defined in thesaurus projects, and also what had the aim of different translations been. Among this group some topics were also referred to in other contexts throughout the interview, such as why they consider multilingual thesauri to be required and how they would translate certain terms (cf. next subchapter 9.3.3).

First, it is important to remember that translating was considered not as an optimal way to construct a thesaurus, but a way determined by practical (i.e. time and money) reasons (cf. previous chapters and Hudón 1997 & 2004). It is also important to note that although the new multilingual thesaurus was often constructed on the basis of a monolingual corpus, it was not considered by the participants to be translating in a traditional sense, but a very special kind of terminology work, where one has to consider various matters of thesaurus use and thus reflect correspondencies towards collections, indexing practices and discourses of information seekers. The participants, in principal, had the possibility of influencing the source corpus, although in practice considerable adaption to the needs of the majority and conforming to the guidelines was made earlier in the project.

"--- [In the project] we discussed it as translating, but it's not translating in its traditional sense, but looking for the correspondences ["vastaavuuksien hakemista"]. (FI TS4)

("*What were you aiming for in your translations in the project?*")

"It was more or less the correspondence ["vastaavuus"], mapping the correspondence, it was not like translating, it was correspondence." (FI TS5)

Second, it is important to remember that Finnish thesaurus specialist represented other views than those of the British thesaurus specialists interviewed (cf. 9.2.2). For the British informants English had had a major role in all the cases – often as a source language in the thesaurus project, and at least as a common working language, whereas for the Finnish participants it often represents a source language and/or a working language or – rather seldom – one of the target languages. Due to the dominance of English as a source language, the Finnish participants spoke mainly from the perspective of translatability issues from English into Finnish.

Therefore, **what was aimed for in the translations?** The informants were asked directly how they understand equivalence in multilingual thesaurus construction and whether they somehow defined it in their projects. In this section the terminological cases were brought into discussion by the participants. - The topic was also studied via terminological examples, when the informants were asked to give examples of difficult terms in their projects, and also in the case of the study, family roles terms, were discussed from the multilingual thesaurus construction perspective.

In the thesaurus projects referred to, equivalence was not explicitly discussed (example 1 below), and it (more or less directly) often surprised the informants in the interview situation (example 2 below).

Example 1:

> "I think that we didn't discuss the matter, but it was somehow obvious that it is rather the material to which we reflect." (FI TS3)

Example 2:

> "In these projects --- surprisingly little we discussed these kind of matters, so it was kind of left to every country or subproject ---" (FI TS2)

In practice, equivalence matters at the theoretical or more principal levels were usually considered to be obvious. When constructing the Finnish version of a multilingual thesaurus some clear aspects or even criteria to follow were found.

The clearest consensus was that thesaurus constructors do not work primarily with terms, but with **concepts** referring to different kinds of abstract ideas and phenomena etc, therefore they were aiming at **conceptual equivalence** (examples 1 and 2 below) and it was not considered important to compare this with dictionaries, but with the actual search results (example 2 below).

Example 1:

> "We were aiming at a kind of factual correspondence that these phenomena should correspond with each other. So that such a concept can be translated smoothly, but what is it corresponding to? For example, we were for a very long time – well let me look for an example – this, it is related to pensions, and after a long time it clicked that oh we cannot translate it like this. Here it is… It is also related to these pensions which were as such tricky [to translate], and in it we'll then probably end up leaving it on a general level, it's no use to include these to this kind of… to a thesaurus, because they vary so much between countries. So anyway, here was a concept we though for a long time that it would refer to State pensions, and when we asked what is the content of the concept we realised, that it is a national pension, so it's not like private pensions and state pensions, but national pensions. ---" (FI TS4)

Example2:

> "It has no significance whether it is correspondent according to a dictionary, but it is particularly "mapping the concept". Because anything else has no significance if we think that it is going to be used for [information] seeking. So if a researcher seeks for a certain phenomenon from material published in different languages and uses thesaurus as a tool, then there is no point in the controlled vocabulary if the other versions do not express the same phenomenon." (FI TS5)

Thus the other clear aspect, very close to conceptual equivalence, was usability and relevance in the Finnish language context – seen by the way the search results with the source and target terms would cause **similar results in an information seeking** situation (cf. also example 2 above). The matches were not supposed to be necessarily exact, but similar. They aimed thus at (as in Nida's terms) **dynamic seeking equivalence** or (as in Koller's terms) **pragmatic seeking equivalence**. When moving form formal into dynamic or pragmatic equivalence the idea of "equivalent enough similar" was challenging to define in practice.

> "Well equivalences are not necessarily exact, we had the concept of sufficient equivalence ["riittävä ekvivalenssi"], so when we considered the distinction to be so small that in practice it has no significance for the users with these nuances… But in some cases we had to think about it very hard where is the definite border for equivalence. In the Finnish group we thought quite a lot – probably because I was the leader - that again usability is more important than whether it is theoretically absolutely pure. So there is the user's perspective, but it was indeed difficult, because in this case there was no user to directly reflect to. And even today the thesaurus is not in use anywhere in Finland. But if somebody uses this as a dictionary and wants to seek from a English file and does not know how to call something in English, so (s)he can find the correspondent in the English language area by looking at the Finnish version. So that way the concepts correspond with each other, although they may not mean exactly the same thing. So that way one can make like clusters or groups and the seeker can then on the basis of them decide herself how relevant the certain document or facta is to her/his question. So we did not aim for this kind of terminological work where one has to define strictly and conceptually and there must not be allowed anything else than what it then includes." (FI TS2)

The third clear aspect was **collection equivalence**, when the thesaurus constructor aims at equivalence between descriptors and collections, i.e. that the terms in the thesaurus would reflect the content of the collection and (in this context) thus descriptors needed in its content description (example 1 below). In practice, collection equivalence is very close to seeking equivalence, because the constructor has a clear idea of the seeker and therefore the seeking context is very much nested into the future indexing process (example 2 below).

Example 1:

> "Well, in fact we didn't think about the target audience that much, but we though… and we had used YSA terms as much as it was possible to find. But who would use this, I think it was never on view, but instead most important was to which material this was going to be used for in indexing, so that the material would be described soundly. Important was also that the material is research data and material, and what would be searchable. So, without question, we thought that it will be the field's student or researcher, who will use the material, so we didn't have to consider just any elementary school pupil who is making a lecture, and their experience of vocabulary and scale of language use ---" (FI TS3)

Example 2:

> ("*So what is equivalence in this kind of multilingual thesaurus work?*")

"Well, whether you use the term in any language, so the material that pops up with this keyword should correspond with what has been expected. That it has to reflect the one and same material. But, of course, materials collected in different organisations, depending on what is considered as a problem, so of course it varies, but anyway, it is what has been aimed for. For example, here in the case of school institution, it was especially obvious. If you seek 'secondary school', which views do you get? Do you get same kind of people from every country or are there age differences, have they been somehow selected to the school with tests or something else, as wealth? We thought that researchers want uniform quality material regardless of used language." (FI TS3)

The informants did not always agree with this practice, but they still tried to conform to it. It was generally considered a reality that a multilingual thesaurus, especially one aiming at multicultural content cannot always "succeed", i.e. one has to be ready for compromises, "to give and take". Therefore, **consensual equivalence** was also often aimed for, which can also be seen as aiming at **existential equivalence** (cf. Koskinen 2000, 83).

"The votings were often about whether a term should be removed or not, for example, is it too culture-bound. It was not a question if a term was impossible for one language, but about the whole. So, in fact, if it is impossible to translate into several languages, then it is a question about culture or typical way to describe things in English language or culture. But then, on the other hand, a term should not be removed from a thesaurus because it is linguistically difficult for one country. So the one country should somehow round the thing, use a narrower.., or let it be a translation mismatch. - So the dictatorship of minorities cannot be allowed." [laughing] (FI TS5)

In the answers it was clear that informants usually needed strategical help in their work. They used other thesauri not only as terminological sources, but they also studied how terms had been translated and how difficult cases had been handled.

"Yes, and then there was the bilingual vocabulary of the Center of Statistics, we looked at it… The Finnish word was not important, but instead how each one had translated English into Finnish… Because there are several possibilities." (FI TS3)

Sometimes a term was left without a Finnish language equivalent, although the source term was left in the thesaurus. This was done because of the usability and relevance – the translation would have been too foreign and/or artificial and there was no need among the planned user group(s) for a direct citation loan. It is also noteworthy here, that the participants did not aim at a term per term –equivalence between different language versions, but they agreed that there are several degrees of equivalence, and non-equivalence was one of the accepted ones (cf. ISO 5964).

It was also a shared opinion that it is not necessary to translate everything in multilingual and multicultural thesaurus projects (examples 1 and 2 below), and this was not discussed as a conflict to the standards and guides, but as a very natural and realistic choice. In cases of non-translation the equivalent provided in the translated version could be the original source language term (example 2) or it could be left without any equivalent (example 3 below).

Example 1:

> "For example, in the Nordic [project], we aimed in equivalence relationships at – I mean in relationships to other languages – we aimed for conceptual correspondencies and we did not begin to translate such concepts, which are clearly not relevant from the perspective of our material." (FI TS2)

Example 2:

> "When there are terms related to systems, occupational titles, geographical divisions, we don't translate them at all. In the Finnish language version there are also English language terms. So it was one clear decision we made that in vain we don't translate." (FI TS2)

Example 3:

> "Sometimes we make exceptions on specificity, according to the demand, for example justice, we leave it in one version as such and in the other we go to single crimes." (FI TS2)

As already stated in the context of construction practices, the participants also experienced thesaurus construction projects as unique learning processes. The process sometimes changed the working methods, but the translation aims also changed sometimes during the process. The question of non-equivalence then had to be considered more explicitly than in the beginning, which led to the increased tendency to make non-equivalence relationships between different language versions. Artificial equivalents were generally not accepted as descriptors due to poor predictability in indexing and in an information seeking situation (cf. Iivonen 1989; Fugmann 1994; ISO 2788-1986; Rowley 1988).

> "It takes time to internalise the aim, why one is making this and that. For me it was also so in the beginning, I didn't quite so precisely understand this, that if we think about a term or a concept which does not have an equivalent in Finnish, it is the same to leave it without translation than to make an artificial correspondent that nobody will use. Nor in indexing nor in seeking. This happens when an equivalent is artificial. So, if we think of descriptors, the indexer --- (s)he must find it, must look at the hierarchy or to look for single words, and if the equivalent is artificial enough even the indexer does not find it, not to mention the seeker then." (FI TS5)

In special cases the participants were willing to accept the foreign language term as a Finnish descriptor as such, i.e. as a direct citation loan but not representing non-equivalence relationships. These cases were of two kinds – the concept and term were already known in the Finnish context as a citation loan (e.g. *open access*) or the concept was foreign in Finnish, and the original foreign language term was considered to be more usable due to the skills, needs and practices of the future users (see example 1 below). In the case of a loan word (adopted into the Finnish language system) vs. a more Finnish word, the loan word was also sometimes selected as a preferred term in the name of user-friendliness or usability. The informants were still very aware of the risks and e.g. of so-called "false-friends" (which in translation science means easily misunderstood terms that look alike, but differ in meanings.) (example 2 below).

Example 1:

"Somehow we thought so, that if somebody wants for example an English language term like *community nurse*, if somebody wants information then (s)he won't seek it in Finnish or Swedish language material, or if (s)he seeks, then it can be with the string community nurse, because we don't have any conceptual equivalent for it. I mean in our languages [Finnish and Swedish], that it would be like a paraphrase, it would need to be explained. And we thought that if someone seeks for information about this topic (s)he will certainly know that it is *community nurse*, because otherwise (s)he couldn't even know to seek, because these are country-bound." (FI TS2)

Example 2:

"And the other thing linguistically in equivalence was, of course, that because it was a multilingual thesaurus, we aimed to include the loan word – if it had the same meaning. But it surprised how often it can happen that eventhough the English word is exactly the same as the Finnish word the meaning has changed. So the loan word is not the same in Finnish as it has been in the original, like, for example, in English. They are often in their origin Greek or Latin based words, so they are of course not like direct loans. But in the case of these concept names we chose the loan word from the two alternatives, because we thought it will make it more understandable, because it has several languages. (FI TS1)

The first case is not represented in thesaurus construction guidelines, but the second case follows the guidelines given in thesaurus literature about selection of a preferred term when discussing loan words and translations of loan words:

"Terms from other languages are sometimes encountered as "loan words", i.e. foreign terms which are accepted as newly-coined terms. If these terms are well-established, they should be incorporated into the thesaurus. Occassionally a loan word and a putative translation co-exist. If the loan word is more widely accepted, it should be treated as the preferred term, but if the translation becomes well-established, this should be preferred. Resiprocal references should be made between the preferred and non-preferred terms." (ISO 2788-1986, 7-8)

As already discussed in previous contexts and examples, the Finnish participants interviewed tended to consult a variety of sources when evaluating the nature of candidate terms. **One of the central (hypothese like) assumptions of the study was that Finnish information specialists tend to aim at harmony in YSA when constructing special thesauri.** In the interviews, the importance of YSA was not questioned directly, but it arose on several occasions, such as when asking about terminological sources. In additon, the significance of YSA in equivalence matters was discussed a great deal by the participants, and each participant had aimed at harmony in their translations – within the realms of possibility.

"All the time we have had YSA here alongside with [the other thesauri], and it has been already used in the construction of our controlled vocabularies. When we add new terms, we aim at, if only possible, harmony with YSA." (FI TS2)

The informants were aware of different discourses and of different ways to approach indexed matters, although they did not directly discuss discourses. They also

reflected their own background to the way they tended to see things (cf. Adler 1997; Hall 1981).

> (*Were there problems between a term and a descriptor?*)

> "Well mostly it was when we were tuning up with YSA, so that it would be YSA term as long as possible, it was the first. And then social science language is so vivid --- The background is always uncovered in the suggested descriptor. What was meant to be indexed was always in mind." (FI TS3)

> "I studied their information seeking --- The researches have used very specific terms. --- The biggest challenge for controlled vocabularies is to describe the world in that way, that seeking is possible. But on the other hand they cannot be too specific, but on the other hand the researchers or information seekers need particularly a certain perspective to the matter. So it is an unsolved question." (FI TS5)

Were the difficulties due to linguistic or cultural reasons? The informants spoke more about differences in cultures and cultural systems, and not very much about linguistic matters (cf. Nida 1964, 130). The differences in culture were considered to cause linguistic differences. The cultural differences led to internationalisation discussions – whether a term is too culture-specific or relevant just to a small cultural group. However, on the other hand, the (more rare) linguistic difficulties were considered even more problematic, since there was "no way to round it".

> "--- And on the other hand, well, the differences are only on the level of single terms, but linguistic differences can sometimes cause great difficulties since the matter is outlined via the language. If there exists a linguistic problem, for example, in the social sciences, where "social" means in Finnish both the social and societal, so it is divided into two parts. And it causes always terribly big problems, because each time there is like "social participation", social this and social that, one has to think is it societal ["yhteiskunnallinen"] or social ["sosiaalinen"]. And it is terribly difficult for the English speakings to understand because they think it as one, they cannot even divide it, so it is our way to even divide it and it is depended on the language which makes the division, and which division their language does not make. And to discuss these linguistic differences is truly interesting, and when someone in the leader team says that this cannot be translated into Greek because we have one word for this in Greek, for example, for these two phenomena, or two words for this one phenomenon, and then everybody is like uhmmmmmmm. So the linguistic [problems] are very difficult." (FI TS5)

Standards regarding thesaurus construction were not discussed in detail, but their influence and degrees of equivalence (see degrees of equivalence in ISO 5964-1985 standard in chapter 3.3.3.2 *Equivalence in multilingual thesaurus construction*) could be seen in the answers of the respondents. The tone was then often critical, and the concepts and guidelines represented in standards can be interpreted to be too linguistic and not sufficiently practical.

> (*Did you define equivalence or correspondence?*)

> "Well, there was yes, in the list of mismatch terms, that if it was a broader concept or narrower concept or… But then there were also that if one word in English was several words in Finnish, it was also there [in the list of mismatch terms] --- But I think it is not [mismatch], it is then a question of interpretation when it is not a mismatch

anymore, because if the correspondencies correspond so it doesn't matter whether it is one or several words, but if the phenomenon spesifically corresponds only partially." (FI TS5)

- **Summary and Conclusions**

Terms that were possible to translate into Finnish literally, i.e. there existed an established term in the Finnish language, were usually translated in that way, and the selected strategies thus had a common emphasis on the source language. The relevance of the studied terms and concepts was often rejected in the Finnish context, but this did not affect the used translation strategy. When it was not possible to translate directly, the translation strategy was considered from a richer variety of perspectives. If the concept was considered clearly foreign from the Finnish perspective, the translation tended to emphasise more the source language context. In each case the informants thought of the Finnish user when looking for an equivalent, but the aims of the translation varied according to the situation (what kind of problem the studied term represented) – in the most difficult case (*homemaker*) the translation aim was sometimes considered to be more of a means of providing an explanation of the content of the source language term, than an expression of an usable descriptor in the source language context.

In thesaurus construction standards and guidelines, vocabulary work is similar to a clear and logical decision-making process, the result of which is one good (or best possible) solution. The Finnish thesaurus specialists talked about it more as a compromise-seeking process, where they stressed, the result could also have been something else entirely.

It was a shared feeling, that thesaurus construction and updating demands a lot of human effort, that there is a lack of resources (money, possibility to engage enough people to do the job) and this produces a feeling of inadequancy in front of this enormouse task. In Finland the General Finnish Thesaurus YSA is generally used in common and special libraries, and its influence was strong in the selection of preferred terms in new multilingual thesauri.

Thesaurus construction was also generally understood as a learning process, where the methods and aims could change accordingly to the experience gained. The working methods tended to move in a more democratic direction and in equivalence the attitude towards non-equivalence was more accepted.

On a more general and theoretical level in the interviews, the strategies found in this group were internationalisation, domestication and foreignisation. Existential equivalence was also found, and the reason for this is that certain terms are needed among the majority of project participants (countries). Translation projects were common, and in these projects the main strategy was to aim for internationalisation, which means here, i.e. in thesaurus construction context, similar results to the source and target term (i.e. descriptor) in information seeking situations. If this was not possible, the domestication was usually prior to foreignisation. - In internationalisation

strategy a broader level was looked for, where the cultural and/or linguistical differences do not cause serious difficulties anymore. – It is still noteworthy that foreignisation was considered to be a natural solution, if the source descriptor was considered to be relevant also in the target culture. In these accepted cases, foreignisation often means, in practice, a direct citation loan – a source term is left without any translation, and included as a clear foreign language term. Otherwise foreignisation was usually not accepted and the thesaurus constructors did not want to make "artificial equivalents". – It is important to keep in mind, that the strategies used and equivalence matters discussed reflected here the most common used method – translating. Although the translation method was commonly used in multilingual thesaurus construction, it was not considered to be the optimal way to construct a thesaurus.

The equivalences aimed for were conceptual, dynamic seeking, collection, and consensual or existential equivalence. The conceptual equivalence was especially emphasised, but the consensual equivalence was considered as a priority, in practice, since *no-one can alone see all the meanings and nuances*. (The result differs from other groups studied, e.g. Finnish researches emphasised the conceptual equivalence, and in it the connotative aspect was very strong, which was not brought up in equivalence discussions by the Finnish thesaurus constructors.)

In terms of Koller's equivalence analysis (1989, see chapter 3.4.1.2 *The idea of equivalence*) the Finnish thesaurus contructors interviewed in this study aimed at *denotative equivalence*. Most propably it is due to the nature of documentary languages that the connotative aspects were not discussed very much within the context of equivalence issues. *Connotation of socially determinded usage* and *connotations of geographical relation or origin* were still mentioned indirectly, and they were considered to be aims in thesaurus construction, but their realities were seen as difficult ones, since there is a lot of significant variety especially within social sciences. Thesaurus translations clearly aimed at *text-normative equivalence*, and in the Finnish context this also means harmonisation with the Finnish general thesaurus YSA. The *pragmatic equivalence* was considered as the primary criterion, and this means, in this context, that a thesaurus constructor aims at correspondence in an information seeking situation between different language versions and for users representing different cultures. Furthermore, the pragmatic equivalence is an organic way related to user-friendliness, aboutness, relevance and usability. *Formal equivalence* was not necassarily aimed at in thesaurus translations. The equivalence among thesaurus terms' relationships were considered important in the sense that they determine the despcriptors, i.e. the meaning and place in the thesaurus, but the thesaurus constructors (in accordance with standards and guidelines) do not aim at e.g. equivalence between non-preferred descriptors.

Moreover, the Finnish thesaurus constructors followed the idea of dynamic or functional equivalence, according to which the function of the source text is the same or similar to the function of the target text (cf. Nida 1964/2000; Vehmas-Lehto 1999). They thought that translation should cause the same reaction in the target audience as the source text does in the original context. Instead of evaluating "is the translation correct", they constantly discussed "for whom". Accordingly to the Skopos theory (cf.

Reiss & Vermeer 1986; Koskinen 2001) the thesaurus construction process was seen as a decision-making process, but especially in the case of most problematic terms the guidelines followed were not always coherent.

9.4 Similarities and differences in thesaurus construction

The studied **terminological case** family roles was moderately well represented in the studied **thesauri** – in the multilingual UNESCO Thesaurus, Eurovoc and ELSST and in the monolingual HASSET (British-English), SOSIG (British-English), CSA Thesaurus of Sociological Indexing Terms (American-English/International), ICPSR Subject Thesaurus (English-International), ERIC (American-English) and YSA (Finnish-Finnish) (about the studied thesauri see Chapter 6.3.1 *Thesauri*.). *Family roles* was represented in six thesauri as a preferred term. *Breadwinners* was represented in five thesauri, and in most cases, that is in four out of five, as a preferred term. *Heads of household* was also represented in five thesauri, in each case as a preferred term. *Homemakers* was represented in six thesauri as a preferred term. The most heterogeneous result was with *housewives*, which was represented in seven of the thesauri, but in three of them as a non-preferred term. None of the terms had an exact equivalent in the General Finnish Thesaurus, YSA. The theme of family roles was still represented in all of the thesauri to some degree, i.e. from more or less different perspective or emphasis, e.g. with terms referring to domestic responsibilities, household duties, sexual division of labour, stay-at-home mothers, parental role etc.

The **overall discourse of thesauri** shows that thesauri conform to traditional language usage and practices. The terms were understood in accordance with general language usage, although their use in thesauri and databases was narrowered to certain aspects. A greater consistency among different thesauri existed more on a conceptual level than on a term level (cf. Iivonen 1989 and 1995, Iivonen & Kivimäki 1998). E.g. *homemaker* is also an occupational title, but in the studied thesauri it was represented solely as a caregiver within a family, and its near-synonym *housewives* was sometimes used as a non-preferred term for it, sometimes as a narrower term. The lexicological richness was greatest on a broader term level (*family roles, homemakers*), which is partly due to narrower terms and partly due to the common character of the terms.

The scope of the studied thesauri and the indexing practices they are part of, was clearly seen in the material (cf. Hjørland 1997; 2002 and 2004). For example in ERIC the educational aspect was most clear, and it explains why e.g. in the case of *heads of household* parenthood was emphasised. HASSET is more likely to include specific terms than e.g. YSA, since it is used in indexing of even variables in statistical material. The trends were, however, not always so obvious – e.g. YSA is a general thesaurus and does not include the studied terms, although it is still very rich and detailed with regard to e.g. different types of household and to possibilities for describing different tasks and responsibilities within them. The discourse of thesauri is thus not unified – although it clearly represents a certain kind of text-type (cf. Reiss 1981/2000), it also includes different kinds of subdiscourses.

The **changes during the study period** 2002-2009 were rather minor, except that SOSIG was replaced by its source model, HASSET. It was slightly more usual to have new terms as related, narrower or broader ones, than to have removals. Additions and removals strengthened the studied terms meaning as a family role term. In the case of *homemaker* a clear trend was also seen of a move to a gender-neutral or gender-evenness representation style. (It is noteworthy, that adding symmetric terms for different genders does not necessarily mean true gender-neutrality, cf. Engelberg 1998.)

The thesauri studied did not provide detailed information concerning how they have understood **equivalence**. It was often mentioned that guidelines that were followed were the thesaurus construction standards provided by ISO and SFS.

The multilingual thesauri studied were in their **nature** rather international, and the underlying basic assumption about equivalence was that each descriptor has an equivalent in other languages, but the non-descriptors may vary and are dependent on the current language version. The dynamic nature of language was thus acknowledged, and formal equivalence was aimed at to a certain degree, and still rather strictly adhered to (cf. Meriläinen 1997, 100; Nida 1964/2000, 129-130). The result conforms to the ideas normally represented in thesaurus construction guides and norms. As discussed in chapters 3.3.3.2 *Equivalence in multilingual thesaurus construction* and 3.4.3 *Equivalence in thesaurus construction* when compared to the main translation strategies such as domestication, foreignisation and internationalisation, the ideology behind the ISO 5964-1985 standards is closest to the idea of internationalisation. In practice, this means that the content of the thesauri should be designed in a way that makes it possible to aim at equivalence between different language versions.

The Finnish (five persons, five interviews) and British (six persons, four interviews) information specialists specialised in social sciences were interviewed as **thesaurus constructors** representing nine different institutions. It is noteworthy that the Bristish thesaurus specialist represented differing view from the Finnish thesaurus specialists interviewed. In all the cases, English had had a major role – as the exclusive language, the exchange language or often the source language in the thesaurus project, and frequently as the common working language (cf. ISO 5964-1985). Most of the British participants had been working monolingually and/or been in charge of a monolingual corpus, and thus they represented a more corpus-based view or a general (monolingual) thesaurus construction view. For the Finnish participants English often represented a source language and/or a working language or – rather seldom – one of the target languages. Due to the dominance of English as a source language, the Finnish informants spoke mainly from the perspective of translatability issues from English into Finnish. (As stated before, the British and Finnish informants thus do not represent true comparison pairs, but the aim of interviewing different relevant groups and the triangulation of methods and material was to gain a holistic view, cf. chapters 5 and 6.)

Together with the translatability and equivalence issues studied the informants often referred to the **text type** – thesauri represent a special kind of discourse (cf. Reiss 1981/2000; Buckland 1999; Lehtonen 1994; Potter & Wetherell 1990; Pälli 2003). They also generally considered the lack of **common understanding** of the nature of a thesaurus to cause problems in thesaurus construction.

The thesaurus constructors acknowledged both **cultural and linguistic reasons** for making the construction of multilingual and multicultural thesauri difficult. The terms which were considered as most difficult to translate into Finnish by the Finnish informants interviewed were also considered to be complicated by the British thesaurus specialists. The difficulty was often due to different discourses, as well as due to time, which changes the meaning and/or also the style of the words (cf. Aitchison 1984; TSK 1989; Varantola 1990; Wierzbicka 1997; Buckland 1999; Lehtonen 2000; Katan 2004; Forsman 2005).

Although the participants were constantly very much aware of the **discourse** of thesauri, they did not always follow the norms of documentary language and in the most exceptional case (the term *housewife*) even a neologism such as "kotivaimo", a literal translation[103], was suggested. The reason for such dispersion and flexibility is supposedly not only due to the problematic terms, but also the lack of proper methodological tools for analysing, the determination and motivation of the aimed for equivalence. The informants themselves raised the fact that the social sciences are an especially difficult field to handle – due to different needs and discourses, and the aims and norms of communication.

Due to the **research design** – the British participants were not asked to translate anything and they represented a more translation corpus based view – the British participants did not discuss **linguistic difficulties** (between different languages) as much as the Finnish informants did. – It was natural for the Finnish informants to emphasise cultural aspects, and to consider the linguistic matters as sometimes even more difficult matters to handle in a multilingual thesaurus. In general, all the informants talked more about differences in cultures and cultural systems, but not very much about linguistic matters (cf. Nida 1964, 130) although the differences in culture were considered to cause linguistic differences. The cultural differences led to internationalisation discussions, and it was commonly stated that if a term is too culture-specific or only relevant to a small cultural group it should not be included in (an international) multilingual thesaurus.

The **nature of documentary languages** and thus also of the descriptors caused many obstacles in the translation process – the foreign concept was considered to be describable in Finnish or in English, but to express it with a term suitable and usable for indexing and information seeking purposes was far more complicated, and it thus represented more of a problem for different discourses rather than different languages. Although a term was easily translatable into Finnish on the term level, the Finnish participants questioned its usability and considered the translation as non-communicative, since the translations were in these cases more like word-for-word translations or at their best adaptations still reflecting the practices and needs of the source language and culture.

---

[103] A basic Google search on 5.5.2009 resulted in 18 200 records with *kotirouva*, and 308 records with *kotivaimo*.

Both the Finnish and the British informants considered that the **translation unit** is a **concept**, and not a term. The differences were therefore more concerned with **how to translate**, than what to translate, and different levels of equivalence were considered to be an unavoidable reality. Furthermore, due to the nature of the studied case – terms supposed to be foreign from the Finnish perspective – the question that invariably arose was whether the translation was needed (cf. dynamic equivalence – to whom) in the Finnish context and/or should its thesaurus context be similar to that in the source version. The British thesaurus constructors also questioned aiming at all other than non-equivalence. Among the Finnish participants this was still more strongly justified because different language versions of a thesaurus may differ in e.g. specificity of content.

However, and as stated before, the respondents did not present translating as an **ideal way** of constructing a multilingual and multicultural thesaurus, but the British informants were not as critical towards it as the Finnish ones were. The gained experience tended to favour not only in principle, but also in practice, more truly multilingual and multicultural working methods. (Cf. previous chapters and Hudón 1997 & 2004.) The thesaurus specialists generally prioritised usability in indexing and in information retrieval. They were not aiming at equivalence with dictionaries, but emphasised aspects of **pragmatic equivalence** (cf. Newmark 1988) – predictability, user-friendliness, coherence with the indexed and retrieved in material. Together with the gained experience, especially the question of non-equivalence was considered more properly than in the beginning, which led to the increased tendency to make non-equivalence relationships between different language versions. Artificial equivalents were generally not accepted as descriptors due to poor predictability in indexing and in information seeking situations (cf. Iivonen 1989; Fugmann 1994; ISO 2788-1986; Rowley 1988).

The informants did not make a distinction between the **different types of multilingual thesauri** which is made in this study – monocultural, multicultural and international. They considered the matters mostly from the perspective of an international and multilingual thesaurus, in which case a multilingual thesaurus is seen as thesaurus reflecting equally the different cultures and thus also aiming at a rather general content. Both the British and the Finnish respondents frequently raised the point that a multilingual thesaurus can also be multilingual within one language, i.e. one language version may also include foreign language terms. The reason for accepting terms other than those of the target language terms is a foreignisation strategy, which was generally considered the best strategy if there was a lack of usable equivalents in accordance with domestication strategy. In principle one culture was not considered to be superior to any other, and despite the major language (which the translation corpus represents) all the other language versions were normally considered to be adaptive to the needs of the majority.

On the general level, the **Finnish** thesaurus specialists **translated** "*according to the sense and meaning*"(cf. Reiss 1981/2000) and thus aimed at dynamic and functional equivalence. The following dynamic and functionalist viewpoint does not convey the

matters considered, and therefore the strategies used were also studied from other, partly different and partly more specific, points of view.

In the case of the practical examples studied, the Finnish thesaurus specialist preferred the **translation strategies** (in the most-favoured order) domestication, internationalisation, foreignisation and existential equivalence (cf. Lindfors 2001, Venuti 1995, Koskinen 2000, Nida 1964/2000 and cf. chapter 3.4.1.3 *Domestication or foreignisation – or internationalisation*). On a more general and theoretical level in the interviews, the strategies found among answers of the interviewed Finnish thesaurus constructors were internationalisation, domestication and foreignisation. Existential equivalence was also found, and the reason for this is that some terms were required among the majority of project participants (countries). The uppermost aim was user equivalence, and thus dynamic equivalence (cf. Nida 1964/2000). In Newmark's (cf. 1988) terms translations vary from word-for-word translations to communicative translations. In the translation projects with little (or even no) possibility to influence the translation corpus, word-for-word and literal translations were common. The Finnish informants were ideally or in principle often aiming for semantic translations and communicative translations, the emphasis being, in most cases, clearly on the target language users (cf. Newmark 1988). In accordance with semantic and communicative translation strategies the Finnish informants generally preferred descriptive phrases over foreign terms (cf. Nida & Reyburn 1981).

The **British** informants preferred internationalisation and foreignisation **strategies**, and the English translation corpuses included as a matter of course, only concepts fluently expressed in English. The commonly accepted idea of consensual equivalence (cf. 9.3.2), which also means a selected translation strategy, was thus in practice more frequently followed by the experts. For the Finnish and British thesaurus specialists the aimed for equivalences were (generally in most favoured order): conceptual, dynamic seeking, collection, and especially for the Finnish also consensual or existential equivalence. It was normal to accept different styles and motives – as long as the differiating meanings were reported in e.g. translation scope notes. The translation strategies were also not firm – there existed several different strategies within one project and those produced by the same person, and the selected stategy was very much dependent on what kind of translation problem the term discussed represented. For example, the same informant could strive for internationationalisation with regard to education systems, domestication with housewives, and foreignisation, when the term was considered as untranslatable into Finnish (cf. FI TS3).

As in the Skopos-theory (cf. Reiss & Vermeer 1989 and chapter 3.4.2.2 SKOPOS-theory) translating was considered to be a **decision-making process** by both the Finnish and British thesaurus constructors. It was also a shared experience, although this was more strongly apparent among the British informants, and during the process the aims became more realistic which also influenced the content and structure of thesauri.

According to the British and Finnish informants not only collegial support is needed in thesaurus construction, but also **co-operation** with users (i.e. information seekers and indexers) and terminological specialists. If the thesaurus is also meant to be used internationally and/or multiculturally, native speakers of all the language versions

were considered essential. This principle is in line with the guidelines presented in the standards and guidebooks, which point out that close cooperation is essential for the successful construction of thesauri (cf. chapter. 3.3 *Thesaurus Construction*). In practice the multilingual projects among the British informants were the constructing of the translation corpus and among the Finnish informants of two kinds – translation projects aiming at multicultural or international content, and multilingual projects, reflecting one culture – i.e. the content was monoculturally Finnish.

The **working methods** among both the Finnish and British thesaurus specialists tended to progress during the thesaurus project in a more democratic direction (from a centralised or de-centralised structure to a semi-centralised structure, cf. ISO 5964-1985). Additionally, the idea of non-equivalence became more accepted as more experience was gained concerning multilingual and multicultural thesaurus construction.

Despite the general views on selected translation strategies there are also **more specific issues** to consider in practical multilingual thesaurus construction work, which can be reflected from more detailed perspectives (like in the study) in terms of Koller's (1989) factors (cf. chapter 3.4.1.2 *The idea of equivalence*).

The Finnish and British thesaurus specialists aimed at equivalence on a **denotative** level, and even though the thesaurus constructors had given stronger negative connotations regarding the studied term(s), they prioritised denotation over connotation (cf. Wierzbicka 1991; Nida 1975). In the case of family role terms it was considered difficult, and sometimes even impossible, to aim for connotative equivalence due to the differences on the cultural level. Different subcultures were also discussed, and the problems were considered not only to appear on the language level, but also on the discourse level.

Text-normative matters caused serious problems – for example *homemakers* could have had a Finnish equivalent which would have been completely satisfying in another type of text. However, in the thesaurus context the nature of documentary language was the main barrier in attaining a true (or exact) equivalent. The text-normative factors were connected to the pragmatics – if the translation equivalent was not considered to be in terms of style, relevance and usability etc. a descriptor, it could not be predictable for the users and thus a competent term in indexing and in information seeking situations.

Formal factors discussed by the respondents concerned terminological issues and the terms' context in a thesaurus - a well-established term being translated with a descriptive phrase or a thesaurus context seen as too foreign or narrow. The interviewed thesaurus specialists were aware of e.g. ISO and SFS standards regarding thesaurus construction, but in their translations they did not motivate their decisions referring to literature or guidelines about thesaurus construction. The terms typical for thesaurus literature and standards such as non-equivalence and partial equivalence were still constantly used in the arguments.

Overall, the differences among British, Finnish and international thesaurus discourses were rather minor. The Finnish participants that were interviewed were more culturally homogeneous. On the term selection level, harmonising with The General Finnish Thesaurus YSA was aimed at, but in conflict situations the tendency

was, among the Finnish participants, to also favour the practices and suggestions received from the specific group of potential users. In thesaurus construction standards and guidelines vocabulary work comparable with a clear and logical decision-making process, the result of which is one good (or best possible) solution. However, especially the Finnish thesaurus specialists interviewed spoke about this as more of a compromise-seeking process, where they stressed, their result could have also been something quite different.

10 Conclusions and Discussion

"*There is no unbiased look when one starts ---*" (Finnish thesaurus constructor)

This final chapter summarises the research, reflects the process and presents some ideas for future research. First (in chapter 10.1), the aims of the study and research design are reported and discussed, and reflections on the process are also given. Second (in 10.2), the research results by different discourse types are summarised, which is followed (in 10.3) by the theoretical consideration concerning different types of multilingual thesauri, translation strategies, and pragmatic indexing term equivalence. Finally (in 10.4), some ideas for the further research are indicated.

10.1 Aims and research design

"*The place of practice and of theory, for the translation of any text, is the place of its practice.*" (Meschonnic 2003, 344)

Different types of languages can be approached from many perspectives. This study approached the challenges of multilingual thesaurus construction from the perspectives of semantics and pragmatics. In semantic analysis the focus is on what the words conventionally mean and in pragmatics on the 'invisible' meaning - or how we recognise what is meant even when it is not actually said (or written) (Yule 1996; Hudson 1998). When an information-seeker makes an information search in a database (s)he operates with at least five different languages: the authors, the indexers, the synthetic structure, the users and the search strategy (Buckland 1999), which all represent a type of discourse (cf. Fairclough 1992) and which should be considered in thesaurus construction.

In the research, multilingual thesauri were approached as cultural products and the focus was twofold: On the empirical level the focus was placed on the translatability of certain British-English social science indexing terms into the Finnish language and culture at a concept, a term and an indexing term level. On the theoretical level the focus was placed on the aim of translation and on the concept of equivalence. In accordance with modern communicative and dynamic translation theories (see Nida 1964/2000; Reiss & Vermeer 1986; Vermeer 1989; Vehmas-Lehto 1999) the interest was on the human dimension and translating was seen as a decision-making process. The research is qualitative.

As a starting point for the empirical section, terms related to family roles (*family roles, breadwinners, heads of household, homemakers, housewives*) have been used. The studied terms were selected from the working version of the European Language Social Science Thesaurus (ELLST, see Forsman & Keränen 2002; Jääskeläinen 2006

and appendices 5, 6, 8) which were considered to be typical social science translation problems representing culture-bound and abstract concepts. Thus the study focused on the examination of the function of translation in multilingual thesaurus construction, and on the equivalence type driven by the focus. Translation strategies have not been studied before within the framework of information science, whereas equivalence has been defined (see ISO 5964-1985; Aitchison et. al. 1997).

The theoretical framework used was in its nature cross-disciplinary and especially the theories from translation science have been extended to deal with a new area of application: thesaurus construction. The reader of the study should be aware of the fact that translation theories are often formed on the basis of literature, which as a translation unit differs greatly from thesauri. It is also noteworthy to mention that the study itself was not multicultural, that its starting point was in fact monocultural. (A multicultural study would require not only multicultural material, but also multicultural research design and analysis.)

Languages and ideas expressed by languages are created mainly in accordance with expressional needs of the surrounding culture (Suojanen 1993) and thesauri were considered to reflect several subcultures and consequently the discourses which represent them. The research material consisted of different kinds of potential discourses: dictionaries, database records, and thesauri, Finnish versus British social science researches, Finnish versus British indexers, simulated indexing tasks with five articles and Finnish versus British thesaurus constructors. In practice, the professional background of the two last mentioned groups was rather similar. It became even more clear that all the material types had their own characteristics, although naturally not entirely separate from each other. It is further noteworthy that the different types and origins of research material were not used to represent true comparison pairs, and that the aim of triangulation of methods and material was to gain a holistic view (cf. Denzin 1978).

Multicultural thesaurus construction faces severe challenges. A monocultural thesaurus should reflect the discourses of indexed material and answer to the needs and practices of end-users both as indexers and information seekers. The success is dependent on the aims and focus of thesaurus. The research design is in line what has been previously found. Lykke Nielsen (2002, 16) has stated that thesaurus constructors must make up their mind whether it is possible to assign a firm definition to any word or whether words have fuzzy meanings, and that the decision made will obviously affect our view of the role of thesauri and the semantic information, needed in the thesaurus. Meriläinen (1997, 110) has studied the concept of equivalence in the ISO 5964 – standard and stated, that when "considering the equivalence concept found in the ISO 5964 in light of the linguistic discussion on equivalence, it seems evident that the descriptor equivalence is an instance of dynamic equivalence. This conclusion is based on the fact that the thesauri include also such equivalencies which could never be tolerated in natural language use. These equivalence relationships depend on the function of thesaurus and are based on non linguistic arguments." (Ibid.)

10.1.1 Research questions

"*In the study of languages, one can safely assume nothing*." (Hall 1981, 101)

The general **research questions** were (see also chapter 5 *The Study*):

1. Can differences be found between Finnish and British discourses regarding family roles as thesaurus terms, and if so, what kinds of differences, and what are the implications of these differences for multilingual thesaurus construction?

2. What is the pragmatic indexing term equivalence?

The **results** were clear:

1) It was possible to identify different discourses. There also existed sub-discourses. For example within the group of social scientists the orientation to qualitative versus quantitative research had an impact on the way they reacted to the studied words and discourses, and indexers placed more emphasis on the information seekers whereas thesaurus constructors approached the construction problems from a more material based solution. The differences between the different specialist groups i.e. the social scientists, the indexers and the thesaurus constructors were often greater than between the different geo-cultural groups i.e. Finnish versus British. The differences occurred as a result of different translation aims, diverging expectations for multilingual thesauri and variety of practices. For multilingual thesaurus construction this means severe challenges. The clearly ambiguous concept of multilingual thesaurus as well as different construction and translation strategies should be considered more precisely in order to shed light on focus and equivalence types, which are clearly not self-evident. The research also revealed the close connection between the aims of multilingual thesauri and the pragmatic indexing term equivalence.

The different main discourses are discussed in more detail in chapters 10.2.1-10.2.3.

2) The pragmatic indexing term equivalence is very much context-depended. Although thesaurus term equivalence is defined and standardised in the field of library and information science (LIS), it is not understood in one established way and the current LIS tools are inadequate to provide enough analytical tools for both constructing and studying different kinds of multilingual thesauri as well as their indexing term equivalence. The tools provided in translation science were more practical and theoretical, and especially the division of different meanings (cf. Koller 1989; Vehmas-Lehto 1999) of a word provided a useful tool in analysing the pragmatic equivalence, which often differs from the ideal model represented in thesaurus construction literature (cf. esp. exact equivalence in ISO 5964-1985; Aitchison et. al. 1997).

The study thus showed that the variety of different discourses should be acknowledged, there is a need for operationalisation of new types of multilingual thesauri, and the factors influencing pragmatic indexing term equivalence should be discussed more precisely than is traditionally done.

The theoretical implications and new operationalisation of multilingual thesauri, translation strategies and the pragmatic indexing term equivalence are discussed in more detail in chapter 10.3.

10.1.2 The process

"*It takes time to internalise the aim, that what for is one making this and that.*" (Finnish thesaurus constructor)

The **material** was selected and gathered with the aim of reflecting what the different needs and practices in multilingual and/or multicultural thesaurus construction involve from a holistic perspective. It was considered necessary to limit the focus in order to make the practical research work realistic. As a language-pair (and perspective) translatability from British-English into Finnish was selected and as a thematic terminological starting point family roles terminology within social sciences was selected as to the way it is represented in the working version of ELSST.

The limitation was considered to be successful, as the selected theme allowed the societal and cultural context to be taken into consideration (see chapters 3.1 *Language, culture and meaning* and 4 *Linguistic and social context*). It also represented a typical problematic case within the most common working form of multilingual thesaurus construction represented in the material – translating (cf. Hudón 1997 and 2001) and was acknowledged as such especially by the Finnish respondents. The problems occurred on different levels of the studied case – on the concept, term and indexing term levels. If the cases would have been more concrete, for example, then the result could have been different. The terminological case thus represents abstract culture-bound social science concepts.

Halverson (1998, 7) states in the context of translation studies and representative corpora that the decisions made as regards the use of various data sources that it must "be made on the basis of theoretical notions pertaining to the relevance of various types of text, and the relative significance of the different types. It must be also mentioned that practical concerns such as time and money quite often factor into the decision as well." (Ibid.) While the case was understood as optimal, the proceedings with British respondents was considered as somewhat risky. All the British interviews were conducted consecutively and there was no time for reflecting on the first ones and making improvements to the interview schema. The risk was taken due to financial reasons. It transpired that the British social scientists were generally not as skilful in information seeking as the author had expected and thus practices (i.e. use of thesauri) axiomatic to the author were foreign to the respondents (cf. Adler 1997, 78-87; Hall 1981, 20-33; Rowley 1988, 48-49; Forsman 2005,36). Only a few of the social scientists interviewed (one British and two Finnish) said that they have sometimes used a thesaurus in information seeking. The use of thesauri in keywording of their own documents was also unfamiliar, although all of them were used to giving keywords.

They often experienced information seeking as an oppressive and time consuming thing to do. More importantly, the respondents (thesaurus constructors, indexers, social scientists) interviewed as specialists were true professionals in their field, and aware of different discourses and of different ways to approach indexed matters, although they did not necessarily speak directly about discourses. They also reflected on their own background as to the way they tend to see things (cf. Adler 1997; Hall 1981).

The focus on translation in the answers provided by respondents was in a more constant flux than expected. This led to changes in the analysis and made the principal and pragmatic matters (aims of translations) more important and consequently the importance of the semantic analysis at the indexing term level was reduced. After adapting to the idea of different kinds of material received, the result was considered to be fruitful.

The discourse analysis is more of a theoretical framework than a clear method (Potter & Wetherell 1990; Suoninen 1992; Jokinen & al. 1993) and its usefulness depends on the conceptual tools used. After the interviews were conducted, the role of translation science as an essential part of the theoretical framework became even more important, since more functional analysis tools were needed, and thus a rather linguistic orientation was selected and a modification made into a more attitudinal study. The theoretical and empirical part of the study were thus in a process of interaction during the entire research process.

It was a surprising fact for the author that it was easier to obtain information specialist with experience of thesaurus construction as respondents, than respondents on indexing. It become apparent that in many organisations indexing was done on the principle of "copy-paste" (by looking up indexing in another database, that had already been completed, and copying it) and thus indexing as a means of providing descriptors was not necessarily seen as intellectual work.

The material was also in some cases so small that it was not possible to make any deeper analysis. However, the information was still considered valuable as a means of describing the use and/or potential usefulness of a certain term in a certain context.

The **methods** used were in general functional. Focused interviews (see Hirsjärvi & Hurme 2000), as a method, are very time-consuming and include several phases before the actual analysis can take place such as, finding relevant respondents, time scheduling, travelling, interviewing, transcribing - altogether the 27 material collecting interviews with the 29 respondents took 41 hours 15 minutes (in addition two test interviews and two consulting interviews had been made). It was also a particularly time-consuming method for indexers, who in addition to the interviews participated in a prior, simulated indexing task. Without a thematic case it would not have been realistic to conduct material as in the study with social science experts (as researchers, but also as indexers and thesaurus constructors). The theme also influenced the time – with more heterogeneous terminology the time aspect would have become an even greater issue.

The co-word analysis (see Persson 1991) was used in an untraditional way to illustrate the qualitative material. In the case of this type of qualitative material, the bibliometric maps give both a compact and a rich picture of the lexical network and co-

occurrences of the studied terms. The problems related to this type of analysis were traditional - e.g. the use of harmonised versus authentic material (cf. ibid.) and the limitations and quality of databases (cf. von Ungern-Sternberg 1998: 1994) and were therefore a question of constant balancing. – In order to illustrate main stream or general phenomena, it is necassary to harmonise different versions of linguistic expressions, but at the same time a lot of nuances can be lost and this also includes a risk of misinterpretation in the data handling process. With the interview material it was an even more demanding and problematic method to use than expected. First, the word associations given were often not single words or phrases, but rather long considerations. Second, when given in Finnish, they needed to be translated into English, and the source material included a lot of problematic words, which were not always easy to translate. When the material was slightly modified in order to be presented as a co-word map, some nuances were lost. Due to different indexing practices the analysis was made more on a term level than on an indexing term level by separating pre-coordinated indexing terms into separate terms (e.g. "*kotityöt*/homework – *miehet*/men, *koti-isät*/stay-at-home fathers" were in the analysis handled as three terms: homework, men, stay-at-home fathers). The clear benefit of co-word analysis is that it makes the contextual equivalence more observable. Its usefulness is thus at the contextual information level. It does not provide answers as to how to translate, but it can be used to test different alternatives. By comparing several different kinds of contexts (e.g. databases) and also by interviewing several groups the missing aspects can be found – often the aim is reversed and thus it is possible to find out what is missing in the maps.

The word association method resulted in different, but mainly positive reactions. In some cases a clarification and an assurance was needed that there were in fact no wrong answers, and that in the analysis the author was after a holistic image, not word associations provided by single respondents. The motives criticised may be due to the way the word association method is used in psychology (cf. information science in Lykke Nielsen 2002) as in the description provided by the *Center for Psychological Studies* (2005): "Because verbalizations reflect ideation, it is assumed that the systematic study of associative thought can reveal information about an individual's personality characteristics, areas of emotional disturbances, and the like."

Despite the methodological shortcomings, the word association method was considered fruitful in this context. The attitude of the respondents towards making word associations was generally positive, although the context was clearly more controlled, and some respondents stayed more within the role of a thesaurus constructor rather than giving free associations. – In those cases more comments were given during the discussion of definitions. The associations were also in another way "thesaurus-biased" - the thesaurus specialists involved in ELSST project tended to partly follow the experience gained in the project. However, the more problematic the term, the more soul-searching the discussions became, and the word associations provided not only contextual information about the content of the studied terms, but also functioned as good stimulus for further discussions, particularly when the informants went back to their associations and explained how a certain term feels in use and why.

It became clear that the challenge of our field is to try to make the search of multicultural and human related abstract topics less oppressive. The norms and expectations towards documentary language and thesaurus construction should be considered more carefully and not highly concentrate on the linguistic aspects such as whether a descriptor is short, unambiguous etc. One should consider in more detail not only the practices, but also the aims of indexing in a multilingual context.

10.2 The discourses

"*And then language changes, doesn't it, as social perceptions change the use of word changes*." (British indexer)

The different discourses are represented in a material-based way. The method of proceeding reflects more precise insight into thesaurus construction. The discourses begin with the social scientists, and then proceeds to the indexers and ends with the thesaurus constructors.

10.2.1 Academic discourse

"*So I think it is quite dangerous to lots of words… ---*" (British social scientist)

The academic discourse within a theme (family roles terminology) was explored and analysed via six English language online dictionaries and one Finnish-English online dictionary (all provided for academic use on Finnish university web-pages) and by interviewing twelve social scientists (equally British and Finnish) whose research areas represent different aspects of family roles (such as reconciliation of work and family, being stay-at-home father, socio-economic differences of family members in different countries etc).

The dictionary definitions are poor when considering the pragmatic meanings of the studied terms (see also Wierzbicka 1996). The interviewed respondents tended to provide much more information about the connotative aspects, style and gender-relatedness of the studied terms. *Breadwinner*, for example, was generally considered to be a male-related term, whereas only two of the studied dictionaries indicated (with dynamic variations) it to be a male-related term, and paradoxically, one a feminine-related term. The discourse of dictionaries versus the discourse of researches was in this sense heterogeneous, that in the studied social science discourses the meanings are more precise, although not necessarily more narrow than in the common language discourse of the dictionaries studied.

In the terminological case of family roles the Finnish social scientists emphasised more similarly and equally the pragmatic and the connotative meaning rather than the

semantic and denotative meaning, and thus – although aiming at communicative equivalence (cf. Newmark 1989) the studied terms were not usually accepted as descriptors. This has something to do with the thematic case, but also with the barrier with (documentary) thesaurus language as opposed to (natural) common and scientific language. The result might have been different if the respondents had been more familiar with the function and practice of thesauri.

In general, the studied English terminology was more acceptable to the British informants than to the Finnish respondents as words, terms and descriptors, although the concepts were also often regarded as problematic by the British respondents. For example *head of household* was generally not accepted as a word in common language usage, but considered to have a negative connotation.

Selection of a translation strategy was a complex issue and influenced by personal and general reasons. The researchers' qualitative versus quantitative research orientation influenced the selected translation strategy – the terms were more acceptable for respondents using quantitative material and their translations were more semantic than for the respondents using qualitative material. - It is noteworthy, that the origin and using context of the ELSST and therefore also of the terminological case used in this study reflect mostly perspectives of quantitative material; the aim of the ELSST is to provide a tool for indexing and retrieving mainly quantitative material. The respondents, especially those representing a more qualitative research orientation tended to favour more general terms (also in indexing and in information seeking) such as parenthood. Both the Finnish and the British informants thought that it is better to use politically correct terms (cf. Ulrych 1992; Venuti 1998) and thus, for example, to avoid genderism, although they thought, that it is in a way artificial – in real language usage for instance *homemaker* refers to women (cf. Engelberg 1993 and 1998).

Pragmatic and communicative equivalence was often prioritised over to semantic translations and meanings (cf. Koller 1989). In general, the British interviewees were against internationalisation of terminology and some even clearly prioritised domestication in the sense that they considered a rich and nunceful English terminology an essential and preferable to a more general and easily translatable terminology. The Finnish informants followed rather consistently a domestication strategy when trying to find equivalents relevant from the perspective of Finnish practices (cf. Venuti 1998; Lindfors 2001). They did not provide equivalents according to the existential equivalence (cf. Koskinen 2000), but followed the ideas of the Skopos-theory (cf. Reiss & Vermeer 1986; Vermeer 1989; Koskinen 2001) and thus clearly considered the function of the translation they provided. If thesauri are to also to reflect social science discourse the connotative aspects should be considered more carefully than the common guidelines indicate.

10.2.2 Indexing discourse

"What does it actually mean, does not probably refer to anything concrete…
[laughing]" (Finnish indexer)

The discourse of indexing was explored by studying how a certain terminology was used in three databases, by interviewing three Finnish and three British indexers and by conducting a simulated indexing task using five journal articles.

The emphasis in the analysis of the Finnish ARTO and LINDA databases and the British COPAC database was the time period between 1990-2003, and on descriptors and title words. In the analysis the limitations of the systems were clearly seen, these were found to be various types of practices and in the COPAC database also the lack of common vocabulary tool used in indexing (cf. von Ungern-Sternberg 1994).

The terms studied were known to be problematic and they were not common indexing terms in the studied databases. However, some clear tendencies could still be found. It became apparent that the same terms and their translations were used differently in the databases, and that there existed also differences between the Finnish databases. The discourse of databases is thus not unified within one terminological case. Even when the terms were understood similarly the context used might differ and often the terms were used in a broader or narrower sense than in the thesaurus discourse.

The number of terms representing a transliteration problem rather than a translation problem was less than one tenth in ARTO and LINDA, and in COPAC about one fifth. The indexing style was more gender-neutral in the Finnish databases than in the British, and thus clearly gender-related terms were not as much used in the indexing of the Finnish databases. The reason for this is thought to be the cultural context (see chapter 4 *Linguistic and social context*), which is also reflected in the General Finnish thesaurus YSA.

Issues such as domesticity, modernity and male gender were commonly "hidden" in the indexing of studied samples and consequently in bibliometric maps unobservable or biased matters (cf. Chen 2003; Venuti 1995; Yule 1996). It can be questioned whether a male gender, domesticity and/or a recent time aspect can be considered to be a norm in the indexing – but how are they later seen in the information seeking results, do they become "invisible"? A concrete question in an information retrieval situation is how is it possible to aim for preciseness in searches if the information required is e.g. about men's roles in families today.

Together with these problematic matters it is also reasonable to ask which time frame is to be reflected in thesauri and thesaurus translations, should it only be one time period? Could thesauri be constructed and updated in a way that allows different time aspects? The Sociological thesaurus of Cambridge Sociological Abstracts (CSA), for example, also includes information about a descriptor's history (e.g. when the descriptor *housewives* was replaced with *homemakers*). Providing information about past practices is a new trend that is also found in HASSET (Humanities and Social Science Electronic

thesaurus, e.g. within the scope note of the term *homemakers* - "term created Sept. 2000. Prior to this the term "domestic responsibilities" might have been used).

In a multilingual thesaurus construction context the British and Finnish indexers considered knowledge on subject domain to take priority over the language knowledge.

The simulated indexing task revealed clearly that the Finnish and British indexers considered their everyday indexing differently. The British indexers were, in general, more European or international, whereas the Finnish indexers considered their perspective in indexing as evidently Finnish (cf. Hall 1981; Adler 1997; Koskinen 2000). This has to do with the working environment and potential database audiences, and with the languages (Finnish as Finnish, and English also as a global lingua franca, see chapter 4.1.2 *On Finnish and English*).

In indexing, the aim was to describe both the general and specific level of the topic, and it was also a clear guideline (cf. Rowley 1988). Different discourses were also considered (cf. Buckland 1999; Lehtonen 1994; Venuti 1998; Vehmas-Lehto 1999). Together with different discourses time aspect (cf. Aitchison 1991; Suojanen 1993; Hirsjärvi and Hurme 2000) was also often discussed and it was seen as a construction problem as well as an update problem. The problems become more challenging especially in the context of gender-neutral language use (cf. guides for non-discriminatory language use as UCL 2000; Brunel 1999; School of Psychology 2002; University of Southampton 1997; Engelberg 1993, 1998). In all the cases considerable disparity occurred in the terminological and conceptual differences (cf. Iivonen & Kivimäki 1998), and the indexing was remarkably more coherent on general theme level (cf. Iivonen 1989). As with the database indexing studied (of ARTO, LINDA and COPAC) there were variations especially as to whether to express the indexed study's time, place and methods of research.

The Finnish and British indexers often made (in practice) a clear difference between different types of meaning and especially between semantic and pragmatic meanings (cf. Koller 1989; Lehtonen 1994; Buckland 1999; Vehmas-Lehto 1999) when discussing the terminological case. In most cases, a clear difference was seen between semantic and pragmatic meanings especially on gender-relatedness (cf. Lehtonen 1992; Engelberg 1993; Yule 1996; Venuti 1998; Vehmas-Lehto 1999). The studied cases as concepts, terms and descriptors were, in general, considered more broadly than in the thesauri studied.

There was a gap between theoretical and practical discussions regarding translatability into Finnish. Discussion on the more principal level of the aspects discussed by the respondents can be related to Koller's (1989) equivalence types such as denotative, pragmatic and formal equivalence, and the connotative and text-normative aspects of equivalence were not raised when discussing the terminological case. The respondents considered equivalence similar to dynamic equivalence (cf. Nida 1964/2000; Vehmas-Lehto 1999). Their representations were closest to the domestication strategy (cf. Venuti 1998; Lindfors 2001) when they were aiming at translations that reflected the target culture, but in practice (cf. simulated indexing task situation) all preferred internationalisation (cf. Koskinen 2000). In the Finnish context,

the impact of the general nature of the used controlled vocabulary tool YSA (The General Finnish Thesaurus) was clearly seen.

It also became obvious that the aims of multilingual indexing should be discussed more precisely. One Finnish respondent mentioned that multilingual indexing may not necessarily need to aim at giving equivalents in many languages, but be used to aim at complementing monolingual indexing.

10.2.3 Thesaurus discourse

> "*But it's often terribly difficult to choose what would be the wise thing to do*." (Finnish thesaurus constructor)

In order to illustrate the discourse of thesauri from the perspective of multilingual thesaurus construction nine thesauri (mono- and multilingual, general and of social sciences) were analysed and six British (two double interviews) and five Finnish thesaurus constructors were interviewed.

The study of thesaurus discourse in comparison to academic discourse revealed that thesauri operate on the denotative level of the words but, as discussed in the context of academic language usage, if we want to improve predictability in an information seeking context we should also pay more attention to the connotative level, since the connotative meanings can cause a major obstacle in the use of multicultural thesaurus within social sciences (cf. Ulrych 1992).

The overall discourse of thesauri also shows that thesauri conform to the traditional language usage and practices. The terms were understood in accordance with general language usage, although their use in thesauri and databases was narrowed in order to cover certain aspects. A greater consistency among different thesauri existed on a more conceptual and/or theme than on an indexing term level (cf. Iivonen 1989 and 1995, Iivonen & Kivimäki 1998). The thesauri studied also represent different sub-discourses. The scope of the studied thesauri and the indexing practices they are part of was clearly seen in the material (cf. Hjørland 1997; 2002 and 2004).

The thesauri studied did not provide detailed information about how they have understood equivalence, which may be detrimental to their use since in practice the idea of equivalence, the norms followed, and needs and expectations of respondents varied considerably. The frequently mentioned guidelines to follow were thesaurus construction standards provided by the International Organization for Standardization (ISO) and by The Finnish Standards Association (SFS), but they do not provide information as to how the aim of the thesaurus has been understood.

The multilingual thesauri studied were in their nature rather international and the underlying basic assumption about equivalence was that each descriptor had an equivalent in other languages, but the non-descriptors may vary and are dependent on the current language version. The dynamic nature of language was thus acknowledged,

and formal equivalence was aimed at to a certain degree, but this was felt to be still rather strict (cf. Meriläinen 1997; Nida 1964/2000). The result conforms to the ideas commonly represented in thesaurus construction guides and norms. It is noteworthy that when compared to the main translation strategies such as domestication, foreignisation and internationalisation, the ideology behind the ISO 5964-1985 standards is closest to the idea of internationalisation. In practice this means that the content of the thesauri should be designed in such a way that it is possible to aim at equivalence between different language versions, and therefore clearly culture-bound words should not be included.

Together with the translatability and equivalence issues studied the thesaurus experts often brought up the text type and thesauri were considered to represent a special kind of discourse (cf. Reiss 1981/2000; Buckland 1999; Lehtonen 1994; Potter & Wetherell 1990; Pälli 2003) with its own traditions. The respondents also considered that the lack of common understanding of the nature of a thesaurus (between indexers and information seekers) caused problems in thesaurus construction and use.

The thesaurus constructors acknowledged both cultural and linguistic reasons for making the construction of multilingual and multicultural thesauri difficult. The terms which were considered most difficult to translate into Finnish by the Finnish informants interviewed were also considered to be complicated by the British thesaurus specialists. The difficulty was often considered to be caused by different discourses, as well as by time, which changes the meaning and/or the style of the words (cf. Aitchison 1984; TSK 1989; Varantola 1990; Wierzbicka 1997; Buckland 1999; Lehtonen 2000; Katan 2004; Forsman 2005).

Although the informants were constantly very much aware of the discourse of thesauri, they did not always follow the norms of documentary language and in the most exceptional case even a neologism was suggested. The reason for such dispersion and flexibility is purportedly not only due to the problematic terms, but also to the lack of proper methodological tools for analysing, determination and motivation of the aimed equivalence, which has not previously been discussed in detail in LIS prior to this research. The informants themselves raised the point that social sciences are an especially difficult field to handle due to different needs and discourses as well as aims and norms of communication. The spontaneous interview situation may also have influenced the results, and the thesaurus experts as well as other interviewed specialists might have acted in a different way, if more time had been allowed. However, it is worth noting that the topic was not new to the respondents.

The British informants did not discuss linguistic difficulties as much (between different languages) as the Finnish informants did, due to the research design – the British informants were not asked to translate anything and they represented a more translation corpus based view. It was usual for the Finnish informants to emphasise cultural aspects, and to consider the linguistic matters as sometimes even more difficult matters to handle in a multilingual thesaurus. In general, all the informants talked more about differences in cultures and cultural systems, and not very much about linguistic matters (cf. Nida 1964, 130) and the differences in culture were considered to cause linguistic differences. The cultural differences led to internationalisation discussions,

and it was generally stated that if a term is too culture-specific or relevant to only a small cultural group it should not be included in (an international) multilingual thesaurus.

The nature of documentary languages and thus also of descriptors caused many obstacles in the translation process – in some cases the foreign concept was considered to be describable in Finnish or in English, but to express it with a term suitable and usable for indexing and information seeking purposes was far more complicated, and it represented thus more a problem of different discourses than different languages.

Both the Finnish and the British respondents considered the translation unit to be a concept, and not a term. The differences, therefore, were more concerned with how to translate rather than what to translate, and different levels of equivalence were considered to be an unavoidable reality. Furthermore, due to the nature of the studied case – terms supposed to be foreign from the Finnish perspective invariably led to questions as to whether a translation was needed (cf. dynamic equivalence – to whom) in the Finnish context and/or should its thesaurus context be the same as the one used in the source version.

The thesaurus experts did not consider translating as an ideal way of constructing a multilingual and multicultural thesaurus, but the British respondents were not as critical towards this as the Finnish ones. Gained experience was a factor that tended to favour not only in principle, but also in practice, more truly multilingual and multicultural working methods (cf. Hudón 1997 & 2004 and ISO 5964-1985.) The thesaurus specialists generally prioritised usability in indexing and in information retrieval. They were not aiming at equivalence with dictionaries, but emphasised aspects of pragmatic equivalence (cf. Newmark 1988; Koller 1989) – predictability, user-friendliness, coherence with the indexed and retrieved in material. Together with the practice of gaining experience, the question of non-equivalence was especially considered in more detail than in the beginning of the projects, which led to an increased tendency to make non-equivalence relationships between different language versions. Artificial equivalents were generally not accepted as descriptors due to poor predictability in indexing and in information seeking situations (cf. Iivonen 1989; Fugmann 1994; ISO 2788-1986; Rowley 1988).

The respondents did not make a distinction between the different types of multilingual thesauri which has been made in this research – monocultural, multicultural and international. They considered the matters mostly from the perspective of an international and multilingual thesaurus, in which case a multilingual thesaurus is seen as a thesaurus reflecting equally different cultures and thus also aiming at a rather general content. Both the British and the Finnish informants introduce the topic that a multilingual thesaurus can also be multilingual within one language version, i.e. one language version may also include foreign language terms. The reason for accepting other terms rather than a target language term is a selected foreignisation strategy, which was commonly considered to be the best strategy if there is a lack of usable equivalents in accordance with a domestication strategy. In principle one culture was not considered superior to any other, and despite the major language (which the translation corpus represents) all the other language versions were generally considered to be adaptive to

the needs of the majority. The problem here is that usually the working language is often a major language such as English and in that way already initially represents the majority, especially if counted in the numbers of potential language users. Thus this aspect is related to problems common to language and power in general (see e.g. Mälkiä & Stenvall 1997).

On the general level Finnish thesaurus specialists aimed at dynamic and functional equivalence. The followed dynamic and functionalist viewpoint does not relate all the matters considered, and therefore the strategies used were also studied from other, partly different and partly more specific, points of view.

In the case of the practical examples studied the Finnish thesaurus specialists preferred these translation strategies (in the most-favoured order): domestication, internationalisation, foreignisation and existential equivalence (cf. Lindfors 2001, Venuti 1995, Koskinen 2000, Nida 1964/2000 and cf. chapter 3.4.1.3 *Domestication or foreignisation – or internationalisation*). On a more general and theoretical level in the interviews, the strategies found among the answers of the Finnish thesaurus constructors were internationalisation, domestication and foreignisation. Existential equivalence was also found, and it was justified by the fact that some terms are essential among the majority of project participants (countries). The uppermost aim was user equivalence, and thus dynamic equivalence (cf. Nida 1964/2000).

In Newmark's (cf. 1988) terms translations vary from word-for-word translation to communicative translations. In the translation projects with little (or even no) possibility of influencing the translation corpus, word-for-word and literal translations were common. The Finnish informants were ideally or in principle often aiming for semantic translations and communicative translations, the emphasis being in most cases clearly on the target language users (cf. Newmark 1988). In accordance with semantic and communicative translation strategies the Finnish informants generally preferred descriptive phrases over foreign terms (cf. Nida & Reyburn 1981), although at the principal level were also willing to include the translation unit as a direct citation loan and clearly as a foreign concept to the Finnish culture.

The British informants preferred internationalisation and foreignisation strategies, and the English translation corpuses included as a matter of course only concepts fluently expressed in English. The commonly accepted idea of consensual equivalence, which also means a selected translation strategy, was thus, in practice, followed more by the Finnish thesaurus experts. For the Finnish and British thesaurus specialists the aimed equivalences were (generally in the most favoured order) conceptual, dynamic seeking, collection, and especially for the Finnish also consensual or existential equivalence. It was normal to accept different styles and motives – as long as the differiating meanings were reported in e.g. translation scope notes. The translation strategies were also not fixed – several different strategies also existed within one project and by the same person, and the selected strategy was very much dependent on what kind of translation problem the term discussed represented.

As in the Skopos-theory (see Reiss & Vermeer 1989) translating was considered to be a decision-making process by both the Finnish and British thesaurus constructors. It

was also a shared experience, although this was more strongly apparent among the British informants, and during the process the aims became more realistic and also influenced the content and structure of thesauri.

Thesaurus experts emphasised the importance of rich co-operation and stated, that not only collegial support is needed in thesaurus construction, but also co-operation with users (i.e. information seekers and indexers) and terminological specialists. If the thesaurus is also meant to be used internationally and/or multiculturally, native speakers of all the language versions were considered as essential. The principle is in line with the guidelines represented in standards and guidebooks, which point out close cooperation as essential for the successful construction of thesauri (cf. chapter. 3.3 *Thesaurus Construction*). In practice, the multilingual projects were performed by the British informants constructing the translation corpus and those performed by the Finnish informants were of two kinds – translation projects aiming at multicultural or international content, and multilingual projects, reflecting one culture – i.e. the content was monoculturally Finnish.

The working methods among both the Finnish and British thesaurus specialists tended to develop during the thesaurus project in a more democratic direction (from a centralised or de-centralised structure to a semi-centralised structure, cf. ISO 5964-1985). The idea of non-equivalence also became more accepted as more experience was gained in multilingual and multicultural thesaurus construction.

Despite the general views on selected translation strategies there are also more specific issues to consider in practical multilingual thesaurus construction work, which can be also ascertained from more detailed perspectives (as in the study) in terms of Koller's (1989) factors (cf. chapter 3.4.1.2 *The idea of equivalence*).

The Finnish and British thesaurus specialists aimed at equivalence at a denotative level, and even though the thesaurus constructors had given even stronger negative connotations to the studied term(s), they prioritised denotation over connotation (cf. Wierzbicka 1991; Nida 1975) and acted thus in an opposite way to social scientists. In the case of family role terms it was considered difficult, and sometimes even impossible, to aim at connotative equivalence due to the differences at the cultural level. Different subcultures were also discussed, and the problems were considered not only to appear at the language level, but also at the discourse level.

Text-normative matters caused serious problems and in the thesaurus context the nature of documentary language was the main barrier in attaining a true (or exact) equivalent. The text-normative factors were connected to the pragmatics – if the translation equivalent was not considered to be a descriptor in terms of style, relevance and usability etc. it could not be predictable for the users and thus a competent term in indexing and in information seeking situations.

Formal factors, discussed by the thesaurus experts, concerned terminological issues and the terms' context in a thesaurus - a well-established term being translated with a descriptive phrase or thesaurus context seen as too foreign or narrow. The interviewed thesaurus specialists were aware of e.g. ISO and SFS standards regarding thesaurus construction, but in their translations they did not motivate their decisions referring to

the literature or guidelines on thesaurus construction. However, the terms typical for thesaurus literature and standards such as non-equivalence and partial equivalence were still constantly used in the discussions.

Overall, the differences among British, Finnish and international thesaurus discourses were rather minor. The Finnish informants were more culturally homogeneous. At the term selection level, harmonising with the General Finnish Thesaurus YSA was aimed for, but in conflict situations the tendency among the Finnish participants was to favour the practices and suggestions received from the specific group of potential users. In thesaurus construction standards and guidelines, vocabulary work was comparable with a clear and logical decision-making process, the result of which is one good (or best possible) solution. Nonetheless, especially the Finnish thesaurus specialists talked about it as more of a compromise-seeking process, where they stressed, that their results could have been somewhat different.

The respondents gave several suggestions of advice to the thesaurus constructors, based on their experience. The need for scope notes and rich co-operation were those that received the most emphasis. The reason for these needs is the acknowledged variety of different discourses and word meanings, which were considered to cause a risk of too heterogeneous indexing and for unwanted search results.

10.3 The thesauri

"--- *does it cover also this and does it truly mean that*." (Finnish thesaurus constructor)

The common problems discussed in the research were in line with those presented by Hudón (1997): stretching the target language to make it fit a foreign conceptual structure where it becomes non-usable to the users of the target language; transfering a whole conceptual structure from one culture to another even when it is not appropriate; translating literally terms from the source language into meaningless expressions in the target language.

As stated previously equivalence was usually not clearly defined in existing thesauri or thesaurus projects referred to by the informants. Thesaurus constructors emphasised the acceptance of uncertainty – the process was not considered to be systematic as presented in thesaurus construction guidelines and standards, but the idea of thesaurus construction as a compromise-making process was clearer than the idea of thesaurus construction as a well-structured and operationalised decision-making process.

More practical conceptual tools are needed in order to be able to analyse and express the varied aims and practices of multilingual thesauri, which should be of help to both the users and researchers of thesauri. In the next subchapters, the types of multilingual thesauri are discussed as well as the pragmatic indexing term equivalence.

10.3.1 Different types of multilingual thesauri

"*I doubt that most users would go to that effort of looking up one*." (British indexer)

"*If you want precise results, a thesaurus is absolutely essential*." (British thesaurus constructor)

The history of multilingual thesauri is one that has included target-language bias, that has not included true multicultural design, and that has reflected incorrectly an idea of cultural neutrality (Hudón 2001, 68-69). This was also the shared common experience of the Finnish thesaurus experts.

In the ISO standard *Documentation - Guidelines for the establishment and development of multilingual thesauri* the concept and term multilingual thesaurus is defined as: "A thesaurus (---) containing terms selected from more than one natural language. It displays not only the interrelationships between terms, but also equivalent terms in each of the language covered" (ISO 5964-1985, 3). This conforms to the way multilingual thesauri are generally understood in practice and in LIS literature. As stated before the informants did not make distinction between **different types of multilingual thesauri**. Based on this research new definitions are suggested. The distinction is considered to be fruitful as it helps in determining what the actual cultural content is – does it reflect one or several or any specific culture. If the type and aim of a thesaurus were stated clearly the user would know better how to orientate and what kinds of concepts and words to use in information seeking and in indexing. The different types of multilingual and/or multicultural thesauri are: monocultural, multicultural and international.

In the study, a **multilingual thesaurus** is thus considered to be a tool for information retrieval and documentation where the indexing terms have equivalents in one or more foreign languages. It is not necessarily multicultural in a sense that it may be bound to the surrounding culture, e.g. solely to the British.

A **monocultural thesaurus** is considered as including several languages, but reflecting one specific culture. E.g. in Finland web-version of the General Finnish Thesaurus (YSA) includes also Swedish language equivalents, but it is culturally only Finnish. It is primarily not meant for cross-cultural searching.

A **multicultural thesaurus** refers to the multicultural content and the using environment of the thesaurus studied – it is not designed with only one cultural group in mind and it includes usually several languages. It may also be monolingual, e.g. English, but still reflect several English-speaking cultures and discourses (such as special lexicology from England, the United States, Canada) or different countries and cultures of Africa which share a common language. It is primarily meant for cross-cultural search. The construction of multicultural thesaurus naturally requires special multicultural co-operation (cf. ISO 5964-1985).

An **international thesaurus** is designed to serve the needs of many cultures without any clear culturally-specific content, and it may be monolingual or multilingual.

In it, the obviously cultural words and concepts are removed during the construction of the corpus.

It is also important to discuss accurately what multilingual indexing is and the implications that the different methods may lead to. Is it indexing in one language and providing equivalents in other languages with the help of a multilingual and/or multicultural thesaurus for cross-lingual and cross-cultural search? Or is it indexing multilingually within one cultural context (e.g. Finnish indexing complemented with e.g. English descriptors in a Finnish database).

10.3.2 Translation strategies

"*--- we talked about it as translating, but it's not translating in its traditional sense, but looking for the correspondences*" (Finnish thesaurus constructor)

"*--- it's something that is Europe wide, although they may mean slightly different things in different countries.*" (British indexer)

The respondents were usually not aiming for literal translation or at least did not consider a literal translation to be useful. Different kinds of strategies existed, and also the aim and focus were rather flexible (cf. Zethsen 2004, 126-127), and the function of the target text was not necessarily the same as the source text in its original context has (cf. Reiss & Vermeer 1986, 54-59; Koskinen 2001, 380). Venuti (1995, 306) has stated that translating should never aim to remove dissimilarities between different cultures entirely, and in the context of multilingual thesauri this is often the aim – but it cannot of course be fulfilled entirely, although it can nevertheless be the sound aim in the construction, in order to serve the needs and practices of the varied user groups. In the following the different thesaurus types previously illustrated are discussed together with translation strategies as well as some of the limitations.

Monocultural thesauri conform to **domestication** as a translation strategy. The descriptors are made recognisable and familiar and the foreign culture is brought closer to the thesaurus user in the target culture (cf. Lindfors 2001, 6). The risk is that the descriptors are extended too much and in the new form no longer serve as usable words (cf. Hudón 1997). If, for example, the Finnish word *sauna* is replaced with a *heating room* the concept may have changed too much, so that it is no longer recognisable to the Finnish users and those familiar with Finnish culture.

In the case of an **international thesaurus** the aim is to be culture-neutral. In the case of *nationalism*, for example, no cultural subterms would be provided as narrower terms (such as *Scottish nationalism*, *Fennomania*). The translation strategy used at the conceptual level – where different concepts are selected and interpreted as descriptors – is **internationalisation** and **consensual equivalence** (cf. Koskinen 2000). In these strategies the relevance of the terms are not necessarily the same, but all the concepts are familiar to all the culturally different user groups. The problem in indexing and in information seeking is, that if all clearly culture-specific expressions are removed, what

happens to the traditional values in LIS as literary and user warrant? Are the linguistic expressions no longer drawn from the literature and do they determine the formulation of descriptors? Does the representation conform to the way users would naturally think or prefer to think? (cf. ISO 5964-1985, Hjørland 2007)

A **multicultural thesaurus** is the most challenging one to construct – how should the variety of needs be truly multiculturally fulfilled? In a multicultural thesaurus the strategies are similar to both **domestication and foreignisation.** The translation is done at a later stage than in monocultural thesaurus construction. The translation corpus includes concepts representing several cultures. The concepts are then interpreted into several languages (or discourses). The primary strategy is domestication – if the concept is familiar in the target culture it receives an equivalent that conforms to the target language and culture. However, if the concept is unfamiliar in the target culture, but important to other cultural groups, it is included by using foreignisation strategy, in which case the user of a thesaurus is taken over to the foreign culture and made to appreciate the cultural and linguistic differences (cf. Lindfors 2001, 6). The foreignisation strategy is complimented with a translation scope note, which clarifies the meaning in the other culture(s). For example, sharia law (in Finnish *šaria-laki*/*sharia-laki*) is not in its origin either a British or a Finnish concept, but it has established linguistic expressions in English and in Finnish. Thus user of the thesaurus would also be able to seek material reflecting only specific (sub)cultures. In extreme cases, different language versions would also include foreign language terms, which as the ISO 5964 equivalence degree remind us, the most unavoidable solution is non-equivalence. However, if the aim is for a truly multicultural content and accepting the potential non-equivalence between different languages and cultures, the attitude towards aimed equivalence degrees should be extended to also cover non-equivalence. It is presumed that the original source language term also works better in indexing and in information seeking, especially in the context of scientific communities. It is assumed that truly multicultural thesauri are only realistic to attain within rather specific contexts (domains).

If a thesaurus is characterised only as a **multilingual thesaurus** without any information about its nature and construction and translation strategies used, it can be of many kinds – and usually not in a coherent way. The problem here is also predictability – if the thesaurus uses several strategies, which also differ radically, how is it possible to find and use the proper descriptors?

The following step is to highlight the different aspects of the concept of the pragmatic indexing term equivalence.

10.3.3 The pragmatic indexing term equivalence

"*One just tends to be sceptical and say it [exact equivalence] is impossible*." (British thesaurus constructor)

"--- *we naively thought that we would be able to have exact matches* --- [laughing] *that came soon disagreed I think*." (British thesaurus constructor)

"--- *in details usability goes ahead*" (Finnish thesaurus constructor)

Attention in the research has been placed on human effort and special attention has been paid to different choices and actions in the definitions of equivalence and the choices of translation strategy. In the empirical part the starting point of the analysis has been translatability issues from British-English into Finnish and also the more theoretical levels (ideals, principles and practices) have been studied. Koller's division (1989, 1995, see also Zethsen 2004) of different types has been understood as kind of layers and perspectives for the analysis of the empirical material. The division led to further questions regarding the pragmatic indexing term equivalence from five different perspectives. These perspectives are considered useful in the new context of thesaurus construction:

1. Denotative

- Have the two indexing terms the same truth value? Do they refer to a same concept at the denotative level?

2. Connotative

- How does the connotative level have an impact? Are the two indexing terms equally positive or neutral? Can the connotative level be superior to the denotative level?

3. Text-normative

- Thesauri represent documentary language. How does the text-type have an affect? Do the indexing terms represented as equivalent represent similarly both documentary languages?

4. Pragmatic

- Is the translation accepted and usable in a target context (in the way it was in the source context)? Is the effect the same? - Does the translation produce similar results e.g. when information seeking as in the source version?

5. Formal

- Are the hierarchies the same in the source and target languages? - How is the hierarchical thesaurus system considered?

It was obvious, that translatability is, on the one hand, strongly connected to relevance and on the other hand to equivalence. In a thesaurus context there are three

questions to consider: first, one has to consider the relevance of the concept in a vocabulary, secondly the aim of translation has to be considered and thirdly the equivalence is important. In other words, the questions to be asked are: 1) to translate or not (corpus based relevance issue) -> if to translate 2) what to aim for -> (translation strategy issue) and ->3) how to achieve it (pragmatic equivalence type and equivalence degree issue).

Koller's division (here interpreted in a thesaurus context) can be used as a basis for evaluating different aspects of equivalence by providing different perspectives to consider. It helps to express the concrete aims and also to make the necassary translation scope notes, in which information is provided about how the specific translation differs from the other language equivalents (e.g. "*social security* has a positive connotation in culture X, whereas in culture Y it is generally understood as negative and therefore it may not be used in similar meaning and context").

The principal aim was generally pragmatic equivalence – the indexing terms represented in different languages should result in similar material in the search, despite the language used. To achieve this means that in some situations thesaurus constructors should stop aiming for a translation in its traditional sense and endeavour to transfer the denotative and central meaning from one language to another (cf. Nida 1975, 228), but also to aim at correspondence in an information seeking situation. The emphasis is then clearly on the documentary language type, which does not necessarily conform to the natural language. It was also natural that the entire aim cannot always be fulfilled –if not operating at a very general level (extreme cases of international thesauri types, which would then miss the specificity often needed in successful indexing and information seeking).

As stated previously, the perspectives found were conceptual, dynamic seeking, collection and consensual or existential equivalence. Conceptual refers to same truth value at the denotative level and at the connotative level (cg. Koller's terminology), and it thus fulfils the criterion of dynamic seeking and collection equivalence. Dynamic seeking refers to obtaining similar search results when using indexing terms in different language and presented as equivalent, and in this case a phenomenon that was clearly British should result in similar material also in a Finnish context, but which would then have different significance in the Finnish context. Collection equivalence refers to database-based meaning and is the most organisation-specific and thus probably reflects the perspective of domestic more and also more that of an indexer than an information seeker. Existential equivalence means that the status of the indexing term in the target language has not been considered to be high and often the source terms are not possible to translate as indexing terms, but another kind of translation (such as a descriptive phrase or citation loan, which often do not work when seeking or indexing target language material) is included in order to provide an explanation of the source term - instead of aiming at a functional equivalence type. This last mentioned type is very much related to user equivalence – e.g. in a multicultural thesaurus some indexing terms are inevitably relevant to only one section of the users.

The different aspects and perspectives required are more probably involved if thesaurus projects include different kinds of user groups. Both the information seeker-

centered and material-based indexing perspective is necessary in order to construct a thesaurus that is functional for both indexing and information seeking. As also discussed earlier in various contexts, the achievement of connotative equivalence is one of the most difficult problems in translation, and can seldom be absolute (see Koller 1989). Thesaurus constructors, and therefore, also thesaurus translators aim to achieve neutral, descriptive and precise equivalents, but the users cannot be expected to work against the norms of common language usage, and thus thesaurus terminology should also carefully consider the norms of non-discriminatory language (cf. especially Ulrych 1992; Fairclough 1994). As stated previously, if thesauri are also to reflect social science discourse the connotative aspects should be considered more carefully than the common guidelines indicate.

As with aboutness (cf. Ingwersen & Järvelin 2005, 381) equivalence is not an inherent feature of indexing terms represented as "equivalent", but dependent on the person who determines the aims during the thesaurus construction and those seeking information in a time-space continuum. It transpired that due to the contextual nature when aiming for equivalence it is not possible to clearly define the pragmatic indexing term equivalence, but it would help both the thesaurus constructor and different users as indexers or information seekers to know what kind of ideology lies behind the thesaurus. As pointed out, translating is not a value-free action and it always involves the necessity of choice (see especially Koller 1989; Venuti 1995; Ruokonen 2004) and therefore a translator should first set up a hierarchy of values to be preserved in the translation and from this to derive a hierarchy of equivalence requirements for the thesaurus or indexing terms in question (cf. Koller 1989). It would also be useful to provide this information for the different users in order to help them to select a suitable indexing or information seeking strategy. An increase in the complexity of the using environment intensifies the amount of work that needs to be done. In human communication, vocabulary tools are still only tools, and cannot provide a complete interpretation.

10.4 The future

Whatever is said about the limitations of translating, it is and it will be one of the most important and valuable actions in the world. (Johann Wolfgang von Goethe, 1749-1832, cited here Saksa 2004, 9)

First, it would be useful to concentrate on further operationalising the idea of different translation strategies and the concept of the pragmatic indexing term equivalence by e.g. using different kinds of material (terminology, languages, and/or cultures). Certainly more research would increase understanding of the phenomenon, and as the field is so little studied there is a need for proper tools for both practical and theoretical evaluation. The research should also not only be theoretical, but also empirical in its nature. This means that in order to develop the concept of pragmatic indexing term equivalence in various contexts it would be reasonable to test the theoretical definitions in several specific kinds of multilingual thesaurus projects, and then re-formulate the definitions. Operating within a framework of multicultural

research would naturally be of benefit. -> **Translation strategies and pragmatic indexing term equivalences in different thesaurus contexts.**

Second, it would be necessary to also study multilingual and/or multicultural indexing practices. One interesting platform could be the international Cambridge Sociological Abstracts database (CSA), in which descriptors are provided by American indexers with the help of continuously updated and informative thesauri and (possible) identifiers by the authors of the articles. (Preliminary studies made showed that the quality and possibilities of the system are of a higher quality than that found, for example, in the Finnish and British databases used in the research.) It would also be an interesting alternative to gather a pool of documents and see how they are indexed in different databases (cultural contexts). A British-driven alternative for an international platform would be the International Bibliography of the Social Sciences (IBSS), which also uses a thesaurus.

-> **Practices of multilingual and/or multicultural indexing**.

Third, it would be important to study matters invisible in indexing – looking at what kind of aspects are left without descriptors and how they are handled in information seeking situations. In this context, it can be questioned whether male gender, domesticity and/or a recent time aspect is considered to be a norm in the indexing – and how are they seen later in information seeking results or whether they become "invisible". A concrete question for an information retrieval situation is, how to aim for preciseness in searches if the information needed is e.g. about men's roles today in families.

-> **Aspects which are invisible versus transparent in documentary languages and indexing**.

Fourth, it would be valuable to also study genderism in documentary language discourse, since e.g. English thesauri often have descriptors such as *married women workers* without a male equivalent. The questions to be asked might include: How does the genderism occur? What are the topics and is the trend similar in different contexts? Does the indexing discourse conform to the thesaurus discourse and/or to the common language discourse?

-> **Genderism in documentary language discourse**.

Fifth, it would be valuable to concentrate on the methodological and analytical tools used in the evaluation of the results of multilingual and multicultural thesaurus construction.

-> **Tools and methods useful in thesaurus construction evaluation**.

'Tell me about the snow,' Moomintroll said and seated himself in Moominpappa's sun-bleached garden chair. 'I don't understand it.'

' I don't either,' said Too-ticky. 'You believe it's cold, but if you build yourself a snowhouse it's warm. You think it's white, but at times it looks pink, and another time it's blue. It can be softer than anything, and then again harder than stone. Nothing is certain.' (Jansson 1971, 28)

REFERENCES

Adler 1997 = Adler, Nancy 1997 (1991): *International Dimensions of Organizational Behaviour*. 3rd ed. South-Western Canada: College Publishing. 331 pp.

Ahlgren & Kekäläinen 2007 = Ahlgren, Per & Kekäläinen, Jaana (2007): "Indexing strategies for Swedish full text retrieval under different user scenarios". *Information Processing & Management* 43(1): 81-102. (Retrieved at ScienceDirect 29.9.2008.)

Airio 2009 = Airio, Eija 2009: *Morphological Problems in IR and CLIR. Applying linguistic methods and approximate string matching tools*. Tampere University Press. Acta Electronica Universitatis Tamperensis; 842. 97 pp.

Aitchison 1991 = Aitchison, Jean 1991 (1984*): Language Change: Progress or Decay?* 2nd ed. Suffolk: Fontana Paperbacks. 258 pp.

Aitchison et. al. 1997 = Aitchison, Jean & Gilchrist Alan & Bawden David1997 (1972): *Thesaurus construction and use: a paractical manual*. 3rd ed. London: Aslib. 212 pp.

Alasuutari 1994 = Alasuutari Pertti 1994 (1993): *Laadullinen tutkimus*. 3rd ed. Tampere: Vastapaino. 281 pp.

Alaterä & Halttunen 2002 = Alaterä, Anu & Halttunen, Kai 2002: *Tiedonhaun perusteet - osa lukutaitoa*. Btj Kirjastopalvelu,

Allardt & Starck 1981 = Allardt, S. & Starck, C. 1981: *Vähemmistö, kieli ja yhteiskunta*. Juva: WSOY.

Anderson 1997 = Anderson, J.D 1997: "Organization of knowledge". IN: J. Feather & P. Sturges (Eds.). *International Encyclopedia of Information and Library Science*, p 336-353. London: Routledge.

Bassnett 1995 = Bassnet, Susan 1995: *Teoksesta toiseen. Johdatus kirjallisuuden kääntämiseen*. Ed. Riitta Oittinen. Jyväskylä: Vastapaino. [Original work: Translation Studies 1980 (1991), Routledge.]

Bates 1986 = Bates, Marcia J. 1986: "Subject Access in Online Catalogs: A Design Model." *Journal of the American Society for Information Science* 37:357-376 (November 1986).

Bibexcel 2001 = *Bibexcel - a toolbox for bibliometricians*. Version 2001-05-21. Developed by Olle Persson, Inforsk, Umeå University, Sweden. Uploaded from the World Wide Web <URL: http://www.umu.se/inforsk/Bibexcel/index.html>

BIS 2009 = Department for Business Innovation & Skills (BIS) 2009: Additional Paternity Leave and Pay. Retrieved in 18th September 2009 from the World Wide Web <URL: http://www.berr.gov.uk/whatwedo/employment/workandfamilies/add-paternity-leave/index.html>

Blair 2003 = Blair, D. C. 2003: "Information retrieval and the philosophy of language". *Annual Review of Information Science and Technology*, 37 (1): 3-50.

Borlund 2000 = Borlund, Pia 2000: *Evaluation of interactive information retrieval systems*. Doctoral dissertation, 276 pp. Åbo: Åbo Akademi University Press.

Brake & al. 1995 = Brake, Terence, Walker, Danielle & Walker, Thomas 1995: *Doing Business Internationally* : *The guide to cross-cultural success*. Burr Ridge, IL: Irwin Professional Publishing. 225 pp.

Brannen 1992 = Brannen, Julia 1992: "Combining qualitative and quantitative approaches: An overview." IN: Brannen, Julia (ed.) *Mixing methods: Qualitative and quantitative research*, pp. 3-38. Aldershot: Avebury.

Brunel University 1999 = Human Resources, Brunel University 1999*: Non-Discriminatory Language Guide. A guide for students and staff.* Retrieved in 2002 from the World Wide Web <URL: http://www.brunel.ac.uk/admin/humanresources/equalopp/langguid.html>

Buckland 1999 = Buckland, Michael 1999: "Vocabulary As A Central Concept In Library And Information Science." Preprint of paper published as "Vocabulary as a Central Concept in Library and Information Science" IN: *Digital Libraries: Interdisciplinary Concepts, Challenges, and Opportunities. Proceedings of the Third International Conference on Conceptions of Library and Information Science* (CoLIS3, Dubrovnik, Croatia, 23-26 May 1999. Ed. by T. Arpanac et al. Zagreb: Lokve, p. 3-12.

Busha & Harter 1980 = Busha Charles H. & Harter Stephen P. 1980: *Research Methods in Librarianship. Techniques and Interpretation*. Orlando, Florida: Academic Press. 417 pp.

Callon & al.1991 = Callon, M., Courtial, J.P., Laville, F. 1991: "Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry." *Scientometrics*, Vol.22. No.1 (1991) pp.155-205.

Cambridge International Dictionary of English 2001 & 2002 = *Cambridge International Dictionary of English* (9.4.2001, 3.9.2002). Cambridge Dictionaries Online. Cambridge University Press. Consulted in the World Wide Web <URL: http://dictionary.cambridge.org/>

Carmen 2004 = Carmen Guarddon Anelo 2004: "Philosophy, Anthropology, and Linguistics in Translation." *Translation Journal*, Volume 8, No. 1, January 2004. Retrieved in 2004 from the World Wide Web <URL: http://accurapid.com/journal/27philosophy.htm>

Catford 1965 = Catford, J.C. 1965: *A Linguistic Theory of Translation , an Essay in Applied Linguistics*. London: Oxford University Press. 103 pp.

Center for Psychological Studies 2005 = Nova Southeastern University - Center for Psychological Studies, "Word Association Test." Retrieved in January 17th 2005 from the World Wide Web <URL: http://cps.nova.edu/~cpphelp/WAT.html>

CD-Perussanakirja 1997 = *CD-Perussanakirja*. Kotimaisten kielten tutkimuskeskuksen julkaisuja 94. Kotimaisten kielten tutkimuskeskus.

Chan et al. 2000 = Chan, Lois Mai & Lin, Xia & Zeng, Marcia Lei 2000:" Structural and Multilingual Approaches to Subject Access on the Web." *IFLA Journal*, Vol. 26(2000)3.

Chang & Lee 2001 = Chang, S.-J. L. & Lee Y.-Y. 2001: "Conceptualizing context and its relationship to the information behaviour in dissertation research process". *The New review of Information Behaviour Research*, vol. 2, 2001: 79-94.

Chesterman & Arrojo 2000 = Chesterman, Andrew & Arrojo, Rosemary 2000: "Shared Ground in Translation Studies." *Target* 12:1, 151-160.

Chesterman 2005 = Chesterman, Andrew 2005: *Englanti globaalina ja eurooppalaisena lingua francana*. IN: Johansson & Pyykkö 2005, pp. 115-128.

Council for Equality 1999 = Council for Equality 1/1999: *A Millennium of Equality*. 16pp. Retrieved in March 4th 2002 from the World Wide Web: <http://www.tasa-arvo.fi/wwweng/publications/publications1.html>

Council for Equality (2002) = "Gender equality in Finland". 8pp. Retrieved in March 4th 2002 from the World Wide Web: <http://www.tasa-arvo.fi/www-eng/publications/publications2.html>

Cousins & Tang 2003 = Cousins, C. and Tang, N. 2003: ' *Working Time and Family Life in the UK*.' Business School Working Papers UHBS 2003-3, Employment Studies Paper 47. 25 pp.

Creswell 1994 = Creswell, J. W. 1994: *Research design :Qualitative & quantitative approaches*. Thousand Oaks: Sage.

Crystal 1997 = Crystal, David 1997: *English as a Global Language*. 2nd ed. Cambridge University Press, Cambridge.

CSA 2001 = Cambridge Scientific Abstracts (2.5.2001): "Thesaurus of Sociological Indexing Terms. Introduction." Retrieved in May 2nd 2001 from the World Wide Web URL: <http://www.csa.com/edit/sociothes.html>

CSA 2002a = Cambridge Scientific Abstracts (15.1.2002): "Keyword Search." Retrieved in January 15th 2002 from the World Wide Web URL: <http://www.csa2.com/csa/HelpV5/keyword.shtml> [no update information available]

CSA 2002b = Cambridge Scientific Abstracts (15.1.2002): "Field Codes and Search Examples." Retrieved in January 15th 2002 from the World Wide Web URL: <http://www.csa2.com/csa/HelpV5/fielded.shtml> [no update information available]

Denzin 1978 = Denzin, Norman K. 1978: *The research act: A theoretical introduction to sociological methods*. 2nd edition. New York: McGraw-Hill.

Dictionary.com = Lexico LLC: Dictionary.com. Consulted 2001-2009 in the World Wide Web <http://www.dictionary.com/>

Ding et al. 2000 = Ding, Ying & Chowdhury, Gobinda & Foo, Schubert 2000: "Organising Keywords in a Web Search Environment: A Methodology based on Co-Word Analysis." *Proc. 6th International Society for Knowledge Organization (ISKO 6) Conference*, Toronto, Canada, 28-34. Retrieved in from the World Wide Web March 30th 2001 URL: <http://islab.sas.ntu.edu.sg:8000/user/schubert/publications/2000/00isko.htm>

Doerr 2001 = Doerr, Martin. (2001): "Semantic Problems of Thesaurus Mapping." *Journal of Digital information*, vol. 1 issue 8. Retrieved in April 23rd 2001 from the World Wide Web, URL: <http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Doerr/>

EC 2001 = European Communities 2001: *Presentation of the thesaurus*. Retrieved in April 5th 2001 from the World Wide Web URL: <http://europa.eu.int/celex/eurovoc/index.htm>

EC 2009 = European Communities 2009: *Presentation of the thesaurus*. Retrieved in January 29th 2009 from the World Wide Web URL: <http://europa.eu/eurovoc/sg/sga_doc/eurovoc_dif!SERVEUR/menu!prod!MENU?langue=EN>

ECE 1/2000 = Economic Commission for Europe. Annual report 6/May 2000 - 11/May 2001. Economic and Social Council. Official Records. *Economic Survey of Europe 1/2000*. Available in the World wide Web (last visited 24.9.2009) URL: < http://www.unece.org/commission/2001/e_ece_1387e.pdf>

Eco 1995 = Eco, Umberto 1995 (1993): *The search of the perfect language*. Translated by James Frentress. Oxford: Blackwell.

Engelberg 1993 = Engelberg, Mila (1993): "Maskuliini, feminiini, virkamies?" *Naistutkimus* 4/93, pp. 39-50.

Engelberg 1998 = Engelberg, Mila (1998): ”Sukupuolistuneet ammattinimikkeet.” *Virittäjä* 1/1998, pp. 74-92.

EOC 2002 = Equal Opportunities Commission (EOC) 20.10.2002: “News release: Dads still see breadwinner as their main role in the family.” Retrieved in October 30th 2002 from the World Wide Web URL: <http://www.eoc.org.uk/cseng/news/20_oct_fathers.asp>

ERIC 1999 = Educational Resources Information Center 1999: *About ERIC and the ERIC Thesaurus*. Retrieved in April 11th 2001 from the World Wide Web URL: <http://ericae.net/scripts/ewiz/aewiza.htm#about>

Eskola & Suoranta 1988 = Eskola, Jari & Suoranta, Juha 1988: *Johdatus laadulliseen tutkimukseen*. Jyväskylä: Vastapaino.

Eskola & Suoranta 1998 = Eskola J. & Suoranta, J. 1998: ”Laadullisen tutkimuksen arviointi.” IN: Eskola, J. & Suoranta J.: *Johdatus laadulliseen tutkimukseen*, pp. 209-231. Tampere: Vastapaino.

Fairclough 1992 = Fairclough, Norman 1992: *Discourse and Social Change*. Cambridge, England: Polity Press. 259 p.

Fiske 2000 = Fiske, John 2000 (1992): *Merkkien kieli : Johdatus viestinnän tutkimiseen*. 6.ed. Tampere: Vastapaino. [English original: Introduction to Communication Studies. 2nd ed. Routledge, 1990.]

Forsman & Keränen 2002 = Forsman, Maria & Keränen, Susanna. “From HASSET to MADIERA : FSD develops multilingual content description.” [electronic resource] *FSDneWWWs* 2002:3. URL: < http://www.fsd.uta.fi/tietoarkistolehti/english/09/madiera.html> [Consulted February 27th 2005]

Forsman 2005 = Forsman, Maria 2005: *Development of research networks: The case of social capital*. Doctoral dissertation. Åbo: Åbo Akademi University Press. 237 pp.

Forssén & Hakovirta 2001 = Forssén, Katja & Hakovirta Mia 2001: “Family Policy, Work Incentives and Employment of Mothers.” IN: Sigg, R. (ed): *Social Security in the Global Village*. New Bunswick.

Fugmann 1994 = Fugmann, Robert 1994: “Representational Predictability: Key to the Resolution of Several Pending Issues in Indexing and Information Supply.” In *Knowledge organization and quality management proceedings of the Third International ISKO Conference*, 20-24 June 1994, Copenhagen, Denmark. Eds. Hanne Albrechtsen & Oernager, Susanna. INDEKS Verlag, Frankfurt/Main, Germany.

Gephart 1999 = Gephart, R. 1999: “Paradigms and research methods”. *Research Methods Forum*, 4. Retrieved in September 20th 2005 URL:

<http://www.aom.pace.edu/rmd/1999_RMD_Forum_Paradigms_and_Research_Methods.htm>

Gilchrist 2003 = Gilchrist 2003: “Thesauri, taxonomies and ontologies – an etymological note.” *Journal of Documentation*. Vol. 59 No.1, 2003.

Google Help 2002 = Google 2002: *Google Help*. Retrieved in September 20th 2002 from the World Wide Web URL: <http://www.google.com/help/refinesearch.html#domain>

Gordon 2005 = Gordon, Raymond G., Jr. (ed.), 2005: *Ethnologue: Languages of the World*, Fifteenth edition. Dallas, Tex.: SIL International. Retrieved in November 15th 2006] URL: <http://www.ethnologue.com/>.

Grefenstette & Nioche 2000 = Grefenstette, G., & Nioche, J. (2000): "Estimation of English and non-English Language Use on the WWW." *RIAO'2000* (Recherche d'Informations Assistee par Ordinateur), Paris, April 12-14, 2000. Retrieved in February 1st 2002 from the World Wide Web, URL: <http://www.xrce.xerox.com/research/mltt/publications/Documents/P19137/content/RIAO2000gref.pdf>

Haarala 1981 = Haarala, Risto (1981). *Sanastotyön opas* [Guide to terminology work]. Kotimaisten kielten tutkimuskeskuksen julkaisuja 16. Helsinki: Valtion painatuskeskus.

Haataja & Nyberg 2006: Haataja, Anita & Nyberg, Anita 2006: "Diverging paths? The dual earner - dual carer model in Finland and Sweden in the 1990s." IN Ellingsæter, Anne Lise & Arnlaug, Leira (eds.) *Politicing Parenthood in Scandinavia*. *Gender relations in welfare states*. Policy Press, Bristol. P. 217-239.

Hall 1981 = Hall, Edward T. (1981, org. 1959): *The Silent Language*. New York: Doubleday. 192 pp.

Hall 1982 = Hall, Edward T. (1981, org. 1966): *The Hidden Dimension*. New York: Doubleday. 217 pp.

Halverson 1998 = Halverson, Sandra 1998: "Translation studies and representative corpora: establishing links between translation corpora, theoretical/descriptive categories and a conception of the object of study."* (*Forthcoming in *Investigating Translational Behavior and Other language Phenomena with the tools of Corpus Linguistics,* Special edition of *Meta*. Guest editor S. Laviosa.) In Halverson Sandra, *Concepts and Categories in Ttranslation Studies*. Department of English, university of Bergen, Norway. Doctoral Dissertation.

Hedemark & Sundin 2005 = Hedemark, Å., Hedman, J. & Sundin, O. 2005: "Speaking of users: on user discourses in the field of public libraries". Information Research, 10(2) paper 218, 19 pp. Retrieved in October 31st 2005 from the World Wide Web URL: <http://InformationR.net/ir/10-2/paper218.html>

Helsingin yliopiston kirjasto 1999 = Helsingin yliopiston kirjasto 1999: "VESA – verkkosanasto. Yleisesittely eli mikä on yleinen suomalainen asiasanasto (YSA)". Retrieved in May 2nd 2001 from the World Wide Web URL: <http://vesa.lib.helsinki.fi/ysa/ysayleis.html> 5 pp.

Helsinki University Library 2005 = Helsinki Universtiy Library 2005 (last updated) 15.2.05): "LINNEA database services – LINDA". Retrieved in 2005 from the World Wide Web, URL: < http://www.nationallibrary.fi/libraries/linnea/linda.html>

Hert & Jacob & Dawson 2000 = Hert, Carol A. & Jacob, Elin K. & Dawson, Patrick. A usability assessment of online indexing structures in the networked environment. *Journal of the American Society for Information Science*, 2000, 51 (1): 971-988.

Hiidenmaa 2000 = Hiidenma a, Pirjo 2000: "Suomen kielen eloonjäämisoppi". pp. 45-56. In *STUDIA GENERALIA*. *Suomi EU:n puheenjohtajana, suomi Euroopan kielenä*. Heino, Päivi & Westermarck, Harri (eds.). Helsinki: Helsingin yliopiston vapaan sivistystyön toimikunta, Yliopistopaino. 123 pp.

Hirsjärvi & Hurme 2000 = Hirsjärvi, Sirkka & Hurme, Helena 2000: *Tutkimushaastattelu*. *Teemahaastattelun teoria ja käytäntö*. Helsinki: Helsinki University Press.

Hjørland, 1997 = Hjørland, Birger. (1997): Information Seeking and Subject Representation. An Activity-Theoretical Approach to Information Science. London: Greenwood Press. 213 p.

----- 2002 = Hjørland, Birger. (2002): "Domain analysis in information science. Eleven approaches - traditional as well as innovative." *Journal of Documentation*, 58(4), 422-462.

----- 2004 = Hjørland, Birger. (2004): "Domain analysis in information science." IN: *Encyclopedia of Library and Information Science*. New York: Marcel Dekker. Pp. 1-7.

----- 2005 = Hjørland, Birger (2005): "Core Concepts in Library and Information Science (LIS)". (Last edited: 27-07-2006) Retrieved in September 15th 2008 from the World Wide Web URL: <http://www.db.dk/bh/core%20concepts%20in%20lis/home.htm>

Hjørland & Nicolaisen 2005 = Hjørland, Birger & Nicolaisen, Jeppe (eds.) 2005: *The Epistemological Lifeboat. Epistemology and Philosophy of Science for Information Scientists*. Retrieved in January 23rd 2007 (last updated 22.1.2007) from the World Wide Web URL: <http://www.db.dk/jni/lifeboat/home.htm>

Hjørland & Pedersen 2005 = Hjørland, Birger & Pedersen, Karsten Nissen 2005: "A substantive theory of classification for information retrieval". *Journal of Documentation*, Vol. 61 No.5, 2005, pp. 582-597. Emerald Group Publishing Limited.

Hjørland 2007 = Hjørland, Birger 2007 (last edited 20.8.2009): "Literary warrant (and other kinds of warrant)". In Core *Concepts in*

Library and Information Science (LIS). Retrieved in September 15th 2009 from the World Wide Web URL: <http://www.db.dk/bh/Lifeboat_KO/CONCEPTS/literary_warrant.htm>

Hjørland 2008 = Hjørland, Birger 2008: "What is Knowledge Organization (KO)?" *Knowl.Org*. 35 (2008) No. 2/No. 3. p. 86-101.

Hooper 1997 = Hooper, Tracy 1997 (2002): "Planet SOSIG: The new SOSIG Interface." *Ariadne*, Web version, Issue 10, July 1997. Retrieved in July 8th 2002 from the World Wide Web URL: <http://www.ariadne.ac.uk/issue10/planet-sosig/>

Horton et al. 1998 = Horton, Thomas & Coulter, Neal & Grant, Emanuel 1998: "Applying Content Analysis to Humanities Computing Research Literature." Joint International Conference ALLC/ACH'98, Association for Literary and Linguistic Computing, Association for Computers and the Humanities, 5 – 10 July 1998, Lajos Kossuth University Debrecen, Hungary. Retrieved in April 23rd 2001 from the World Wide Web URL: <http://lingua.arts.klte.hu/allcach98/abst/abs20.htm>

HS 2004 = Helsingin Sanomat. NYT-liite 8/2004.

Hudón 1997 = Hudón, Michele 1997: "Multilingual thesaurus construction: integrating the views of different cultures in one gateway to knowledge and concepts." *Information Services & Use*, Vol. 17 Issue 2/3, 1997, p. 111, 13 pp.

Hudón 2001 = Hudón, Michele 2001: "Relationships in Multilingual Thesauri." IN: Bean Carol, & Green Rebecca (eds.) *Relationships in the Organization of Knowledge*. pp. 67-80 Springer, 2001.

Hutchins 1975 = Hutchins, W.J. 1975: Languages of indexing and classification: A linguistic study of structures and functions. London: Peter Peregrinus.

Huttunen 2001 = Huttunen, Jouko 2001: *Isänä olemisen uudet suunnat*. Jyväskylä: PS-kustannus.

Hypertext Webster Gateway. Consulted in the World Wide Web 2001-2009 URL: <http://smac.ucsd.edu/cgi-bin/http_webster?>

Häkkinen 1989 = Häkkinen, Kaisa 1982 (1989): *Yleisen kielitieteen peruskurssi*. 10th ed. Turun yliopiston suomalaisen ja yleisen kielititeen laitoksen julkaisuja. Turku: Turun yliopisto.

Häkkinen 1994 = Häkkinen, Kaisa 1994: *Kielitieteen perusteet*. Tampere: Suomalaisen kirjallisuuden seura. 188pp.

Iannella 1998 = Iannella, Renato 1998: "Mostly Metadata: A Bit Smarter Technology." *VALA Conference*, Jan. 1998. The Resource Discovery Unit (RDU), The University of Queensland, Australia. Retrieved in December 21st 2000 from the World Wide Web URL: <http://www.dstc.edu.au/RDU/publications/m_vala98/>

ICPSR 2009a = The Interuniversity Consortium for Political and Social Research (ICPSR) 2007: *About ICPSR*. Last visited in September 29th 2009 in the World Wide Web, URL: <http://www.icpsr.umich.edu/ICPSR/org/index.html>

ICPSR 2009b = The Interuniversity Consortium for Political and Social Research (ICPSR) 2009: "Thesaurus". Last visited in September 29th 2009 in the World Wide Web, URL: <http://www.icpsr.umich.edu/ICPSR/thesaurus/index.html>

Iivonen 1989 = Iivonen, Mirja 1989: *Indeksointituloksen riippuvuus indeksointiympäristöstä*. University of Tampere, Department of Library and Information Science, Studies No. 26/1989.

---------- 1995 = Iivonen, Mirja 1995: *Hakulausekkeiden muotoilun yhdenmukaisuus onlineviitehaussa* (Consistency in the formulation of query statements in online bibliographic retrieval). Tampereen yliopisto, Acta Universitatis Tamperensis. Ser. A, vol. 443, Tampere. [Includes summary in English.]

Iivonen & Kivimäki 1998 = Iivonen, Mirja & Kivimäki, Katja 1998: "Common Entities and Missing Properties: Similarities and Differences in the Indexing of Concepts." *Knowl. Org.* 25(1998)No.3, pp.90-102

Independent 10.6.2002 = Independent Digital (UK) Ltd 10 June 2002: "Unemployed single mothers to be trained as gas fitters". By Marie Woolf, Chief Political Correspondent. Retrieved in October 30th 2002 from the World Wide Web URL: <http://www.independent.co.uk.>

Independent 8.7.2002 = Independent Digital (UK) Ltd 8 July 2002: "The best man for the job". By Nick Druerden. Retrieved in October 30th 2002 from the World Wide Web URL: <http://www.independent.co.uk>.

Independent 20.10.2002 = Independent Digital (UK) Ltd 20 October 2002: "What kind of dad are you? Forceful? Useful? Or just plain embarrassing?" By James Morrison. Retrieved from the World Wide Web October 30th 2002 URL: <http://www.independent.co.uk>.

Ingo 1982 = Ingo, Rune 1982 (1980): *Kääntämisen teoriaa ja sen sovellusta*. 3rd ed. Åbo: Åbo Akademi.

----- (1990): Ingo, Rune 1990: *Lähtökielestä kohdekieleen. Johdatusta käännöstieteeseen*. Porvoo-Helsinki-Juva: WSOY. 324 pp.

Ingwersen 1994 = Ingwersen Peter 1994: "Polyrepresentation for information needs and semantic entities: elements of a cognitive theory for information retrieval interaction." IN: Croft, W. B. & van Rijsbergen, C. J. (eds.). *SIGIR 94: Proceedings of the seventeenth annual international ACM-SIGIR conference on research and development in information retrieval*, organised by Dublin City University, 3-6 July 1994. Dublin, Ireland. Lindon: Springer-Verlag, p. 101-110.

----- 2002 = Ingwersen Peter 2002: "Cognitive perspectives of document representation." IN: Emerging Frameworks and Methods : Proceedings of the Fourth International Conference on Conceptions of Library and Information Science (CoLIS4) .- Bruce, Harry et al. [ed]. Greenwood Village, Colorado : Libraries Unlimited , 2002, p. 285-300.

Ingwersen & Järvelin 2005 = Ingwersen, P. & Järvelin, K. (2005). *The Turn: Integration of Information Seeking and Retrieval in Context*. Heidelberg: Springer.

INRA 2001 = International Research Associates INRA (Europe), European Coordination Office S.A. 2001: Eurobarometer 54 Special "Europeans and Languages". Report produced by INRA for The Education and Culture Directore-General. Managed and organized by The Education And Culture Directorate-General Unit <<Centre for the citizen – Analysis for public opinion>>. February 2001. (Originally published in French.) Retrived in November 29th 2002 from the World Wide Web URL:
<http://europa.eu.int/comm/education/languages/lang/barolang_en.pdf>

International Encyclopedia of Information and Library Science 1997. Edited by J. Feather & P. Sturges. London: Routledge.

ISI Citation Databases Help 2000 = Institute for Scientific Information 2000: "ISI Citation Databases Help. Documentation version 4.3", last modified 10/6/2000. Retrieved in August 9th 2001 from the World Wide Web URL: <http://webofscience.com/help/helpdefs.html> Moved into new address, ISI Web of Knowledge, URL: < www.isiknowledge.com>, last visited September 28th 2009.

ISO 5963-1985 = International Organization for Standardization 1985: *Documentation – Methods for examining documents, determining their subjects, and selecting indexing terms*. Geneve: International Organization for Standardization.

ISO 2788-1986 = International Organization for Standardization 1986: ISO International Standard 2788. *Documentation - Guidelines for the establishment and development of monolingual thesauri*. Geneve: International Organization for Standardization.

ISO 5964-1985 = International Standards Organization (ISO). *Guidelines for the Establishment and Development of Multilingual Thesauri*. 1st ed. Geneva: ISO, 1985 (ISO 5964-1985-02-15).]

Jallinoja 2006 = Jallinoja, Riitta 2006: *Perheen vastaisku*. Helsinki: Gaudeamus.

James 2002 = James, Kate 2002: "Cultural Implications for Translation". In *Translation Journal* Volume 6, No. 4 October 2002. Retrieved in 2005 from the World Wide Web <URL: http://accurapid.com/journal/22delight.htm>

Janesick 2000 = Janesick, Valeri J. 2000: "The Choreography of Qualitative Research Design: Minuets, Improvisations, and Crystallization." Pp. 379-399. IN: Denzin, Norman K. & Lincoln, Yvonna S. (eds.) *Handbook of qualitative research*. 2.ed. Thousand Oaks : Sage. Pp. 1065.

Jansson 1971 = Jansson 1971 (1958): *Moominland Midwinter*. (Translated by Thomas Warburton.) England: Puffin Books.

Johansson & Pyykkö 2005 = Johansson, Marjut & Pyykkö, Riitta (eds.) 2005: *Monikielinen Eurooppa : Kielipolitiikkaa ja käytäntöä*. Helsinki: Gaudeamus. Pp. 374.

Johansson & Wiberg 2005 = Johansson, Marjut & Wiberg, Matti 2005: "Institutionaalinen monikielisyys Euroopan uniossa: monesta kohti yhtä?" Pp. 29-47. IN: Johansson, Marjut &

Pyykkö, Riitta (eds.) 2005: *Monikielinen Eurooppa : Kielipolitiikkaa ja käytäntöä*. Helsinki: Gaudeamus. 374 pp.

Jokinen & al. 1993 = Jokinen, Arja, Juhila, Kirsi & Suoninen, Eero 1993: *Diskurssianalyysin aakkoset*. Tampere: Vastapaino. 238 pp.

Jorna & Davies 2001 = Jorna, K., & Davies, S. 2001): "Multilingual Thesauri For The Modern World - No Ideal Solution?" *Journal of documentation* 2001 57 (2) pp.284-295.

Juvakka & Viljamaa 2002 = Juvakka, Essi & Viljamaa, Janne 2002: *Miehen mittainen isä*. Helsinki: WSOY. 154 pp.

Järvelin 1995 = Järvelin, Kalervo 1995: Tekstitiedonhaku tietokannoista. Johdatus periaatteisiin ja menetelmiin. Espoo: Suomen ATK-kustannus.

Jääskeläinen & Forsman 2003 = Jääskeläinen, Taina & Forsman, Maria; "Finnish Version of the Multilingual Thesaurus ELSST" [electronic resource] *FSDneWWWs* 12 2003:3. Retreieved in from the World Wide Web February 27th 2005 URL: < http://www.fsd.uta.fi/tietoarkistolehti/english/12/elsst.html>

Jääskeläinen 2006 = Jääskeläinen, Taina 2006: "ELSST - monikielisen tesauruksen kääntämisen haasteet." *Tietoarkisto* 19 (1/2006).

(FSD-artikkeli) Numero 19 (1/2006) Electronic Resource, Retrieved in July 20th 2006 from the World Wide Web URL: <http://www.fsd.uta.fi/tietoarkistolehti/19/elsst.html>

Karlsson 2000 = Karlsson, Fred 2000: "Maailman kielitilanne, englannin kieli ja USA:n hegemonia." pp. 69-88. IN: *STUDIA GENERALIA. Suomi EU:n puheenjohtajana, suomi Euroopan kielenä*. Heino, Päivi & Westermarck, Harri. Helsinki: Helsingin yliopiston vapaan sivistystyön toimikunta, Yliopistopaino. 123 pp.

Katan 2004 = Katan, David 2004 (1999): *Translating Cultures : An introduction for Translators, Interpreters and Mediators*. 2nd. ed. Manchester, UK: St. Jemore Publishing. 380 p.

Katzner 2002 = Katzner, Kenneth 2002 (1977): *The Languages of the World*. 3 rd edition. London, UK: Routledge Taylor & Francis Group.

Kela 2009a = The Social Insurance Institution of Finland (Kela): "Paternity allowance". Retrieved in September 18th 2009 from the World Wide Web URL: < http://www.kela.fi/in/internet/english.nsf/NET/180708132828HS?OpenDocument>

Kela 2009b = The Social Insurance Institution of Finland (Kela): "Daddy month." Retrieved in September 18th 2009 from the World Wide Web URL: < http://www.kela.fi/in/internet/english.nsf/NET/210708145722HS?OpenDocument>

Kekäläinen 1999 = Kekäläinen, Jaana (1999): *The effects of query complexity, expansion and structure on retrieval performance in probabilistic text retrieval*. PhD dissertation, Department of Information Studies, University of Tampere. Acta Universitatis Tamperensis 678.

Keränen 2000 = Keränen, Susanna 2000: *Tiedonhaun sanastoja sanastotyössä ilmenneet kielenhuolto-, vierasperäisyys- ja ekvivalenssiongelmat*. Un-published master thesis. University of Tampere, Department of Information Studies.

Keränen 2001 = Keränen, Marja 2001: "Vertaileva ja poikkikulttuurinen tutkimus. Kaksi tapaa lähestyä muita." [" Comparative and Cross-cultural Research. Two Ways of Approaching Other Countries."] *Politiikka* 43:2, p. 82-92, 2001.

Kielikone Ltd. = Kielikone Ltd. – Kielikone Oy: *MOT Finnish-English*. Versions 4.0 and 4.7. Consulted 2001-2009 on the World Wide Web URL: <http://mot.kielikone.fi/>

Kinnunen 2001 = Kinnunen, Merja 2001: *Luokiteltu sukupuoli*. Jyväskylä: Osuuskunta Vastapaino. 187 pp.

Koller 1989 = Koller, Werner 1989: "Equivalence in Translation Theory." IN: Chesterman (Ed.), *Readings in Translation Theory*, Finland, Oy Finn Lectura Ab, pp. 99-104.

----- 1995 = Koller, Werner 1995: "The Concept of Equivalence and the Object of Translation Studies", *Target*, 7 (2), pp. 191-222.

Koskinen 2000 = Koskinen, Kaisa 2000: *Beyond Ambivalence*. *Postmodernity and the Ethics of Translation* [Ambivalenssin tuolle puolen. Postmoderni aika ja kääntäjän etiikka]. Acta Electronica Universitatis Tamperensis; 65, Tampereen yliopisto 2000. ISBN 951-44-4941-X, ISSN 1456-954X. 130 pp. Retrieved in from the World Wide Web, URL: <http://acta.uta.fi/pdf/951-44-4941-X.pdf>] (last visited 7.12.2005)

----- 2001 = Koskinen, Kaisa 2001: "Ekvivalenssista erojen leikkiin – käännöstiede ja kääntäjän etiikka." IN: Oittinen, Riitta & Mäkinen, Pirjo (eds.) 2001: *Alussa oli käännös*. Tampere: Tampere University Press. pp. 374-387.

Kotimaisten kielten tutkimus tutkimuskeskus 2003 = Kotimaisten kielten tutkimus tutkimuskeskus 2003 (2004): *Kielitieteen asiasanasto : suomi – englanti – viro – saksa*. Eds. Kristiansson-Seppälä, Anna-Liisa & Kulonen, Ulla-Maija. 235 pp. Kotimaisten kielten tutkimuskeskuksen julkaisuja 128, ISSN 0355-5437, ISBN 952-5446-03-4, Helsinki 2003. Retrieved in from the World Wide Web, URL: <http://www.kotus.fi/kirjasto/asiasanasto.shtml> (Last updated 20.2.2004, last visited 24.1.2006)

--- 2007 = Kotimaisten kielten tutkimuskeskus 2007 (14.8.2007, last visited 25.2.2009): *On the Research Institute*. Retreieved in August 14th 2007 from the World Wide Web URL: <http://www.kotus.fi/index.phtml?l=en&s=2>

Kramsch 1993 = Kramsch Claire 1993: *Context and Culture in Language Teaching*. Oxford: Oxford University Press. 304 pp.

Kristensen & Järvelin 1990 = Kristensen, Jaana & Järvelin, Kalervo 1990: "The Effectiveness of a Searching Thesaurus in Free-Text Searching of a Full-Text Database." University of Tampere, Department of Library and Information Science, *Research Notes*.

Kuhlthau 2004 = Kuhlthau, Carol Collier 2004: "Seeking meaning. A process approach to library and information services." Westport: Libraries Unlimited. 247 pp.

Kärki & Kortelainen 1996 = Kärki, Riitta; Kortelainen, Terttu 1996: *Johdatus bibliometriikkaan*. Tampere: Informaatiotutkimuksen yhdistys, 1996.113 pp.

Lahelma & Martikainen & Rahkonen 2005 = Lahelma, Eero, Winter, Thorsten, Martikainen, Pekka, Rahkonen, Ossi 2005: "Työn ja perheen väliset ristiriidat ja niiden taustatekijät naisilla ja miehillä." IN: Kauppinen Kaisa (main editor): *Työ ja ihminen* 2005; 19: 34-48. Helsinki: Työterveyslaitos. 162 pp.

Lahtinen 2000 = Lahtinen, T. 2000: *Automatic Indexing: An Approach Using an Index Term Corpus and Combining Linguistic and Statistical Methods*. Helsinki: University of Helsinki, Department of General Linguistics. (Publications: 34).

Landry 2004 = Landry, Patrice 2004: "Multilingual Subject Access: The Linking Approach of MACS". In *The Thesaurus: Review, Renaissance and Revision*. Edited by Alan R Thomas and Sandra K Roe. The Haworth Information Press, NY, USA. 209 pp., p. 177-191.

Layder 1993 = Layder, Derek. 1993: New strategies in social research : An Introduction and Guide. London: Polity Press.

Lehtokangas, R. & Järvelin, K. 2001 = Lehtokangas, R. & Järvelin, K. 2001: "Consistency of textual expression in newspaper articles: An argument for semantically based query expansion". *Journal of Documentation*, 57 (4): 535-548.

Lehtonen 2000 = Lehtonen, Mikko 2000 (1996): *Merkitysten maailma. Kulttuurisen tekstintutkimuksen lähtökohtia*. 3rd ed. Tampere: Vastapaino. 255 pp.

Leininger 2000 = Leininger, Kurt 2000: Interindexer consistency in PsycINFO. *Journal of Librarianship and Information Science*, 32 (1) March 2000. pp. 4-8.

Leiwo 2000 = Leiwo, Matti. (2000). "Suomen kieli 2000-luvulla: Voiko kielen kehitystä ennustaa." IN: Kalaja, P, & Nieminen, L. (eds.) (2000). *Kielikoulussa – kieli koulussa. AfinLAn vuosikirja 2000*. Jyväskylä: Suomen soveltavan kielitieteen yhdistyksen julkaisuja no. 58, pp. 387-404.

Lewis 2009 = Lewis, M. Paul (ed.) 2009: "Statistical summaries" IN *Ethnologue: Languages of the World*. 16th edition. Dallas, Tex.: SIL International. Online version. Last visited in September 29th 2009 in the World wide Web, URL: <http://www.ethnologue.com/ethno_docs/distribution.asp?by=area>

Lindfors 2001 = Lindfors, Anne-Marie 2001: "Respect or ridicule: Translation strategies and the images of a foreign culture." Helsinki English Studies. *The Electronic Journal of the Department of English at the University of Helsinki*. Volume 1, 2001. Retrieved in 2004 from the World Wide Web URL: <http://www.eng.helsinki.fi/hes/Translation/respect_or_ridicule1.htm>

Glossary of linguistic terms 2004/2009 (2.6.2009). Eds. Loos, Eugene E & Anderson Susan & Dwight H. & Day, Jr. & Jordan Paul C & Wingate Douglas J. SIL International (Partners in Language Development). Retrieved in June 2nd 2009 in the World Wide Web URL: <http://www.sil.org/linguistics/GlossaryOfLinguisticTerms/contents.htm>

López-Huertas 2008 = López-Huertas, María 2008: "Some Current Research Questions in the Field of Knowledge Organization." *Knowl. Org*. 35(2008) No.2/No.3. pp. 113-136.

Lykke Nielsen 2002 = Lykke Nielsen, Marianne 2002: *The word association method : a gateway to work-task based retrieval*. Åbo : Åbo Akademi University Press. Doctoral dissertation.

---------- 2004 = Lykke Nielsen, Marianne 2004: "Thesaurus Construction: Key Issues and Selected Readings." IN: *The Thesaurus: Review, Renaissance and Revision* ,pp. 57-74. Edited by Alan R Thomas and Sandra K Roe. The Haworth Information Press, NY, USA. 209 pp..

---------- 2005 = Lykke Nielsen, Marianne 2005: "Special libraries and specialized vocabularies in the digital age. " *Encyclopedia of library and information science* (ELIS). New York, N.Y.: Marcel Dekker.

MADIERA 2009 = The Madiera Portal 2009: "Madiera Portal User Guide". Retrieved in September 29th 2009 from the World Wide Web, URL: <http://www.madiera.net/userguide.shtml>

Mai 2001 = Mai, Jens-Erik 2001: "Semiotics and indexing: an analysis of the subject indexing process." *Journal of Documentation*, 57 (5) 2001: 591-622.

Meriläinen 1984 = Meriläinen, Outi 1984: *Perustiedot kirjastoluokitusten ymmärtämiseksi*. Helsinki: Kirjastopalvelu Oy.

------------- 1985 = Meriläinen, Outi 1985: "Indeksoinnin ulottuvuuksia." IN: *Kirjastovuosikirja 1984*. Ed. Inkeri Hovi [et al.], pp. 48-63. Helsinki: Kirjastopalvelu.

------------- 1996 = --------------------- 1996: Asiasanaekvivalenssi ja tesaurusten yhteensopivuus. *Informaatiotutkimus* 15(3)-1996 , pp. .83-91.

------------- 1997 = Meriläinen, Outi 1997: "Descriptor Equivalence in the Context of Bilingual Indexing." *Libri* 1997, vol. 47, pp. 107-113.

Merriam-Webster Online Dictionary. Merriam-Webster Online.

Retrieved in from the Wide Web 2001-2009 URL: <http://www.m-w.com/.>

Merton & al.1990 = Merton, Robert K & Fiske, Marjorie & Kendall, Patricia L.1990 (1956): *The Focused Interview : A Manual of Problems and Procedures*. 2nd ed. New York: Free Press.

Meschonnic 2003 = Meschonnic, Henri 2003: "Texts on translation." [Introduction and translation by Anthony Pym.] *Target* 15:2 (2003), 337-353.

Miller 2003a = Miller, Uri 2003: "Thesaurus Construction". IN: Encyclopedia of Library and Information Science (ed. By Miriam A. Drake), Volume 4: 2379-3144, Second Edition, pp. 2800-2810 New York: Dekker.

Miller 2003b = Miller, Uri 2003: "Thesaurus and New Information Environment." IN: *Encyclopedia of Library and Information Science* (ed. By Miriam A. Drake), Volume 4: 2379-3144, Second Edition, pp. 2811-2819. New York: Dekker.

Miller & Matthews 2001 = Miller, Ken & Matthews, Brian 2001: "Having the Right Connections: the LIMBER Project." *Journal of Digital information*, volume 1 issue 8. Retrieved in 2001 from the World Wide Web URL: <http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Miller/>

Milstead 1998 = Milstead, Jessica L. 1998: "Use of Thesauri in the Full-Text Environment." Based on a paper presented at the 34th Clinic on Library Applications of Data Processing in Cochrane, P. & Johnson, E (eds.) *Visualizing Subject Access for 21st Century Information Resources; Proceedings of the 34th Annual Clinic on Library Applications of Data Processing*, March 2-4,1997. Champaign, IL: Graduate School of Library and Information Science, University of Illinois, 1998. pp. 28-38. Retrieved in March 23rd 2001 from the World Wide Web URL: <http://www.jelem.com/useof.htm>

Milstead & Feldman 1999 = Milstead, Jessica & Feldman, Susan 1999: "Metadata: Cataloging by Any Other Name …" *ONLINE*, January 1999. Retrieved in May 7th 2001 from the World Wide Web URL: <http://www.onlineinc.com/onlinemag/metadata/>

Ministry of Social Affairs and Health 1999 = Ministry of Social Affairs and Health August 24th 1999: *Finnish Family Policy*. Brochures 1999:9eng. Retrieved in 4th March 2002 from the World Wide Web URL: <http://www.stm.fi/english/pao/publicat/paocontents9.htm>

Ministry of Social Affairs and Health 2002 = Ministry of Social Affairs and Health 2002: *Early Childhood Education and Care in Finland*. Brochures of the Ministry of Social Affairs and Health 2002:3. 12 pp. Helsinki 2002. Retrieved in from the World Wide Web October 31st 2002. URL: <http://www.stm.fi/english/pao/publicat/earlychildhood/earlychildhood.pdf>

ISSN 1236-2123. ISBN 952-00-1195-1.

MORI 2002 = MORI Social Research Institute 2002: *Dads on Dads: Needs and expectations at home and at work*. Research Discussion Series. By Warren Hatter, Louise Vinter, Rachel Williams. Equal Opportunities Commission (EOC): Manchester, United Kingdom. Retrieved in October 30th 2002 from the World Wide Web, URL: <www.eoc.org.uk>

Moring 2005 = Moring, Tom 2005: "Eurooppalainen vähemmistökielipolitiikka." Pp. 74-97. IN: Johansson, Marjut & Pyykkö, Riitta (eds.) 2005: *Monikielinen Eurooppa : Kielipolitiikkaa ja käytäntöä*. Helsinki: Gaudeamus. 374 pp.

Mäkelä 1990 = Mäkelä, K. 1990. "Kvalitatiivisen analyysin arviointiperusteet." Pp. 42-59. IN: Mäkelä, K. (ed.) *Kvalitatiivisen aineiston analyysi ja tulkinta*. Helsinki: Gaudeamus.

Mälkiä 1997 = Mälkiä, Matti: "Kieli, yhteisöllisyys ja valta. Näkökulmia alan tutkimusperinteen kehittämiseksi." Pp. 17-48. In Mälkiä & Stenvall 1997.

Mälkiä & Stenvall 1997 = Mälkiä, Matti & Stenvall, Jari (eds.) 1997: *Kielen vallassa. Näkökulmia politiikan, uskonnon ja julkishallinnon kieleen*. Tampere: Tampere University Press. Pp.233

Nascimento & Marteleto 2008 = Nascimento Denise Morado, & Marteleto Regina Maria 2008: "Social field, domains of knowledge and informational practice." *Journal of Documentation*, 2008, Volume: 64, Issue: 3, Pp. 397 – 412.

Newbury 2003 = *The Newbury House Online Dictionary*. Monroe Allen Publishers, Inc 1999. Retrieved in from the World Wide Web 2003, URL: <http://nhd.heinle.com/>

Newmark 1988 = Newmark, Peter 1988: *A textbook of translation*. New York: Prentice Hall. 292 pp.

Nida 1964/2000 = Nida, Eugene 1964 (2000): "Principles of correspondence." In Venuti 2000, pp. 126-140.

Nida & Taber 1969 = Nida, Eugene, A. & Taber, Charles, R 1969: "Theory and Practice of Translation". *Helps for Translators*, vol. 8. Leiden: E.J. Brill.

Nida 1975 = Nida, Eugene A. 1975: Componential analysis of meaning. An introduction to semantic structures. Mouton, The Hague, Paris. 272 pp.

Nida & Reyburn 1981 = Nida, Eugene & Reyburn, William 1981: *Meaning Across Cultures*. American Society of Missiology Series, No.4, New York: Orbis Books.

Nida 1989 = Nida, Eugene A. 1989: "Science of translation." IN: Chesterman, Andrew (ed.) 1989: *Readings in translation Theory*, pp. 80-98. Helsinki: Finn Lectura.

Nuopponen 1994 = Nuopponen, Anita 1994: *Begreppssystem för terminologisk analys* . [Concept systems for terminological analysis] PhD dissertation. Vaasa: Acta Wasaensia. University of Vaasa.

----------- 2003 = Nuopponen, Anita 2003: "Käsiteanalyysi asiantuntijan työvälineenä." IN: *Kieli ja asiantuntijuus*. AFinLA- vuosikirja, 13-24. Eds. Koskela, Merja & Pilke, Nina. Jyväskylä: Suomen soveltavan kielitieteen yhdistys.

Nätkin 1997 = Nätkin, Ritva 1997: Kamppailu suomalaisesta äitiydestä. Maternalismi, väestöpolitiikka ja naisten kertomukset. Helsinki: Gaudeamus.

OED Online 2001 = Oxford English Dictionary (9.4.2001). Second edition 1989. Oxford University Press, University of Oxford. Retrieved in from the World Wide Web September 3rd 2001 URL: <http://www.oed.com/>

Oittinen 1995 = Oittinen, Riitta 1995: *Kääntäjän karnevaali*. Vammala: Tampere University Press.

Oittinen & Väyrynen 2001 = Oittinen, V. & Väyrynen, K. 2001: "Englannin hegemonia ja humanistinen tutkimus." *Tieteessä tapahtuu* 8/2001 [What happens in science]. The Federation

of Finnish Learned Societies. Retrieved in December 19th 2001 from the World Wide Web, URL: < http://www.tsv.fi/ttapaht/018/kesk.htm#kes>

Olsson 2004 = Olsson, M.R. 2004: "Understanding Users: Context Communication and Construction". Challenging Ideas: ALIA 2004 Biennial Conference, Gold Coast, Australia, September 2004 IN: *Challenging Ideas*, ALIA, online, pp. 1-9. Retrieved in December 8th 2004 in the World Wide Web URL: <http://conferences.alia.org.au/alia2004/pdfs/olsson.m.paper.pdf>

Persson 1991 = Persson, Olle 1991: *Forskning i bibliometrisk belysning*. Umeå: INUM.

Pettersson 2003 = Pettersson, K. 2003: *Den manliga normen : Genomgång av SAB-systemet och Svenska ämnesord utifrån ett genusperspektiv* . Retrieved in from the World Wide Web June 10th 2008 URL: <http: www.ub.gu.se/kvinn/ansvar/sabao.pdf>

Pirkola & al. 2001 = Pirkola, A., Hedlund, T., Keskustalo, H. & Järvelin, K (2001). "Dictionary-based cross-language information retrieval: problems, methods, and research findings." *Information Retrieval*, 4(3/4): 209-230.

Porter 1986 = Porter, James E 1986: "Intertextuality and the Discourse Community." *Rhetoric Review* 5.1 (1986): pp. 34-47.

Potter 1997 = Potter, Jonathan 1997: "Discourse Analysis as a Way of Analysing Naturally Occurring Talk", in Silverman, D., (Ed), *Qualitative Research: Theory, Method and Practice*, London: Sage Publications, pp.144-160.

Potter & Wetherell 1990 = Potter, Jonathan & Wetherell, Margaret 1990: "Discourse: noun, verb or social practice?" *Philosophical Psychology*, Vol. 3/1990, pp. 205-217.

Puolustusvoimat 2000 = Puolustusvoimat 2000: *Puolustushallinnon asiasanasto*. Maanpuolustuskorkeakoulun kirjasto, Oy Edita Ab, Helsinki. Retrieved in January 24th 2006 in the World Wide Web URL: <http://www.mpkk.fi/attachment/ad9d29e3539815313b364464a41b98a9/56572c33e86964a4209 24073a8e6671f/Asiasanasto.pdf]. 119 pp.

Pälli 2003 = Pälli, Pekka 2003: *Ihmisryhmä diskurssissa ja diskurssina*. [Human group in discourse and as discourse] Acta Electronica Universitatis Tamperensis; 231. 247 pp. Tampere University Press, Tampere.

Reiss 1981/2000 = Reiss, Katharina. 1981/2000: "Type, kind and individuality of text. Decision making in translation." Translated by Susan Kitron. In Venuti 2000, pp. 160-171.

Reiss & Vermeer 1986 = Reiss, Katharina & Vermeer, Hans J. 1986 (1984): *Mitä kääntäminen on*. [Original title: Grundlegung einer allgemeinen Translationstheorie. Transl. and ed. Roinila, Pauli.] Helsinki: Gaudeamus.

Rittschoff & al. 1994 = Rittschoff, KA, Stock, WA, Kulhavy, RW, Verdi, MP, and Doran, JM (1994): "Thematic maps improve memory of facts and inferences: a test of the stimulus order hypothesis." *Contemporary Educational Psychology*, 9, 129-42.

Robinson 1950 = Robinson, Richard 1950: *Definition*. Oxford: Clarendon Press.

Rowley 1988 = Rowley, Jennifer 1988 (1982): *Abstacting and indexing*. 2nd ed. London: Clive Bingley Limited.

Ruokonen 2004 = Ruokonen, Minna 2004: "Schleiermacher, Berman ja Venuti: kolme käännösteoreettista näkökulmaa vieraannuttamiseen." IN: Tommola Jorma (ed.) 2004: *Kieli, teksti ja kääntäminen*. *Language, text and translation*, pp. 63-80. Turku: Turun yliopisto.

SA Bluesheet 2001 = Sociological Abstracts (SA) (9.8.2001) (Last Update To Bluesheet: March 2, 1998). Retrieved in from the World Wide Web 2001 URL: <http://library.dialog.com/bluesheets/html/bl0037.html>

Saariluoma 1992 = Saariluoma, Pertti 1992: *Taitavan ajattelun psykologia*. Helsinki: Otava.

Saarti 1999 = Saarti, Jarmo 1999: Kaunokirjallisuuden sisällönkuvailun aspektit. Kirjastoammattilaisten ja kirjastonkäyttäjien tekemien romaanien tiivistelmien ja asiasanoitusten yhdenmukaisuus. Oulun yliopisto, Oulu. [Includes summary in English.]

Saksa 2004 = Saksa, Silja 2004: *Baabelin perilliset. Kääntäjien ja kääntämisen historiaa*. Jyväskylä: Atena. 191 pp.

Schneider 2004 = Schneider, Jesper 2004: *Verification of bibliometric methods' applicability for thesaurus construction*. Doctoral dissertation. Aalborg: Department of Information Studies, Royal School of Library and Information Science. 356 pp. plus Appendix volume, 125 p.

School of Psychology 2002 = School of Psychology, Queen's University, Belfast 2002 (updated in 22.10.2002): *Guidelines for the Use of Non-sexist Language*. Retrieved in from the World Wide Web 2002 URL: <http://www.psych.qub.ac.uk/ug/guide/nonsexist.html>

Schäffner 1998 = Schäffner Christina 1998: "The concepts of Norms in Translation Studies." *Current Issues in Language and Society*, Vol. 5, No 1&2, 1998.

Seale 1998 = Seale Clive (ed.) 1998: *Researching Society and Culture*. London UK: SAGE Publications. 349 pp.

Seppänen 2001 = Seppänen, Janne. 2001. *Katseen voima: kohti visuaalista lukutaitoa*. Tampere: Osuuskunta Vastapaino. 249 pp.

SFS 5471-1988 = SFS 5471 1988-10-17, 1 p., 10 s. [5] Guidelines for the establishment and maintenance of Finnish language thesauri. Suomenkielisen tesauruksen laatimis- ja ylläpito-ohjeet. Suomen Standardisoimisliitto [The Finnish Standards Association], SFS.

SHKS 2005 = Sairaanhoitajien koulutussäätiö 2005: *Hoidokki – Hoitotieteellinen asiasanasto* v.1.0. Ed. Saranto Kaija, Ensio Anneli, Junttila Kristiina, Pekkala Eija & Tallberg Marianne. Retrieved in January 24th 2006from World Wide Web URL: <http://www.shks.fi/mp/db/file_library/x/IMG/11400/file/web_hoidokki030905.pdf>

Slater 1998 = Slater Dan 1998: "Analysing cultural objects: content analysis and semiotics." IN: Seale Clive (ed.) 1998: *Researching Society and Culture*, pp. 233-260. SAGE Publications, London UK. 349 pp.

Snell-Hornby 1988= Snell-Hornby, Mary 1988: *Translation Studies. An Integrated Approach*. Amsterdam: John Benjamins.

Soergel 1974 = Soergel, Dagobert. 1974: *Indexing Languages and Thesauri: Construction and Maintenance*. Los Angeles: Melville.

Social SciSearch® Bluesheet 2001 = The Dialog Corporation (10.8.2001): *Social SciSearch® Bluesheet*, ONTAP® Social SciSearch® (File 207). Last Update To Bluesheet: September 12, 2001. Retrieved in from the World Wide Web URL: <http://library.dialog.com/bluesheets/html/bl0007.html>

SOSIG 2001 = Social Science Information Gateway (SOSIG) Home Page, URL <http://www.sosig.ac.uk/>

Steiner 1975 = Steiner, George 1975: *After Babel*. London, University Press.

STAT 2001a = Statistics Finland 26.6.2001: *Finland in Figures – Foreign Countries*. *Comparative international data*. Retrieved in March 19th from the World Wide Web URL: <http://www.stat.fi/tk/tp/tasku/taskue_ulkomaat.html>

STAT 2001b = Statistics Finland 21.12.2001: *Finland in Figures – Population*. *Families, 2000, Demoghraphic statistics*. Retrieved in March 19th from the World Wide Web URL: <http://www.stat.fi/tk/tp/tasku/taskue_vaesto.html>

STAT 2002a = Statistics Finland 12.2.2002: *Naiset ja miehet numeroina – Perheet*. *Perheet tyypin mukaan 2000*. Retrieved in March 4th 2002 from the World Wide Web URL: <http://www.stat.fi/tk/he/tasaarvo_perheet.html>

STAT 2002b = Statistics Finland 12.2.2002: *Työvoimaosuus 1970 - 2001, 15 - 74-vuotiaat*. Retrieved in March 18th 2002 from the World Wide Web URL:< http://www.stat.fi/tk/he/tasaarvo_tyo.html>

Strauss & Corbin 1990 = Strauss, Anselm & Corbin, Juliet 1990: *Basics of Qualitative Research*. *Grounded Theory Procedures and Techniques*. SAGE Publications, USA.

Suoninen 1992 = Suoninen, Eero 1992: *Perheen Kuvakulmat: Diskurssianalyysi perheenäidin puheesta*. Tampereen yliopisto, Sosiologian ja sosiaalipsykologian laitos. Tutkimuksia, sarja A, nro 24. 137 pp.

Suojanen 1993 = Suojanen, Päivikki 1993: "Identiteetti, perinne, vallan viestit." IN: Suojanen, Päivikki & Suojanen, Matti K. (ed.): *KULTTUURIN KALEIDOSKOOPISTA : Kirjoituksia kielestä ja kulttuurista*. Tampere: Antrokirjat. 174 pp.

Talja 1997 = Talja, Sanna 1997: "Constituting 'information' and 'user' as research objects: a theory of knowledge formations as an alternative to the information-man theory", IN: Vakkari, P., Savolainen, R., Dervin, B. (Eds),*Information Seeking in Context*, pp. 81-96. LONDON: Taylor Graham, London.

---------- 1998 = Talja, Sanna 1998: *Musiikki, kulttuuri, kirjasto : diskurssien analyysi*. Tampere: TAJU. 289 pp.

---------- 1999 = Talja, Sanna 1999: "Analyzing qualitative interview data: The discourse analytic method." *Library & Information Science Research* 21 (4): 459-477.

Tasa-arvotoimisto 1999 = Tasa-arvotoimisto 1/1999: Tasa-arvon vuosituhat. (Yleisesite tasa-arvokysymyksistä ja naisen asemasta Suomessa. 16 pp.) Retrieved March 4th 2002 from the World Wide Web URL: <http://www.tasa-arvo.fi/julkaisuja/VUOSITUH.html>

Tasa-arvotoimisto 2002 = Tasa-arvotoimisto 2002: *Tasa-arvo Suomessa*. (Esite tasa-arvoviranomaisista ja tasa-arvotyön organisoinnista. 8pp.) Retrieved in March 4th 2002 from the World Wide Web URL:

<http://www.tasa-arvo.fi/julkaisuja/jarjes.html>

Temmerman 2000 = Temmerman, Rita 2000: Towards New Ways of Terminology Description: the sociocognitive approach. Amsterdam: John Benjamin.

The Helsinki University Library 2001a = The Helsinki University Library - The National Library of Finland 2001: *Thesauri*. Retrieved in from the World Wide Web March 29th 2001 URL: <http://www.lib.helsinki.fi/hyk/hul/kte/thesaur.html>

The Helsinki University Library 2001b = The Helsinki University Library - The National Library of Finland 2001: The *Article Index of Finnish Journals*. Retrieved in March 29th 2001

from the World Wide Web URL: <http://renki.lib.helsinki.fi/gabriel/en/countries/finland-index-en.html>

The Helsinki University Library 2001c = The Helsinki University Library – The National Library of Finland 2001: *The Union Catalogue of Finnish Research Libraries*. Retrieved in April 10th 2001 from the Wide Web URL: <http://renki.lib.helsinki.fi/gabriel/en/countries/finland-union-en.html>

Thesaurus.com 2002 = Lexico LLC 2002: Thesaurus.com. Retrieved in from the World Wide Web September 4th 2002 URL:<http://www.thesaurus.com/>

Tonkiss 1998 = Tonkiss Alan 1998: "Analysing discourse." IN: Seale Clive (ed.) 1998: *Researching Society and Culture*. SAGE Publications, London UK. 349 pp, p. 245-260.

Toury 1978 = Toury, G. 1978, revised 1995. "The Nature and Role of Norms in Translation." IN: Venuti, L. *The Translation Studies Reader*. London: Routledge.

Townley & Gee 1980 = Townley, Helen & Gee, Ralph 1980: *Thesaurus-Making : Grow your own word-stock*. London: Andre Deutsch Limited.

Trompenaars 1993 = Trompenaars Fons (1993): Riding the Waves of Culture : Understanding Cultural Diversity in Business. London: Brealey. 192 pp.

TSK 1989 = The Finnish Centre for Technical Terminology (TSK) 1989: *Sanastotyön käsikirja. Soveltavan terminologian periaatteet ja työmenetelmät*. (TSK 14, SFS-käsikirja 50) Helsinki: Suomen Standardisoimisliitto SFS.

---------- 1999 = Nykänen, Olli & Kalliokuusi, Virpi / Tekniikan Sanastokeskus ry (TSK) 1999 (2002): "Sanastotyön sanastoa." IN: *Toimikunnista termitalkoisiin — 25 vuotta sanastotyön asiantuntemusta* (ed. Kuhmonen Kaisa), pp. 170-183. Helsinki: The Finnish Centre for Technical Terminology (TSK).

Tudhope et al. 2001 = Tudhope, Douglas & Alani, Harith & Jones,Christoher 2001: "Augmenting Thesaurus Relationships: Possibilities for Retrieval." *Journal of Digital information*, volume 1 issue 8. Retrieved in from the World Wide Web May 7th 2001, URL: <http:// http://journals.tdl.org/jodi/article/view/181/160>

Töttö 2000 = Töttö, Pertti 2000. Pirullisen positivismin paluu. Laadullisen ja määrällisen tarkastelua. Jyväskylä: Vastapaino.

UCL 2002 = Guide to non-discriminatory language 2002 (1-Mar-00). University College London. Page last amended: 1-Aug-02, date approved: 1-Mar-00. Retrieved in from the World Wide Web September 4th 2002, URL: <http://www.ucl.ac.uk/hr/docs/non_discrim_language.php>

UKDA 1999 = The UK Data Archive, University of Essex 1999 (2002): *Introduction to HASSET*. (Last modified by Sharon Bolton on September 30th 1999, retrieved in February 27th 2002 from the World Wide Web URL: <http://dasun1.essex.ac.uk/services/intro.html>

---------- 2001 = The UK Data Archive, University of Essex 2001 (2005): USER GUIDE TO ELSST THESAURUS. Retrieved in from the World Wide Web January 29th 2005, URL: <www.limber.rl.ac.uk/Internal/ Deliverables/D4_2_final_V2.doc >

---------- 2003 = The UK Data Archive, University of Essex 2003. Unpublished working version of ELLST (European Social Science Thesaurus).

United Nations 2003 = United Nations 2003: *Principles and Recommendations for Population and Housing Censuses*, Revision 1. Series M, No. 67, Rev. 1 (United Nations

publication, Sales No. E.98.XVII.1). (paras. 2.67-2.70) Retrieved in January 24th 2003 from the World Wide Web, URL: <http://unstats.un.org/unsd/cdb/cdb_dict_xrxx.asp?def_code=181>

UNESCO 2001 = United Nations Educational, Scientific and Cultural Organisation (2.5.2001): *UNESCO Thesaurus*. URL: <http://www.ulcc.ac.uk/unesco/index.htm>

von Ungern-Sternberg 1994 = von Ungern-Sternberg, Sara 1994: *Verktyg för planering av tvärvetenskaplig informationsförsörjning. En tillämpning på ämnesområdet bioteknik i Finland*. Åbo: Åbo Akademis förlag, 1994. Diss: Åbo Akademi, 318 pp.

University of Southampton 2002 = University of Southampton, Faculty of Social Sciences 2002 (1999): *Research Students Handbook*. Retrieved in 2002 from the World Wide Web, URL: <http://www.socstats.soton.ac.uk/teach/research-handbook.pdf>

Yule 1996 = Yule, George, *The Study of Language* (2nd edition), Cambridge: Cambridge University Press, 1996, xii + ix.+285 pp. ISBN: 0 521 56851 X (paperback)

Varantola 1990 = Varantola, Krista 1990: *Tekniikan suomi yhdentyvässä Euroopassa : Sanastotyön merkitystä koskeva selvitys*. Helsinki: Tekniikan Sanastokeskus ry (TSK).

Vehmas-Lehto 1999 = Vehmas-Lehto, Inkeri 1999: *Kopiointia vai kommunikointia : Johdatus käännösteoriaan*. Helsinki: Oy Finn Lectura Ab. 143 pp.

Venuti 1995 = Venuti, Lawrence 1995: *The Translator's Invisibility : A History of Translation*. London and New York: Routledge. 368 pp.

---------- 1998 = Venuti, Lawrence 1998: *The Scandals of Translation : Towards an Ethics of Difference*. London / New York, Routdlege. 210 pp.

----------- 2000 = Venuti, Lawrence (ed.). 2000. *The Translation Studies Reader*. London: Routledge. 528 pp.

Vermeer 1989 = Vermeer, Hans. J. 1989: "Skopos and commission in translational action." IN: Chesterman, Andrew (ed.): *Readings in Translation Theory*, pp. 173-187. Loimaa: Oy Finn Lectura Ab.

Vickery 1965 = Vickery, B.C. 1965: *On retrieval system theory*. 2nd ed. Butterwoths, London.

------------ 1975 = Vickery, B.C. 1975 (1958): *Classification and Indexing in Science*. Butherworths, London. 3 rd. ed. 227 pp.

Vuori 2001 = Vuori, Jaana 2001: Äidit, isät ja ammattilaiset. Sukupuoli, toisto ja muunnelmat asiantuntijoiden kirjoituksissa. Tampere: Tampere University Press. 430 pp.

Weinberg 1996 = Weinberg, Bella Hass 1996: "Complexity In Indexing Systems -- Abandonment And Failure: Implications For Organizing The Internet." *ASIS 1996 Annual Conference Proceedings*. October 19 - 24 1996. Available at:

URL: <http://www.asis.org/annual-96/ElectronicProceedings/weinberg.html>

Wetherell & Yates 2001 = Wetherell, M., Taylor, S. & Yates, S. J. 2001: *Discourse as data: A guide for analysis*. London: Sage & The Open University. 344 pp.

Whittaker 1989 = Whittaker, J. 1989: "Creativity and conformity in science: Titles, keywords, and co-word analysis." *Social Studies of Science*, pp. 473-496.

Wierzbicka 1985a = Wierzbicka, Anna. "A semantic metalanguage for a cross-cultural comparison of speech acts and speech genres.", *Language in Society* 14:491-514.

---------- 1985b = Wierzbicka, Anna. "Different cultures, different languages, different speech acts: Polish vs. English", *Journal of Pragmatics* 9:145-178.

---------- 1987 = Wierzbicka, Anna. *English speech act verbs: A semantic dictionary*. Sydney: Academic Press. 397 pp.

---------- 1991 = Wierzbicka, Anna. 1991. *Cross-cultural Pragmatics: The Semantics Of Human Interaction*. Berlin: Mouton de Gruyter. 502 pp.

---------- 1996 = Wierzbicka, Anna 1996: *Semantics Primes and Universals*. Oxford University Press, New York. 500 pp.

---------- 1997 = Wierzbicka, Anna 1997: Understanding Cultures Through Their Key Words. English, Russian, Polish, German, and Japanese. New York: Oxford University Press. 317 pp.

Wilson 2002 = Wilson, David 2002: *The Englishisation of Academe: A Finnish Perspective*. Jyväskyl: Jyväskylän yliopiston kielikeskus (Rapotteja 5).

von Ungern-Sternberg 1994 = von Ungern-Sternberg, Sara 1994: *Verktyg för planering av tvärvetenskaplig informationsförsörjning*. *En tillämpning på ämnesområdet bioteknik i Finland*. (A tool for planning information provision in interdisciplinary fields, applied to the field biotechnology in Finland). Åbo Akademi Förlag – Åbo Akademi University Press, Åbo. 318 pp. [Includes summary in English.]

von Ungern-Sternberg 1998 = von Ungern-Sternberg, Sara 1998: "Knowledge organisation and a macro language for indexing in biotechnology. Proceedings of the 6th ASIS SIG/CR Classification Research Workshop held at the 58th ASIS Annual meeting Chicago, Illinois, October 8, 1995." Ed. by Raymond P.Schwartz. ASIS Monograph Series Information Today, Medford, 1998. *Advances in Classification Research* 6, 1998, pp. 181-194.

WordNet = The WordNet book, WordNet: "An Electronic Lexical Database". Princeton University. URL: <http://wordnet.princeton.edu/>

Worsfold 1999 = Worsfold Emma 1999 (2001): "Subject gateways - fulfilling the DESIRE for knowledge." (Pre-print of a paper presented at the TERENA Networking Conference '98 (TNC'98) in Dresden, Germany and pre-print of the article published in the Elsevier journal Computer Networks and ISDN Systems (Vol 30 Numbers 12-18) 30th Sept 1998.) Retrieved in May 2nd 2001from the World Wide Web URL: <http://www.desire.org/html/research/publications/tnc98gateways/>

Zethsen 2004 = Zethsen, Karen Korning 2004: "Latin-based terms : True or false friends?" *Target* 16:1 (2004), p. 125-142.

Zhangs, Anghelescus & Yuans 2005 = Zhang, X., Anghelescu, H. G. B. & Yuan, X. (2005): "Domain knowledge, search behaviour, and search effectiveness of engineering and science students: an exploratory study." *Information Research*, 10 (2). Retrieved in from the World Wide Web July 7th 2008, URL <http://InformationR.net/ir/10-2/paper217.html>

Appendix 1: Interview guide for the Finnish social scientists

A Background of the study

B Practice and ethics: tape recording, naming the organisation and the interviewee

Theme 1: Background

- Job description, research subject(s), work experience and research history, language skills

Theme 2: Indexing practices

- Experience in key wording own documents

- Tools and principles used, why them

Theme 3: Information seeking

- Experience in information seeking –

as end user and

as information services customer

- Databases, other sources, tools, strategies and terms used

- If contracts information seeking to information specialist, does she/he give any recommendations etc about the terms and databases to be used

- Use of Internet, how, for what purposes

Theme 4: Family roles as concepts

4.1 How do you understand "family roles"? What do they mean and refer to?

4.2 How (and why so) would you divide them (sub-concepts) [The interviewee's own division]?

4.2.1 – in Finland/in Finnish

4.2.2 - in England/in English

4.2.3 – in Europe/in English

4.3 Differences in terminology

- Conceptual differences in Finnish versus British versus European discourses discussion

- Examples about similarities and differences

Theme 5. **Word associations**

- Five word associations (words, synonyms, phenomena, matters etc.) that comes first into mind [ELSST-terms used as stimulus words]

5.1 family roles

5.2 homemakers

5.3 breadwinners

5.4 heads of household

5.5 housewives

Theme 6: Definitions and translations

- definitions and translations for the stimulus words

- reasoning

6.1 family roles

6.2 homemakers

6.3 breadwinners

6.4 head of household

6.5 housewives

Theme 7: Thesauri and ELLST – contextual background information

- Briefly about thesaurus function and manner of representation

- Informing the respondents not familiar with ELSST about European language social science thesaurus, ELSST, and about the European metadata project OR discussing with the respondents familiar with ELSST about the project.

Theme 8: Family roles in ELSST

- translations provided in ELSST, their usability and relevance in Finland

English version:

Family Roles

      NT   Breadwinners

             Heads of Household

             Homemakers

              SN *A person, either male or female, whose role is to look after the home and family instead of earning money from employment*

                NT Housewives

Finnish version:

Perheroolit

      ST   Perheenelättäjät

             Perheenpäät

             Kotitaloustyötä tekevät

              Selite *Palkatonta kotitaloustyötä tekevä henkilö, joka hoitaa kodin ja*

             perheen palkkatyön tekemisen sijasta

                KT Perheenemännät

                NT Kotiäidit + Koti-isät

Theme 9: Free discussion, other comments and questions

- Issues that were mentioned by the interviewee during the interview

- Experiences in the use of international material (from the terminological point of view)

Appendix 2: Interview guide for the British social scientists

A Background of the study

B Practice and ethics: tape recording, naming the organisation and the interviewee

Theme 1: Background

- Job description, research subject(s), work experience and research history, language skills

Theme 2: Indexing practices

- Experience in key wording own documents

- Tools and principles used, why them

Theme 3: Information seeking

- Experience in information seeking –

as end user and

as information services customer

- Databases, other sources, tools, strategies and terms used

- If contracts information seeking to information specialist, does she/he give any recommendations etc about the terms and databases to be used

- Use of Internet, how, for what purposes

Theme 4: Family roles as concepts

4.1 How do you understand "family roles"? What do they mean and refer to?

4.2 How (and why so) would you divide them (sub-concepts) [The interviewee's own division]?

4.2.1 – emphasis on England

4.2.2 – emphasis on Europe

4.2.3 – emphasis on Scandinavia/Nordic countries(/Finland)

4.3 Differences in terminology
- Conceptual differences in (Scandinavian versus) British versus European discussion

- Examples about similarities and differences

Theme 5. **Word associations**

- Five word associations (words, synonyms, phenomena, matters etc.) that comes first into mind [ELSST-terms used as stimulus words]

5.1 family roles

5.2 homemakers

5.3 breadwinners

5.4 heads of household

5.5 housewives

Theme 6: Definitions and translations

- Definitions of the English source terms + reasons to do so

6.1 family roles

6.2 homemakers

6.3 breadwinners

6.4 head of household

6.5 housewives

Theme 7: Thesauri and ELLST – contextual background information

- Briefly about thesaurus function and manner of representation

- Informing the respondents not familiar with ELSST about European language social science thesaurus, ELSST, and about the European metadata project OR discussing with the respondents familiar with ELSST about the project.

Theme 8: Family roles in ELSST

Family Roles

   NT Breadwinners

      Heads of Household

      Homemakers

       SN A person, either male or female, whose role is to look after the home and

      family instead of earning money from employment

       NT Housewives

Their usability and relevance in UK and in a social science thesaurus? Their usability in information seeking? How and in which databases? Expectations about search results? Why so?

Theme 9: Free discussion, other comments and questions

- Issues that were mentioned by the interviewee during the interview

- Experiences in the use of international material (from the terminological point of view)

Appendix 3: Interview guide for the Finnish indexers

A Background of the study

B Practice and ethics: tape recording, naming the organisation and the interviewee

Theme 1 Background

- Job description? Work history? Educational background? Knowledge of foreign languages?

Theme 2: Material indexed and tools and guidelines used

What kind of material do you index?

To which database(s)?

Do you index in other languages than English? – Why? In which languages?

What tools (controlled vocabularies, thesauri e.g.) do you normally use in indexing? What guidelines? Why them?

Theme 3: Equivalence

- What is equivalence? What is aimed at in multilingual indexing?

Theme 4: Family roles as concepts

4.1 What is meant with "family roles"?

4.2 How would you divide them into sub-concepts? Why so?

4.2.1 – in Finland/in Finnish

4.2.2 - in England/in English

4.2.3 – in Europe/in English

4.3 Terminological differences

4.3.1 Conceptual differences in Finnish versus British versus European discourses

4.3.2 Examples about similarities and differences

4.4 How would you represent them in a thesaurus
- a) when thinking of your own working environment and customers/end-users?
- b) In a European social science thesaurus?
  - Why these terms and structure?

Theme 5: Word associations

Five first word associations (words, synonyms, phenomena, matters etc.) that comes into mind [ELSST-terms used as stimulus words]

5.1 family roles

5.2 breadwinners

5.3 heads of household

5.4 homemakers

5.5 housewives

Theme 6: Definitions and translations

- definitions and translations for the stimulus words

- reasoning

6.1 family roles

6.2 homemakers

6.3 breadwinners

6.4 head of household

6.5 housewives

Theme 7: ELLST – contextual background information

Informing the respondents not familiar with ELSST about European language social science thesaurus, ELSST, and about the European metadata project OR discussing with the respondents familiar with ELSST about the project.

Theme 8: Family roles in ELSST

- translations provided in ELSST, their usability and relevance in Finland

English version:

Family Roles

NT   Breadwinners

          Heads of Household

          Homemakers

            SN *A person, either male or female, whose role is to look after the home and family instead of earning money from employment*

              NT Housewives

Finnish version:

        Perheroolit

          ST Perheenelättäjät

            Perheenpäät

            Kotitaloustyötä tekevät

              Selite *Palkatonta kotitaloustyötä tekevä henkilö, joka hoitaa kodin ja perheen palkkatyön tekemisen sijasta*

                KT Perheenemännät

                NT Kotiäidit + Koti-isät

Theme 9: Free discussion, other comments and questions

- Issues that were mentioned by the interviewee during the interview

- Usability of thesauri

- Experiences in the indexing of foreign versus native material

- Internet era (needs for multilingual tools, possible change in construction and representation)

Theme 10: Simulated indexing situation

How would you index these articles a) in your normal working environment? b) in a European multilingual environment like e.g. in a common database for European social sciences?

Please use 9-12 descriptors in both tasks.

Why you used these concepts and terms (will be asked in the interview)? Tools used and principles followed? Were any of the documents specifically difficult or easy? Other experiences?

1. Arnlaug Leira 2002: *Updating the "gender contract"? Childcare reforms in the Nordic countries in the 1990s'* NORA no. 2 2002, Volume 10, s. 81-89.

2. Bridges Judith S. & Etaugh Claire & Barnes-Farrell Janet 2002: *Trait judgments of stay-at-home and employed parents: A function of social role and/or shifting standards?* Psychology of Women Quarterly, 26 (2002), pp. 140-150.

3. Nock Steven L. 2001: *The Marriages of Equally Dependent Spouses*. Journal of Family Issues, Vol. 22 No.6, September 2001, pp. 755-775.

4. Jalovaara Marika 2002: *Socioeconomic Differentials in Divorce Risk by Duration of Marriage*. Demographic Research, Volume 7, Article 16, Published 29 November 2002, URL: www.demographic-research.org, pp. 537-564.

5. Smith Calvin D. 1998: *Men Don't Do This Sort of Thing*. *A Case Study of the Social Isolation of Househusbands*. Men and Masculinities, Vol. 1 No. 2, October 1998, pp. 138-172.

Appendix 4: Interview guide for the British indexers

A Background of the study

B Practice and ethics: tape recording, naming the organisation and the interviewee

Theme 1 Background

- Job description? Work history? Educational background? Knowledge of foreign languages?

Theme 2: Material indexed and tools and guidelines used

What kind of material do you index?

To which database(s)?

Do you index in other languages than English? – Why? In which languages?

What tools (controlled vocabularies, thesauri e.g.) do you normally use in indexing? What guidelines? Why them?

Theme 3: Equivalence

- What is equivalence? What is aimed at in multilingual indexing? (Reasoning, possible authorities like ISO standards)

Theme 4: Family roles as concepts

4.1 What is meant with "family roles"?

4.2 How would you divide them into sub-concepts? Why so?

4.2.1 – in England/in English

4.2.2 - in Europe/in English

4.2.3 – in Scandinavia/ in Finland

4.3 Terminological differences

4.3.1 Conceptual differences in British versus European(/Scandinavian) discourses

4.3.2 Examples about similarities and differences

4.4 How would you represent them in a thesaurus
- when thinking of your own working environment and customers/end-users?
- In a European social science thesaurus?

- Why these terms and structure?

Theme 5: Word associations

Five first word associations (words, synonyms, phenomena, matters etc.) that comes into mind [ELSST-terms used as stimulus words]

5.1 family roles

5.2 breadwinners

5.3 heads of household

5.4 homemakers

5.5 housewives

Theme 6: Definitions and translations

- definitions and translations for the stimulus words

- reasoning

6.1 family roles

6.2 homemakers

6.3 breadwinners

6.4 head of household

6.5 housewives

Theme 7: ELLST – contextual background information

Informing the respondents not familiar with ELSST about European language social science thesaurus, ELSST, and about the European metadata project OR discussing with the respondents familiar with ELSST about the project.

Theme 8: Family roles in ELSST

- usability and relevance in England and in social science thesaurus

Family Roles

    NT   Breadwinners

             Heads of Household

             Homemakers

               SN *A person, either male or female, whose role is to look after the home and family instead of earning money from employment*

               NT Housewives

Theme 9: Free discussion, other comments and questions

- Issues that were mentioned by the interviewee during the interview

- Usability of thesauri

- Experiences in the indexing of foreign versus native material

- Internet era (needs for multilingual tools, possible change in construction and representation)

Theme 10: Simulated indexing situation

How would you index these articles a) in your normal working environment? b) in a European multilingual environment like e.g. in a common database for European social sciences?

Please use 9-12 descriptors in both tasks.

Why you used these concepts and terms (will be asked in the interview)? Tools used and principles followed? Were any of the documents specifically difficult or easy? Other experiences?

1. Arnlaug Leira 2002: *Updating the "gender contract"? Childcare reforms in the Nordic countries in the 1990s'* NORA no. 2 2002, Volume 10, s. 81-89.

2. Bridges Judith S. & Etaugh Claire & Barnes-Farrell Janet 2002: *Trait judgments of stay-at-home and employed parents: A function of social role and/or shifting standards?* Psychology of Women Quarterly, 26 (2002), pp. 140-150.

3. Nock Steven L. 2001: *The Marriages of Equally Dependent Spouses*. Journal of Family Issues, Vol. 22 No.6, September 2001, pp. 755-775.

4. Jalovaara Marika 2002: *Socioeconomic Differentials in Divorce Risk by Duration of Marriage*. Demographic Research, Volume 7, Article 16, Published 29 November 2002, URL: www.demographic-research.org, pp. 537-564.

5. Smith Calvin D. 1998: *Men Don't Do This Sort of Thing. A Case Study of the Social Isolation of Househusbands*. Men and Masculinities, Vol. 1 No. 2, October 1998, pp. 138-172.

Appendix 5: Interview guide for the Finnish thesaurus constructors

A. Background of the study

B. Practice and ethics: tape recording, naming the organisation and the interviewee

Theme 1: Background

- Job description, work experience (also in thesaurus projects as well other experience as e.g. an indexer)? Educational background? Knowledge on foreign languages? Thesaurus languages?

Theme 2: Tools and guidelines used

What guidelines have you followed? What have been the role models, term sources etc.? Working method? Why them? (If not brought up by the respondent, separately the role of General Finnish Thesaurus, YSA.)

3. Equivalence

What is equivalence? What is aimed at in constructing multilingual thesaurus? (Reasoning, possible authorities like ISO standards) (ISO=International Organization for Standardization)

4. Family roles as concepts

4.1 How do you understand "family roles"? What do they mean and refer to?

4.2 How (and why so) would you divide them (sub-concepts) [The interviewee's own division]?

4.2.1 – emphasis on Finland/Finnish

4.2.2 – emphasis on England/English

4.2.3 – emphasis on Europe/English

4.3 Differences in terminology

4.3.1 Conceptual differences in Finnish versus in European discourses

4.3.2 Examples about similarities and differences

4.5 How would you represent them
 -in your normal working environment?
- in a European multilingual environment like e.g. in a multilingual thesaurus for European social sciences?

- Why you used these terms/structure etc?

5. Word associations

Five first word associations (words, synonyms, phenomena, matters etc.) that comes into mind [ELSST-terms used as stimulus words]

5.1 family roles

5.2 breadwinners

5.3 head of household

5.4 homemaker

5.5 housewives

6. Definitions and translations

- definitions and translations for the stimulus words

- reasoning

6.1 family roles

6.2 homemakers

6.3 breadwinners

6.4 head of household

6.5 housewives

7. ELSST – contextual background information

Informing the respondents not familiar with ELSST about European language social science thesaurus, ELSST, and about the European metadata project OR discussing with the respondents familiar with ELSST about the project.

8. Family roles in ELSST

- translations provided in ELSST, their usability and relevance in Finland

English version:

Family Roles

      NT   Breadwinners

              Heads of Household

              Homemakers

               SN *A person, either male or female, whose role is to look after the home and family instead of earning money from employment*

                NT Housewives

Finnish version:

Perheroolit

ST   Perheenelättäjät

       Perheenpäät

       Kotitaloustyötä tekevät

         Selite *Palkatonta kotitaloustyötä tekevä henkilö, joka hoitaa kodin ja perheen palkkatyön tekemisen sijasta*

          KT Perheenemännät

           NT Kotiäidit + Koti-isät

9. Other experiences

- issues that were mentioned by the interviewee during the interview

- other experiences in multilingual thesaurus construction

- usability and future of multilingual thesauri (Internet era, social sciences, linguistic and cultural differences)

Appendix 6: Interview guide for the British thesaurus constructors

A. Background of the study

B. Practice and ethics: tape recording, naming the organisation and the interviewee

Theme 1: Background

- Job description, work experience (also in thesaurus projects as well other experience as e.g. an indexer)? Educational background? Knowledge on foreign languages? Thesaurus languages?

Theme 2: Tools and guidelines used

What guidelines have you followed? What have been the role models, term sources etc.? Working method? Why them?

Theme 3: Equivalence

What is equivalence? What is aimed at in constructing multilingual thesaurus? (Reasoning, possible authorities like ISO standards)

Theme 4: Family roles as concepts

4.1 How do you understand "family roles"? What do they mean and refer to?

4.2 How (and why so) would you divide them (sub-concepts) [The interviewee's own division]?

4.2.1 – emphasis on English/England

4.2.2 – emphasis on English/Europe

4.2.3 – emphasis on Scandinavia/Nordic countries (or Finland)

4.3 Differences in terminology

4.3.1 Conceptual differences in British versus European discourse

4.3.2 Examples about similarities and differences

4.4 How would you represent them
- in your normal working environment?
-in a European multilingual environment like e.g. in a multilingual thesaurus for European social sciences?

- Why you used these terms/structure etc?

Theme 5: Word associations

Five first word associations (words, synonyms, phenomena, matters etc.) that comes into mind [ELSST-terms used as stimulus words]

5.1 family roles

5.2 breadwinners

5.3 head of household

5.4 homemaker

5.5 housewives

Theme 6: Definitions and translations

- definitions and translations for the stimulus words

- reasoning

6.1 family roles

6.2 homemakers

6.3 breadwinners

6.4 head of household

6.5 housewives

Theme 7: ELLST – contextual background information

Informing the respondents not familiar with ELSST about European language social science thesaurus, ELSST, and about the European metadata project OR discussing with the respondents familiar with ELSST about the project.

Theme 8: Family roles in ELSST

- usability and relevance in England and in social science thesaurus

Family Roles

      NT   Breadwinners

              Heads of Household

              Homemakers

                SN *A person, either male or female, whose role is to look after the home and family instead of earning money from employment*

                NT Housewives

Theme 9: Free discussion, other comments and questions

- issues that were mentioned by the interviewee during the interview

- other experiences in multilingual thesaurus construction

- usability and future of multilingual thesauri (Internet era, social sciences, linguistic and cultural differences)

Appendix 7: Original abstracts of the articles

Article 1. Arnlaug Leira 2002: *Updating the "gender contract"? Childcare reforms in the Nordic countries in the 1990s'* NORA no. 2 2002, Volume 10, pp. 81-89.

Abstract:

"Welfare state support for the reconciliation of work and family has long been regarded as a characteristic of the Scandinavian welfare state and trademark of its "woman-friendly" policies. Based on an examination of important Nordic childcare policy reforms of the 1990's, such as the expansion of state-sponsored childcare services, the strengthening of fathers' rights to care and the institution of cash grants for childcare, the impact of reforms on mothers and fathers as workers and carers is discussed. In conclusion, the discussion returns to the reconciliation of work and family, and asks: To what extent have reproduction policies succeeded in updating the traditional gender contract of the male breadwinner family?"

Article 2. Bridges Judith S. & Etaugh Claire & Barnes-Farrell Janet 2002: *Trait judgments of stay-at-home and employed parents: A function of social role and/or shifting standards?* Psychology of Women Quarterly, 26:2 (2002), pp. 140-150.

Abstract:

"Primarily middle class, white college students read brief descriptions of stay-at-home or employed mothers or fathers, estimated how often they performed several parenting behaviors, and rated them on communion and parenting effectiveness. Respondents estimated more parenting behaviors for mothers than fathers in both roles, gave stay-at-home mothers and fathers similar trait ratings, and viewed employed mothers as lower in communion and parenting effectiveness than fathers."

Article 3. Nock Steven L. 2001: *The Marriages of Equally Dependent Spouses*. Journal of Family Issues, Vol. 22 No.6, September 2001, pp. 755-775.

Abstract:

"Marriages of equally dependent spouses (MEDS) are those in which each partner generates between 40% and 59% of the total family earnings. The author argues that such marriages will become increasingly common. Currently, about a fifth (22%) of all marriages are MEDS. This article shows that when marriages become MEDS, wives become less committed to the union although husbands do not. When marriages become MEDS, the odds of divorce also increase. Such divorces are more often at the wife's initiative than are divorces among other couples. These findings are interpreted to suggest that wives are more sensitive to the quality of their marriages than husbands are. This is because men benefit from the status of being married (i.e., being a husband) regardless of the emotional quality of their relationship. The benefits that wives derive from marriage, however, appear to depend on the quality of their unions."

Article 4. Jalovaara Marika 2002: *Socioeconomic Differentials in Divorce Risk by Duration of Marriage*. Demographic Research, Volume 7, Article 16, Published 29 November 2002, URL: www.demographic-research.org, pp. 537-564.

Abstract:

"Using register-based data on Finnish first marriages that were intact at the end of 1990 (about 2.1 million marriage-years) and followed up for divorce in 1991-1993 (n = 21,204), this research explored the possibility that the effect of spouses' socioeconomic position on divorce risk varies according to duration of marriage. The comparatively high divorce risks for spouses with little

formal education and for spouses in manual worker occupations were found to be specific to marriages of relatively short duration. In contrast, such factors as unemployment, wife's high income, and living in a rented dwelling were found to increase divorce risk at all marital durations."

Article 5. Smith Calvin D. 1998: *Men Don't Do This Sort of Thing. A Case Study of the Social Isolation of Househusbands. Men and Masculinities*, Vol. 1 No. 2, October 1998, pp. 138-172

Abstract:

"Prior research has found that househusbands suffer alienation and ostracism from a variety of sources. Based on in-depth interviews with eleven househusbands, this article builds on such research by outlining and analyzing some of the mechanisms of this alienation and some of the adaptations these men made to deal with these experiences. Of particular interest are the problems the men report having with being seen by others as legitimately involved in child care and (to a lesser extent) housework. These data strongly support the idea that the men's sex category overrides other positionings that may be relevant, such as "competent housekeeper" or "full-time househusband and child carer." Put differently, hegemonic conceptions of who ought to be minding the children and the house subvert or thwart these men's attempts to validate themselves and these practices. The consequences are a feeling of illegitimacy on one hand and social isolation on the other."

Appendix 8: Semantic lexical networks in thesauri

Note: In the tables additions are marked with parenthesis (x) and removals with brackets [x]. Scope notes are not included in numbers which describe how many different terms are represented with the studied keyword.

(In updates HASSET represents a new version, 3.0, created in 2006.)

Table 36: Family roles' lexical network in the studied thesauri (situation in 26.2.2003 and in 29.1.2009)

family role(s)	UNESCO		HASSET	SOSIG (GENERAL SOCIAL SCIENCE)[104]	ERIC[105]	CSA SOCIO-LOGICAL THE-SAURUS	TOT.
SN							1 (1)
Refers to the various social roles within family structures (eg, mother, father, grandparent, child). Do not select in reference to the effects of the family as a whole on behaviors or social problems. Added, 1986.						x	
UF							0 (0)
TT/MT							4 (4)
family		x					
family environment			x				
roles			x				
social structure			x				
BT							4 (4)

---

[104] No hits in SOSIG's other thesauri: Government, Politics and Anthropology nor in Social Work and Welfare

[105] Added in 7.1.1966. According to the thesaurus term information in 22.1.2009 in ERIC database 1995 records (*postings*) indexed with thesaurus term "family role".

family	ˎ					
family environment		x	(x)			
role/roles		x	x	x	x	
social structure		x	(x)			
NT						6 (6)
bread(-)winners		x	x			
heads of household		x	(x)			
homemakers		x	(x)			
householders		(x)				
housewives			[x]			
parent(al) role	ˎ	(x)	[x]			
RT						18 (18)
adult children				[x]	x	
care of dependants		x	x			
caregiver role				x		
child role				x		
children					x	
domestic responsibilities		x	(x)			
family (sociological unit)/ families		x	x	x		
family caregivers				[x]		
family environment	ˎ					
family influence				x		
family involvement				x		
family life					x	
family relations					x	
family support				(x)		
fathers					x	
filial responsibility					x	
mothers					x	
parent child relations					x	
parent role				x		

TOT.		40	12 (14)	6 (9)	9 (8)	10 (10)	

Table 37: Lexical network of the term "breadwinner(s)" in the studied thesauri (situation in 26.2.2003 and in 29.1.2009)

BREAD(-)WINNER(S)	ERIC	EUROVOC	ELSST	HASSET	SOSIG	TOT.
SN				(x)		1 (2)
Those whose earnings are the primary source of support for themselves and their dependants. Scope note added April 2006.						
Added 11/111969	x					
UF						0 (0)
USE						1 (1)
Heads of Household	x					
RECORD TYPE						
Synonym	x					
SYNONYMS						(0/3)
breadwinners				(x)		
chief income earners				(x)		
main income earners				(x)		
BT						2 (5)
family		x				
family environment				(x)		
family role(s)			x	x	x	
roles				(x)		
social structure				(x)		
BT2						0
BT3						0
MT/TT						4 (3)
family		x				
family environment			x	[x]		
roles				[x]		

social structure			x	[x]		
NT						0(0)
RT						1 (1)
heads of household			x	x	(x)	
TOT.	2 (2)	2 (2)	4	5 (8)	1 (2)	

Table 38: Head(s) of the household's lexical network in the studied thesauri (situation in 26.2.2003 and in 29.1.2009)

HEAD(S) OF HOUSEHOLD	EUROVOC	ELSST	HASSET	ERIC	CSA SOCIOLOGICAL THESAURUS	TOT.
SN						1 (2)
Added, 1989					x	
'head of household' was originally defined by the Office for National Statistics (ONS) as "The eldest householder, with males taking precedence over females in the case of couples or non-related joint-householders".			(x)			
From 2001 ONS started using 'household reference person' which is defined as "the person responsible for owning or renting or who is otherwise responsible for the accommodation. In the case joint-householders, the person with the highest income takes precedence". It should be noted that this term is not restricted to these precise						

definitions, as they may change over time and across different studies. Scope note added April 2006.					
UF					1
head of family	x				1
breadwinners				x	
displaced homemakers				(x)	
family breadwinners				x	
household heads				x	
SYNONYMS					2
household reference person		(x)			1
HRP		(x)			1
MT					1
family	x				1
TT					3
family environment		x	[x]		2
roles			[x]		1
social structure		x	[x]		2
BT					3
family	x				1
family environment			(x)		1
family roles		x	x		2
groups				x	1
roles			(x)		1
social structure			(x)		
BT2					0
BT3					0
NT					0

RT						19 (22)
bread(-)winners		x	x			2
displaced homemakers				x		1
household(s)	x	x	x		x	4
house husbands			(x)			(1)
family (sociological unit)				[x]	x	1
fatherless family				x		1
fathers				x		1
female headed households					x	1
household heads' economic activity			x			1
household heads' educational background			x			1
household heads' occupational status			x			1
household heads' occupation(s)			x			1
household heads' place of birth			x			1
household heads' social class			(x)			
household heads' socio-economic status			(x)			
household heads' wages			(x)			
householders		x	x			2
housewives		x	x			2
motherless family				x		1
mothers				x		1
one parent family				x		1
parents				x	x	2

TOT:	4 (4)	7	13 (20)	12 (12)	5 (5)	

Table 39: Lexical network of the term "homemaker" in the studied thesauri (situation in 26.2.2003 and in 29.1.2009)

HOMEMAKERS	UNESCO	ELSST	HAS-SET	ERIC	CSA SOCIOLOGICAL THESAURUS	SOSIG	TOT.
SN							3 (4)
"Men or women who carry major responsibilities for household or family management"				X			
A person, either male or female, whose role is to look after the home and family instead of earning money from employment		X					
"Formerly (1964-1985) DC 214425, Homemaker/Homemakers"					X		
A person, either male or female, whose role is to look after the home and family instead of earning money from employment. Term created Sept. 2000. Prior to this the term "Domestic Responsibilities" might have been used.			(X)				
UF							2 (2)
homemaking					X		
housewife/housewives	X				X		
BT							4 (9)
economic activity		X	X			X	3
family (sociological unit)				(X)			

family environment			(X)				
family roles		X	X			X	3
groups				X			1
labour and employment			(X)				
roles			(X)				
social structure			(X)				
women	X						1
BT2							1 (1)
sex	X						
BT3							1 (1)
sex distribution	X						
MT/TT							5 (5)
family environment		X	X				
labour and employment		X	X				
population	X						
roles			X				
social structure		X	X				
NT							1 (2)
house husbands			(X)				
housewives		X	X			X	
RT							17 (16)
displaced homemakers				[X]			1
domestic responsibilities		X	X			X	3
family (sociological unit)				X			1
family life				X	X		2
family structure				X			1
home economics				X			1
home economics education	X						1
home management				X			1
homemaking skills				X			1

households	X						1
housework				X	X		2
married women	X						1
mothers	X						1
sexual division of labor					X		1
spouses				X			1
wives					X		1
workers					X		1
TOT: [106]	9 (9)	7	8 (13)	11	8 (8)		

---

[106] Update in 29.1.05, SOSIG included

Table 40: Lexical network of the term "housewives" in the studied thesauri (situation in 26.2.2003 and in 29.1.2009)

HOUSEWIVES	UNES-CO	ELSST	HASSET	ERIC	CSA SOCIOLOGICAL THESAURUS	SOSIG	TOT.
SN							1
1968 1980				x			1
USE							
homemakers	x			x	x		3
BT							5
family environment			x				1
homemakers		x	x			x	3
labour and employmenl			x				1
roles			x				(1)
social structure			x				(1)
RT							(3)
economic activity			(x)				(1)
heads of household			(x)				(1)
mothers			(x)				(1)
TOT:	1	1	5(8)	2	1	1	

Appendix 9: Example letter of interview request

Dear Professor ----,

I am writing to kindly ask you about your participation in a research interview. Due to your expertise in ---- research it would be most valuable for me to have you as one of the British specialists interviewed.

I am a researcher and doctoral student at the Åbo Akademi University (Finland), Department of Social and Political Sciences / Information Studies. My on-going PhD-project is about multilingual thesaurus construction and indexing practices (translatability of British-English social science terms into Finnish language and culture). It is a part of a bigger research project called "Cultural and linguistic differences in digital storage and retrieval of information". The project leader and my supervisor is professor Sara von Ungern-Sternberg and the project is financed by the Academy of Finland, targeted programme on the Production, Management and Use of Digital Information Resources. The Finnish Cultural Foundation has also financed this subproject. As a case I am using terminology related to families.

I am planning to arrive to England late November in order to make my interviews. Do you think you could find some time in early December (about one-two hours) for the interview?

The interview is about definitions and use of certain British-English terms. I am happy to give more information about my study if needed.

Thank you for your effort!

Yours sincerely, Susanna

--

Susanna Keränen

Dpt. of Social and Political Sciences / Information Studies

Åbo Akademi University

Tavastgatan 13, 20500 Åbo, Finland

e-mail susanna.keranen@abo.fi

tel. +358-2-215 3405

In the present dissertation, multilingual thesauri were approached as cultural products and the focus was twofold: On the empirical level the focus was placed on the translatability of certain British-English social science indexing terms into the Finnish language and culture at a concept, a term and an indexing term level. On the theoretical level the focus was placed on the aim of translation and on the concept of equivalence. In accordance with modern communicative and dynamic translation theories the interest was on the human dimension. The study is qualitative.

The study used multiple cases and several data collection and analysis methods aiming at theoretical replication and complementarity.

The study showed that the variety of different discourses should be acknowledged, there is a need for operationalisation of new types of multilingual thesauri, and the factors influencing pragmatic indexing term equivalence should be discussed more precisely than is traditionally done.

Åbo Akademi University Press

Susanna Nykyri
Equivalence and Translation Strategies in Multilingual Thesaurus Construction

Åbo Akademi University

2010