CiMeC
Center for Mind/Brain Sciences

University of Trento

STaRS.sys

designing and building a commonsense-knowledge enriched wordnet for therapeutic purposes

by

Gianluca E. Lebani

advisor: Emanuele Pianta

Doctoral School in Cognitive and Brain Sciences

XXIV cycle

a dissertation presented in partial fulfillment of the requirements for the degree of *Doctor of Philosophy*

December, 2012

Thesis Committee

CHRISTIANE FELLBAUM
Department of Computer Science
Princeton University
Princeton, New Jersey

RAFFAELLA BERNARDI
Department of Information Engineering and Computer Science
Center for Mind/Brain Sciences
University of Trento
Trento

BERNARDO MAGNINI
Human Language Technology research unit
Fondazione Bruno Kessler
Povo, Trento

ROBERTO BASILI
Department of Computer Science
University of Roma, Tor Vergata
Rome

Abstract

THIS THESIS INVESTIGATES THE POSSIBILITY TO EXPLOIT HUMAN LANGUAGE RESOURCES AND KNOWLEDGE EXTRACTION TECHNIQUES TO BUILD "*STaRS.sys*", A SOFTWARE SYSTEM DESIGNED TO SUPPORT THERAPISTS IN THE REHABILITATION OF ITALIAN ANOMIC PATIENTS.

After an introductory section reviewing classification, assessment, and remediation methods for naming disorders, we analyze the current trends in the exploitation of computers for the rehabilitation of language disorders. Starting from an analysis of the needs of speech therapists in their daily work with aphasic patients, the requirements for the STaRS.sys application are defined, and a number of possible uses identified.

To be able to implement these functionalities, STaRS.sys needs to be based on a lexical knowledge base encoding, in a explicit and computationally tractable way, at least the kind of semantic knowledge contained in the so called "*feature norms*". As a backbone for the development of this semantic resource we chose to exploit the Italian MultiWordNet lexicon derived from the original Princeton WordNet. We show that the WordNet model is relatively well suited for our needs, but that an extension of its semantic model is nevertheless needed.

Starting from the assumption that the kinds composing the feature types classifications exploited for encoding feature norms can be mapped onto semantic relations in a WordNet-like semantic network, we identified a set of 25 semantic relations (~feature types) that can cover all the information contained in these datasets.

To demonstrate the feasibility of our proposal, we first asked to a group of therapists to use our feature types classification for classifying a set of 300 features. The analysis of the inter-coder agreement shows that the proposed classification can be used in a reliable way by speech therapists.

Subsequently, we collected a new set of Italian feature norms for 50 concrete concepts and analyze the issues raised by the attempt to encode them into a version of MultiWordNet extended to include the new set of relations. This analysis shows that, in addition to extending the relation set, a number of further modifications are needed, for instance to be able to encode negation, quantifications or the strength of a relation. Information that, we will show, isn't well represented in the existing feature norms either.

After defining an extended version of MultiWordNet (sMWN), suitable to encode the information contained in feature norms, we deal with the issue of automatic extraction of such semantic information from corpora. We applied to an Italian a corpus state of the art machine-learning-based method for the extraction of common-sense conceptual knowledge from corpora, previously applied to English. We tried a number of modifications and extensions of the original algorithm, with the aim of improving its accuracy. Results and limitations are presented and analyzed, and possible future improvement discussed.

Contents

List of figures

List of Tables

1

Introduction

Identifying the nature and the content of conceptual knowledge is a central issue for the fields of Natural Language Processing (NLP) and Psycholinguistics. This common interest is declined in two different approaches. The main question for Psycholinguistics is: *what does it mean for a person to know a concept? How are concepts represented in the mind?* Whereas NLP is mainly interested in: *what kind of information needs to be encoded in a computer to represent the content of a concept? What is the most efficient way to encoding and exploiting that information?.* A number of works have tried to build a bridge between the two approaches. The best known project is probably the building of WordNet (Fellbaum, 1998b), but the literature is growing from both sides (e.g. Barbu, 2008; Andrews et al., 2009; Steyvers, 2010; Baroni et al., 2010; Kremer and Baroni, 2011; Kelly et al., 2012).

The work presented in these pages adheres to this bridging strategy. It has developed in the context the multidisciplinary project "Human Language Technologies as support for Language Disorders Therapy", that involved researchers from the Human Language Technology research unit at Fondazione Bruno Kessler, from

the Language Interaction and Computation (CLIC) group and from the Center for Neurocognitive Rehabilitation (CeRiN), both belonging to the Center for Mind/Brain Sciences of the University of Trento.

The aim of this project is twofold. From an *exploratory* point of view, its goal is to link recent advantages in two rather independent research areas such as Natural Language Processing and Neuropsychology. Specifically, the project tries to connect the facts investigated by the neuropsychological literature on category-specific semantic disorders[1] with recent advances in the computational commonsense knowledge representation area. As such, this research can be linked to those trying to develop a neuro-cognitively plausible computational model of the human conceptual knowledge, like LSA (Landauer and Dumais, 1997; Laham, 1997), the FUSS hypothesis (Vigliocco et al., 2004) or the Strudel model (Baroni et al., 2010).

On the other (in a sense, more *practical*) side, its outcome is STaRS.sys (*Semantic Task Rehabilitation Support system*), a tool for supporting the work of the therapist in the treatment of naming disorders (Nickels, 2002; Raymer and Rothi, 2002; Horton and Byng, 2002; Springer, 2008). A major difficulty in developing technological aids for aphasic patients is the need to create tools that are able to cope with the great variability of their impairment. Such a flexibility, and that's been our bet, can be achieved only by leading on cognitively motivated models.

Our work developed following this direction, i.e. trying to get insights form cognitive psychology in order to create a cognitively motivated semantic resources for STaRS.sys. We designed this resource by modifying the original WordNet model, a semantic resource built to be "*a dictionary based on psycholinguistic principles*" (Miller et al., 1990). The kind of semantic resource we have in mind is able to encode the kind of semantic information that can be represented by *Featural Descriptions* like <chair> has legs, that have been exploited in cognitive psychology as proxies of the human's semantic memory since the pioneering enquiries by Eleanor Rosch (e.g. Rosch and Mervis, 1975).

Our goals are ambitious, perhaps even unfeasible, given the current state of the art, both in psychology and in NLP. What we really wanted to do with this project, then, is to draw a direction that, we believe, has to be followed to build semantic resources aimed at somehow representing what is in the mind of a speaker.

---

[1]That is, disruptions of the semantic knowledge that selectively (or disproportionately) affect some semantic categories.

This thesis exploits methodologies, hypothesis and ideas from fields as distant as Natural Language Processing and Psychology. As a consequence, some terminological clash was expected. To avoid confusing the reader, we tried to avoid to use ambiguous technical terms, by preferring synonymous terms, whenever possible, or by creating new terms.

This has been quite painful for the main notion of this work, i.e. that of *"feature"*. In cognitive psychology, its meaning can be roughly paraphrased as that of *"concept property"*. In Natural Language Processing, and especially in the Machine Learning field, its meaning is more similar to that of *"attribute"* or *"variable"* of a model. In this thesis, we chose to use the term *"feature"* in its NLP sense. We will instead refer to its psychological meaning through the notion of *"Featural Description"*, that is of a concept-description pair of the form `<cat> is a feline` or `<dog> barks`[2]. We will however refer to the psychological notion of *"feature"* in technical compounds like *"feature generation"* or *"feature norms"*.

1.1 Plan of the Thesis

This thesis is organized in chapters as follows.

Ch 1. The current chapter, **Introduction**, introduces the novel aspects of the thesis and outlines its structure.

Ch 2. **Clinical Practice for Naming Disorders** introduces the reader to the literature on aphasia and reviews the most common methods for anomia assessment and rehabilitation.

Ch 3. **STaRS.sys in a Therapeutic Environment** discusses the role of computers in aphasia rehabilitation and depicts three possible ways of using STaRS.sys for the preparation of a semantic task.

---

[2]Throughout this thesis, concepts and description will be printed in `typewriter` font. When reporting a concept-description pair the concept will be further enclosed by `<angled brackets>`. WordNet synsets will be printed between `{curly brackets}`. Feature types and relations will be reported in *italics*, while concept categories and feature type in Small Capitals.

Ch 4. **building the STaRS.sys Lexical Database** presents the requirements that the STaRS.sys knowledge base has to meet and discusses why the WordNer model is the one that best fit our needs.

Ch 5. **a novel Feature Type Classification** introduces and evaluates a new classification of the kinds of semantic informations that can be associated to a concrete concept in a feature generation task.

Ch 6. **a WN-encoded set of Feature Norms** describes a feature generation experiment aimed at collecting Italian Featural Description to encoded into a dedicated Italian wordnet. Modifications to the original wordnet model are discussed and the outcome of their implementation is analyzed.

Ch 7. **the Automatic Extraction of Featural Descriptions** investigates the usability of a current State of the Art automatic method for knowledge extraction to automatically enrich wordnet with feature-like commonsense knowledge.

Ch 8. **Conclusions** summarizes and criticizes the thesis.

Part of the work reported in chapters 3, 4, 5 and 6 has been previously published in the following articles: Lebani and Pianta (2010a), Lebani and Pianta (2010b), Lebani and Pianta (2010c) and Lebani and Pianta (2012).

2
Clinical Practice for Naming Disorders

Aphasia is an acquired language disorder due to a brain damage. It is better thought as a syndrome, rather than a disease, that can occur as a consequence of a wide range of injuries and pathologies. It is strongly associated with stroke (up to 85% of aphasic patients suffered from a stroke), and it is a common consequence of a cerebrovascular accident; it has been reported a prevalence of aphasia for 21-38% of the stroke patients in the acute phase, see Brust et al. (1976); Wade et al. (1986); Pedersen et al. (1995). Nevertheless, other frequent causes of aphasia are traumatic head injuries, tumors, dementia and brain infections.

Different kinds of aphasia have been identified, depending on the pattern of linguistic difficulties experienced by such patients. Following Goodglass and Wingfield (1997), probably the most pervasive and persistent problem is anomia, that is *"a difficulty in finding high information words, both in fluent discourse and when called upon to identify an object of action by name"*. Naming difficulties, both in production and comprehension, are indeed a symptom that is common to virtually all aphasia types reported by the literature.

2.1   Semantic and Phonological Impairments

Aphasic patients can produce different patterns of naming errors. Together with other behavioural, psycholinguistic and neurolinguistic data, such variability has been interpreted as pointing to the existence of different processes involved in the functional architecture of the lexicon. In such a context, a core distinction that is shared by all theoretical models is that between semantic and phonological processes[1]. Naming problems showed by aphasic subjects can be accounted for as arising from either a mainly phonological or a mainly semantic breakdown, although pure impairment cases, in which patients make only one type of error, are rare.

The diverse symptoms manifested by patient DM (Caramazza et al., 2000) and patient DP (Cuetos et al., 2000), both fluent aphasics[2], can clearly illustrate such an opposition. In his general neuropsychological examination, patient DM showed difficulties in repeating a linguistic input (of any kind: single words, nonwords, phrases and sentences), in namimg object pictures, in reading aloud and in writing by dictation. Caramazza and Hillis described his errors as consisting mainly of phonemic substitutions (e.g. spella (nonword) instead of stella (*"star"*)).

When asked to name 314 pictured objects, DM didn't produce any semantic error, and the 115 errors were either nonwords (the vast majority: 95 cases), formal errors (5 cases - e.g. nuota (*"he swims"*) for suora (*"nun"*)) or no responses (15 cases). An analysis of nonword errors showed that, again, the vast majority of them involved phonological substitution. At the same time, such performance strongly suggests an unimpaired semantic and grammatical processing, in that, stated with thse two scholars, *"it is astronomically unlikely that DM's pattern of phonological substitutions could have been obtained if lexical entries had been selected incorrectly"*.

The reverse pattern of performance is showed by patient DP (Cuetos et al., 2000).

---

[1]For the sake of completeness, it should be mentioned that such models agre on the existence of a third broad functional component: the orthographic one. The early recognized (see Frith, 1980) opposition between orthographic and phonological processes is demonstrated by the existence of different kind of modality specific disorders. As an example, patient RGB (Caramazza and Hillis, 1990) produced semantic errors only in naming and reading tasks, but not in writing and word comprehension tasks, thus suggesting an impairment preserving the orthographic lexicon. On the other side, a subject like SJD (Caramazza and Hillis, 1991) showed a severe impairment in writing verbs (but not nouns) that she could easily pronounce, thus suggesting a selective impairment to this functional module. Anyway, under the hypothesis that the critical stages for word retrieval in conversation are the semantic and the phonological ones, we won't pursue this issue any more.

[2]i.e. patients whose speech is well articulated and grammatical, but semantically inappropriate

In the standard evaluation of his language skills, he performed poorly in the confrontation naming, naming definition and fluency naming tasks. His problems, furthermore, were restricted to naming, in that his reading, writing and comprehension abilities appeared preserved.

When asked to name 420 pictured objects (the test set was made of 140 pictures, his task being to name the full set for 3 sessions), DP gave the correct response for 291 items (69,3%). The vast majority (74 cases) of the 129 errors, furthermore, consisted of semantic substitutions, such as `sun` in response to a picture of the `moon`, and `panther` and `lion` for a picture of a `tiger`. Of the remaining errors, one has been analyzed as a nonword response, two as unrelated word responses, six as mixed responses, one as a formal error and forty-five as others, i.e. no-responses (25), descriptions/circumlocutions (6), visual errors (8), morphological errors (4) and perseverations (2).

Strikingly, while the general performance of DP improved across sessions, the only errors that did not decrease were the semantic ones. Consistently, then, Cuetos and colleagues interpreted DP's breakdown as a selective impairment affecting the semantic processing system.

Kay and Ellis (1987) have been the first to propose criteria for distinguishing between phonological-based and semantic-based anomias. The authors identified the following two triplets of associated behaviours:

- SEMANTIC-BASED ANOMIA

    - failure in semantic categorization tasks;

    - a strong effect of correct phonemic cueing (and an increase in semantic paraphasias after a false phonemic cueing);

    - no "tip-of-the-tongue" phenomenon.

- PHONOLOGICAL-BASED ANOMIA

    - good performance in semantic categorization tasks;

    - no effect of phonemic cueing (at most a very weak effect of the solely correct phonemic cueing);

    - "tip-of-the-tongue" phenomenon (at least sometimes).

Such an approach, however, has to face problems such as the existence of category-specific aphasias and of different production and comprehension impairment degrees (for a review, see Semenza, 1999).

2.2   NAMING DISORDERS ASSESSMENT

In the everyday clinical practice, a variety of methods are exploited for the diagnosis and assessment of naming disorders. Turgeon and Macoir (2008) identify two main approaches. On one side, the clinical-neuroanatomical ones rely primarily on the clinical observation in identifying the relevant symptoms and possible neuroanatomical triggers of the disease.

On the other hand, psycholinguistic approaches are based on cognitive models, whose main purpose is the identification of the different processes involved in language production and comprehension, rather than the classification of the associated symptoms. The above mentioned distinction between semantic, phonological and orthographic processes can be taken as a clear example of the rationale behind functional models.

For a precise characterization of naming disorders, psycholinguistic approaches appear to be much more informative than the clinical-neuropsychological ones. While, indeed, the latter mainly allow the clinician to identify the clinical population to which a patient can be assigned, adopting the former point of view in assessment allows for the identification of both his/her impaired and spared communicative (lexical, in our case) abilities.

Psycholinguistic assessment is usually made through specific tasks and test batteries. One of such batteries consist of the 60 controlled tasks that form the Psycholinguistic Assessment of Language Processing in Aphasia (PALPA: Kay, 1992), whose flexibility is well suited for an effective differentiation of the impaired and preserved abilities[3]. Batteries of solely lexical tasks for experimental purposes has been developed as well (e.g. the Florida Semantic Battery: Raymer and Rothi (2000)).

Following Raymer and Rothi (2002), at least three factors have a critical weight in the psycholinguistic characterization of the naming difficulties experienced by a

---

[3]In their review of the Semantic Therapy practices, Horton and Byng (2002) report a marked preference for such battery (63%) in the set of formal assessments they considered. However, many authors have remarked its lack of standardization (e.g. Raymer and Rothi, 2002; Horton and Byng, 2002; Turgeon and Macoir, 2008).

patient:

- CROSS-MODALITY COMPARISONS. As it is common practice to test a patient by assigning him/her tasks that exploit mostly one processing system (e.g. categorization vs. picture naming) or stage (e.g. categorization vs. repetition). In such tasks, it is crucial to vary the input (e.g. written vs. spoken words) and output modalities (e.g. gesture vs. repeating).

- CHOICE OF THE LEXICAL ITEMS. Given the different sources of variability that can influence the performance of a patient, it is important to choose the correct lexical items to exploit. This is especially true for patterns of impairment like category-specific semantic disorders (Capitani et al., 2003) and grammatical category-specific deficits (Shapiro and Caramazza, 2003).

- ERROR ANALYSIS. Even if not sufficient per se, an analysis of the error patterns produced by a patient can provide useful clues about the nature of his/her impairment. An example can be what has been done in the above comparison of DM's phonological substitutions with DP's semantic errors.

2.3   AN OVERVIEW OF REMEDIATION METHODS

Moving to the every-day therapeutic practice, the most common (behavioural) approaches divide themselves into those trying to repair to an impairment and those trying to circumvent it (Nickels, 2002; Raymer and Rothi, 2002).

In the latter kind of strategies ("Substitutive", "Strategic", "Re-Organizational" or "Compensatory"), the patient is trained to use his/her spared communicative and cognitive mechanisms to accomplish his/her communicative goals. Roughly speaking, then, they work by training the patient to avoid the impairment. In the case of naming disorders, this could be done, e.g., by using the orthographic form of a word in order to retrieve its spoken form (i.e. letter-sound conversion cueing). Following Raymer and Rothi (2002), this approach can be very useful in acute and chronic stages of recovery, and its choice has to be based on a systematic lexical assessment of the spared and of the impaired skills of the patient.

However, patients showing a pattern of impairment that allows for the exploitation of such strategies are quite uncommon, so that the identification of effective "Restitutive" (also called "Facilitation", "Repair" or "Retaining") techniques is the

crucial issue for this branch of research. In this approach, the therapeutic intervention concentrates primarily on the damaged cognitive functions of the patient, in order to remediate to his/her impairment.

Therapies for naming disorders are commonly characterized as mainly phonological or semantic. As pointed out by Nickels, however, such classification is misleading, if not further specified. Indeed, in labelling a therapy as simply semantic or phonological we introduce a level of ambiguity between the type of tasks exploited and the type of impairment addressed (Nickels, 2002).

Following her review, two "semantic vs. phonological" oppositions have been adopted in the literature. The first is the one between "therapies for semantic impairment", that tries to remediate to semantic-based anomias, and "therapies for phonological impairment", that address phonological-based anomias.

Nowadays it's common practice, in talking about semantic and phonological therapy, to refer to the other distinction, that is, the one between semantic and phonological tasks as therapy. Here, therapies for naming disorders are classified according to the inner nature of the tasks they exploit.

Given that lexical processing is a complex task, such an opposition is somehow an approximation. That is, no pure semantic or phonological therapy can exist, as both tasks entail an unbalanced mixture of both kinds of processing. Clearly phonological tasks, say oral word reading or word repetition, always entails some form of semantic comprehension. There is no way to repeat a word without understanding it, at least for an unimpaired speaker. The other way round is equally true. How to sort written or spoken words without accessing to their phonological representations? The bulk of the semantic-phonological therapy opposition, then, is the kind of processing that is addressed to a greater extent.

Furthermore, also the opposition between the two kinds of anomias themselves should be thought as the maximum spread of a continuum, in that patients usually show mixed patterns of impairment, rather than pure disorders.

As a consequence, we should expect both kinds of patients to benefit from both semantic and phonological tasks. Actually, even if issues related to the relative impact of each therapeutic approach are still open, such phenomenon has been well documented in the literature (for a review see Nickels, 2002; Raymer and Rothi, 2002; Springer, 2008, inter alia).

2.4 THE DAY-TO-DAY SEMANTIC THERAPEUTIC PRACTICE

In exploiting tasks that address the semantic knowledge of a patient, semantic therapies try to tap into the semantic context of a word, in order to activate its meaning. Typically, treatments of this kind make use of metalinguistic tasks like semantic judgments, description of word meanings, identification of semantic categories and property generation (Springer, 2008).

Describing some of the exercises that compose the lexical semantic therapy programme BOX (Visch-Brink et al., 1997; Doesborgh et al., 2004) can be useful for explanatory purposes. In the Semantic Category task a number of semantically related words (e.g. `letter`, `postcard` and `bill`), plus one belonging to a different category (e.g. `cigar`), are presented to the patient, who is requested to pick the odd one out. In the Semantic Gradation task, the patient is requested to match words (e.g. `chestnut`, `Easter`, `harvest-time`, `blossom`) with one of two antonyms (e.g. `Spring` or `Autumn`); while in the Syntagmatic and Paradigmatic Relationship task the patient is asked to choose, out of a group of two or three words (e.g. `actor`, `translator`, `courier`), the one that is syntagmatically or paradigmatically related to a probe (e.g. `interpreter`). Notably, such exercises usually articulate along different levels of difficulty, that can be increased by modulating either the semantic distance of the distractors, either the distance or lexico-semantic relation between the probe and the target.

Moving to the everyday clinical practice, the utility of standardized therapies is somehow set against the need to adapt the treatment to the subjects' needs. As an example, consider the case of patients showing similar, but not identical, patterns of category-specific difficulties, such as MD (Hart et al., 1985) and HJA (Riddoch and Humphreys, 1987). Both suffered from a stroke resulting in a *"primary biological categories impairment"*. Nevertheless, their lexical difficulties overlap only partially: while patient MD's impairment affects only FRUIT AND VEGETABLES (sparing other categories such as ANIMALS and FOOD), HJA's impairment involves also ANIMALS and BODY PARTS (Capitani et al., 2003). To be effective, therapeutic aids have to fit the difficulties of these patients .

In their review, Horton and Byng (2002) identified twelve main kinds of semantic therapy[4] tasks (further grouped by type) exploited in the literature:

---

[4]These authors adopted a definition of semantic therapy that is broader than ours. Some of their

- Judgment Tasks

  - category sorting;

  - odd-one-out;

  - word-picture matching;

  - verb/sentence semantics;

  - choose written word (spoken definition);

  - choose written word from spoken word;

  - yes/no attributes question answering (picture stimuli).

- Production Tasks

  - spoken verb production;

  - clause production;

  - spoken "semantic information";

  - words with picture stimuli.

Preparing most of these semantic tasks (e.g. semantic questionnaires[5], but also category sorting, odd-one-out etc.) often requires the therapist to fill out by hand a list of concept-attribute pairs of the kind exemplified in Appendix A.1, illustrating a sample of some features (also known as concepts features, features, featural descriptions) used in the CIMeC Center for Neurocognitive Rehabilitation (CeRiN).

In compiling such lists, therapists perform an activity similar to the one carried out by subjects participating to a property generation task (for a review, see Murphy, 2002), although with a different degree of know-how. In both cases, featural representations are useful because they provide a window into the semantic memory of the patient/subject, rather than an exact description of their knowledge (for a similar point, see Cree and McRae, 2003).

Several works proved that various measures derived from featural descriptions (e.g. feature distinctiveness, semantic relevance, concept similarity, feature correlation) and different feature type categorizations can be able (at least) to account for

---

tasks, indeed, *"form part of a therapy [...] primarily aimed at addressing other systems"*. However, given our explanatory purposes, we report the whole set of tasks they reviewed.

[5]A task in which the patient has to judge whether a concept-attribute pair is true (e.g. `<il cammello> vive nel deserto` (*"<the camel> lives in the desert"*)) or false (e.g. `<il cavallo> è testardo` (*"<the horse> is stubborn"*))

the various patterns of category-specific semantic deficits (McRae and Cree, 2002; Cree and McRae, 2003; Vinson et al., 2003; Sartori and Lombardi, 2004).

In the CeRiN norms, apart from the concept-description pairs themselves, three kinds of semantic information are available. First, a flat taxonomical organization of concepts into semantic classes. Our sample is made of concepts belonging to only four categories: ANIMAL, CLOTHING, FOOD and TOOL.

Second, a broad classification of description types into *visual*, e.g. `ha la pelle rosea` (*"its skin is pink"*) and *nonvisual*, e.g. `fa bene agli occhi` (*"improves eyesight"*) features. Third, attributes are further classified as instances of one of the following six types: *color*, e.g. `ha un mantello marrone` (*"has a brown coat"*); *dimension*, e.g. `è un animale piccolo` (*"is a small animal"*); *matter*, e.g. `è di spugna` (*"is made of terry cloth"*); *morphology*, e.g. `ha otto zampe` (*"has eight legs"*); *natural environment*, e.g. `vive in Australia` (*"lives in Australia"*); *taxonomic category*, e.g. `è un utensile` (*"is a tool"*); *function*, e.g. `si usa per conservare il cibo` (*"is used to preserve food"*); *other encyclopaedic information*, e.g. `si indossa sotto i pantaloni` (*"is worn under the trousers"*). Given the different rationales behind therapeutic practice and experimental collection of norms, it's not surprising that the CeRiN classification is notably different from all the others proposed in the psychological literature (e.g. Garrard et al., 2001; Cree and McRae, 2003; Vinson and Vigliocco, 2008; Wu and Barsalou, 2009).

Semantic features and derived measure, however, cannot account for the whole range of variability observed in the performance of impaired and unimpaired speakers. Other possible dimensions of variation have been proved to be word frequency, word familiarity, age of acquisition, grapheme regularity, morphological complexity, abstractness, visual complexity and word length (Laiacona et al., 1993b; Raymer and Rothi, 2002; McRae and Cree, 2002; Springer, 2008).

The preparation of a therapeutic task is a complex and time-consuming work, that cannot be fully standardized because of he great variability of impairment showed by the aphasic patients, typically performed manually by the therapist that is in charge of controlling, when possible, for many different variables. In such a context, it is natural to ask to what extent and how modern computers can help the therapist by taking charge of part of the manual works or even by enhancing his/her work by providing new possibilities.

3

STaRS.sys in a Therapeutic Environment

Computers can support the rehabilitation of language disorders in many ways: from enhancing assessment to assisting administrative management, from helping the clinician during the therapeutic session to alleviating the communicative difficulties of a patient by exploiting his/her unimpaired abilities (Petheram, 2004; Petheram and Enderby, 2008; van de Sandt-Koenderman, 2011).

General characteristics of computers like vast data storage and retrieval capability, connectivity and ergonomics allows for software applications flexible enough to be adapted to the peculiar needs of each patient. The same characteristics allows for the collection of longitudinal data that can give a more comprehensive description of the patient's abilities and of the treatment evolution. They also can free therapists from many low-level tasks like analyzing scores and submitting repetitive tasks, thus allowing for a more effective therapy (see Petheram and Enderby, 2008).

Electronic devices and software systems built (or converted) for the rehabilitation of aphasic patients can be divided in two broad families: those providing some therapeutic rehabilitation, and those trying to compensate for the patient's loss of

communicative skills. Even if the pioneering studies in this field date back to approximately thirty years ago, (e.g. Colby et al., 1981; Katz and Nagy, 1982; Lincoln et al., 1984; Johannsen-Horbach et al., 1985), most of the literature focused on therapy (van de Sandt-Koenderman, 2004). The present work is no exception. However, as technologies develop in unforeseen ways, and as treatment practices evolve, new possibilities and new needs do emerge. As an example, it is easy to see the spreading of portable devices like tablets and smart phones, together with the facility of creating "apps" for such devices, as a great opportunity to solve some of the problems that slowed the growth of communication aids for aphasia (see van de Sandt-Koenderman, 2004).

3.1 HLTs as Therapeutic Devices

Following Petheram and Enderby (2008), we can divide therapeutic approaches exploiting electronic devices into Computer Only Therapy (COT) and Computer Assisted Therapy (CAT). COT systems allow the patient to practice without the presence of the therapist, while CAT systems are developed for improving the quality of the treatment offered to the patient, as can be the exploitation of multimodal multimedia materials or virtual reality (Wallesch and Johannsen-Horbach, 2004).

In the clinical practice, however, these two approaches often overlaps. An example can be the already cited semantic therapy program BOX (Doesborgh et al., 2004), that is supplied both as a paper and pencil version to be used with the therapist and as an electronic version to be used at home. This example is useful for illustrating one of the key requirements that therapeutic systems must meet: psychological plausibility. Given the vulnerability of the aphasic population, therapeutic exercises and materials automatically supplied to them have to be based on theoretical models and principles (Petheram and Enderby, 2008), and a goal that this line of research has to reach is the electronic implementation of the treatment practices whose efficiency has been well-proved (van de Sandt-Koenderman, 2011).

Another example concerns what probably is the most long-living of these systems: Lingraphica® (Katz, 2009), the commercial version of C-VIC (Computerized Visual Communication: Steele et al. (1989)) one of the earliest multimedia programs developed for aphasic persons. Such a system has been developed as a comprehensive tool that can be used both as a communication device (see 3.2) and as a therapeutic

tool by supplying a formal training program with clinical exercises for categorization, naming, spelling etc.

In an extensive study with 60 patients, Aftonomos et al. (1999) used Lingraphica® for evaluating the efficacy of community-based, real-life, aphasia therapy programs. In addition to the standard one-hour-per-week therapeutic session using materials from Lingraphica®, their patients were prescribed a two-hours per day home practice with the same system. The significant improvement showed by these subjects illustrates a key point of using computers in therapy: they are the only feasible way to reach the minimum therapeutic intensity that, as discussed by van de Sandt-Koenderman (2011), cannot be met in the standard clinical practice.

3.2   High Tech AACs

The most urgent need of the aphasic patient, however, is to communicate. Prior to computers, aphasics could benefit from some low-tech AAC (Alternative and Augmentative Communication) strategies like writing, drawing or pointing to words, pictures or images in communication books or communication boards. Low- tech AAC strategies, however, didn't become very popular for many reasons (extensively discussed by van de Sandt-Koenderman, 2004), among which is the lack of training, the scarce flexibility of these systems and their slowness in a real life scenario.

These limitation can be partly overcome by exploiting high-tech AAC devices developed for helping word retrieval, sentence construction or conversation tout-court. Flexibility and the possibility to easily handle different kinds of multimedia information are probably the key advantage of using such tools.

Well known systems of this kind are the already cited Lingraphica® and the EU project PCAD (Portable Communication Assistant for People with Acquired Dysphasia: Wahn (2002)), that is evolved into the system Touchspeak™. Such devices can be indeed used by an aphasic patient as multimedial pointing board exploitable for constructing sentences. As an examples, the AllTalk™ and TouchTalk™ Lingraphica® devices are built on a vocabulary of images, animations, sounds and texts. Such items (e.g. microwave) are presented to the user in a familiar environment (e.g. a kitchen), and can be combined in a storyboard to construct a phrase.

A serious problem common to all such devices, however, is that the hierarchical organization of their vocabularies is difficult to navigate. The project ViVA (Vi-

sual Vocabulary for Aphasia: Nikolova et al. (2009a,b)) tries to address this issue by building a system whose vocabulary items are connected by the *evocation* relation, that is, how a concept evoke others (Boyd-Graber et al., 2006). In details, this vocabulary has been built moving from the original WordNet (Fellbaum, 1998b) concepts encoded by Boyd-Graber et al. (2006) and a subset of the Lingraphica® vocabulary, and successively extended by exploiting the Amazon Mechanical Turk (Snow et al., 2008) annotators (for details, see Nikolova et al., 2009b).

Being based on such an associative vocabulary, the final user can interact with ViVA in two different ways. In the "user preference module" it is possible to organize the vocabulary in a customized way, so that it is possible to organize concepts in ad-hoc categories, to associate sentences with them, to add and remove concepts and so forth. The goal of the "active learning module" is to organize the vocabulary on the basis of the user's past interactions with the system, of his/her preferences, and of the semantic associations encoded in the vocabulary. Given an input such as `doctor`, then, the ViVA "active learning module" allows the system to suggest related or previously used concepts like `medication`, `dentist` or `pain`.

Finally, it should be pointed out that the communication needs of aphasic patients can benefit not only from the use of ad-hoc created special devices, but also from a different use of popular softwares like PowerPoint or Internet technologies like blogs, chats, e-mails and forums or dedicated Portals (see Egan et al., 2004; Kitzing et al., 2005).

Notwithstanding these positive characteristics, even High-tech AAC systems are not widespread among aphasics. Just a minority of these patients, indeed, actually uses them in the everyday life. As pointed out by van de Sandt-Koenderman (2011), a reason could be the limited immediate benefits of using them in a real communication environment, also due to the fact that their use makes the communication too slow and frustrating. This limitation, however, is counterbalanced by the improvement of the communicative skills due to the indirect linguistic training that follows from the use of an AAC system.

To our knowledge, however, no existing software has been designed for selecting therapeutic stimuli by controlling efficiently the most important variables that can affect the performance of their patients. In particular, nothing similar exists for Italian. In this thesis, we introduce STaRS.sys (*Semantic Task Rehabilitation Support system*), a software system for this purposes.

3.3 STaRS.sys: Use Case Scenario

STaRS.sys is an Italian tool thought to assist the therapist in the preparation of a semantic task. In the framework depicted in the first part of this chapter, than, we can describe it as a CAT tool. Our system can be exploited by a therapist (1) to retrieve concepts possessing certain properties, (2) to retrieve semantic information associated to concepts or (3) to compare concepts. This led us to the identification of the three possible use cases sketched out in the following pages, each characterized by a possible need or class of needs (e.g. *"to find a concept such as..."*), a typical interaction occurring between the user and the system. Such a connection is highlighted in the use case stories below, all of which tell about a fictional therapist (let's call him EP) preparing some specific semantic task for a patient (gL) with a naming deficit selectively affecting animal concepts.

Furthermore, every scenario is an exemplification of the prototypical therapeutic use of one of the three main functionalities of STaRS.sys: the "Find Concept", the "Describe Concept" and the "Compare Concept" one. These functions are directly available in the top level interface to the user. Alternatively, the "Describe Concept" and "Compare Concept" functionalities can be accessed from the output of the "Find Concept" section, so as to take as input the concepts the system found from previous user's specifications.

3.3.1 UCS1 - Get Concept from Specifications

In a first scenario[1], the goal of the user is to find concepts that match some specifications. In the case of our therapist EP, this is very useful for controlling for some of the variables that can affect gL's performance in the selection of the stimuli for a feature generation task, that is, a task in which the patient is required to generate featural descriptions for a set of concepts.

---

[1]Simple queries are enclosed in [square brackets]. Two joining operator are defined: the ampersand is used when both values are referred to the target concept, while the WITH operator is used when one value is a specification of the other. Complex queries, on the other hand, are obtained by joining simple queries through the Boolean operators **AND**, **OR**, **XOR** and **NOT**. In the third scenario, we adopt the notation "comp(*)" (where * is the specification label) to mark the specification label that will be used as term of similarity comparison: e.g. [comp(color) = red] if concepts are compared for their redness, [comp(semantic category) = PREDATOR] if concepts are compared for their being or not predators.

Use Case: Preparing a Feature Generation Task	

Primary Actor: the therapist (EP)

Interests: to find words that meet some intended characteristics

Needs: the retrieval process has to be quick and accurate

many possible concepts have to be proposed to the therapist

variables to be controlled have to be handled carefully

Time Span: before or during the therapeutic session

Interactions

Step	Action
1	**Before the therapeutic session**. Therapist EP starts the organization of the next gL's therapeutic session by reading the patient's case history: gL's rehabilitation program schedules, among the others, a feature generation task. Given gL's disease, the stimuli that have to be presented to the patient should be concepts such that: 1. they belong to the ANIMAL semantic class; 2. they are highly frequent; 3. they are associated to highly distinctive color features; 4. they have a high mean feature distinctiveness.
2	To collect a set of stimuli to submit to the patient, EP opens up STaRS.sys and selects the "Find Concept" option in the initial panel three item-menu
3	In the "Find Concept" modality, EP submits STaRS.sys the complex query: 1. [semantic category = ANIMAL] **AND** 2. [frequency = "high"] **AND** 3. [color = "*" WITH relevance = "high"] **AND** 4. [mean feature distinctiveness = "high"]
4	STaRS.sys elaborates EP's request and supplies a set of concepts matching the search criteria
5	From the STaRS.sys output, EP selects and prints five concepts: zebra (*"zebra"*), orso polare (*"polar bear"*), tigre (*"tiger"*), leopardo (*"leopard"*) and mucca (*"cow"*).
6	**Therapeutic session**. During the therapeutic session, EP submits gL with the list of five concepts selected from the STaRS.sys output
7	Patient gL tries to describe the five concepts supplied by EP

Extensions

By using the "Find Concept" modality, EP is allowed to choose among a range of possible specifications or variables to control. In details, the possible choices can be:

- given values for features: e.g. [color = red];
- values of prototypicality for given semantic categories: e.g. [semantic category = FURNITURE & prototypicality = "high"];
- values of distinctiveness (Devlin et al., 1998) for given features or feature types: e.g. [color = red WITH distinctiveness = "high"] for the feature red, [feature type = *color* WITH distinctiveness = "high"] for the type *color*;
- values of mean feature distinctiveness (Cree and McRae, 2003): e.g. [mean feature distinctiveness = "high"] ;
- values of semantic relevance (Sartori and Lombardi, 2004) for given features: e.g. [color = red WITH relevance = "high"];
- values of frequency: e.g. [frequency = "high"];
- any meaningful combination of these.

recap
S1. the therapist specifies a combination of concept specifications and submits a query to the system   S2. the system retrieves all concepts matching the specifications   S3. the therapist refines his research   S4. the system prints the concepts that match the search terms   S5. the therapist asks the patient to generate short descriptions of (a selection of) the concepts found by the system

3.3.2   UCS2 - GET FEATURES FOR A CONCEPTS

In a second scenario, STaRS.sys is exploited to retrieve featural descriptions for some given concepts. Such information is useful in preparing therapeutic tasks like semantic questionnaires, in which a patient is required to mark as true or false a list of concept-description pairings of the kind reported in Appendix A.1[2].

---

[2]Actually, the CeRiN set in Appendix A.1 is a reduced version of the original list used for preparing a similar task, from which the wrong pairing have been removed.

Use Case: Preparing a semantic questionnaire	

Primary Actor: the therapist (EP)

Interests: to find descriptions for a given set of concepts

Needs: the retrieval process has to be quick and accurate

       many kinds of description should be available to the therapist

Time Span: before or during the therapeutic session

Interactions	
Step	**Action**
1	**Before the therapeutic session**. Therapist EP starts the organization of the next gL's therapeutic session by reading the patient's case history: gL's rehabilitation program schedules, among the others, a semantic questionnaire.
2	Given gL's disease, the therapist fills by hand a list of concepts to be used in the task. Such a list includes the concepts `leopardo` (*"leopard"*), `cammello` (*"camel"*), `giraffa` (*"giraffe"*), `canguro` (*"kangaroo"*).
3	EP decides to look for highly relevant taxonomical or perceptual descriptions of the chosen concepts.
4	To find the intended specifications, EP opens up STaRS.sys and selects the "Describe Concept" option in the initial panel three item-menu.
5	In the "Describe Concept" modality, EP submits the concept and the relevant complex query, that is: 1. [relevance = "high"] **AND** 2. [feature type = *color*] **XOR** [feature type = *isA*]
6	STaRS.sys elaborates EP's request and supplies a set of features matching the search criteria.
7	From the STaRS.sys output, EP selects two attributes to be paired with the concept `leopardo`. These are: è giallo con macchie nere (*"is yellow with black spots"*), and è un felino (*"is a feline"*).
8	EP reiterates steps 5-7 for each of the other chosen concepts (i.e. `cammello`, `giraffa` and `canguro`), thus obtaining the following concept-attribute pairs: <cammello> ha due gobbe sulla schiena (*"<camel> has two humps on its back"*), <giraffa> ha un collo lungo e sottile (*"<giraffe> has a long thin neck"*), <canguro> è un marsupiale (*"<kangaroo> is a marsupial"*).

Continued on next page

9	EP prints the five concept-attribute pairs selected from STarS.sys output and fills out by hand the following list of five false concept-attribute pairs: <leopardo> vive nel mare *("<leopard> lives in the sea")*, <giraffa> è un felino *("<giraffe> is a feline")*,<cammello> ha la barba *("<camel> has a beard")*, <leopardo> ha un corno *("<leopard> has a corn")*, <canguro> è un uccello *("<kangaroo> is a bird")*.
10	**Therapeutic session**. During the therapeutic session, EP submits gL with a list of ten concept-attribute pairs and asks him to judge their correctness. These pairs are partly hand crafted by him, partly obtained by exploiting STaRS.sys.
11	Patient gL tries to mark as true or false the ten pairs supplied by EP.

Variations

Step	Action
2	Given gL's disease, the therapist exploits STaRS.sys to identify a set of concepts to be used in the task. fills by hand a list of concepts to be used in the task (INCLUDE: steps 1-5 of UCS1). Such a list includes the concepts leopardo *("leopard")*, cammello *("camel")*, giraffa *("giraffe")*, canguro *("kangaroo")*.

Extensions

By using the "Describe Concept" modality, EP is allowed to choose among a range of possible semantic characteristics to look for or to control. In details, these can be:

- feature types: e.g. [feature type = *color*];
- values of frequency: e.g. [frequency = "high"];
- values of distinctiveness (Devlin et al., 1998): e.g. [distinctiveness = "high"];
- values of relevance (Sartori and Lombardi, 2004): e.g. [relevance = "high"];
- any meaningful combination of these.

recap|

> s1. the therapist specifies an input concept
>
> s2. the system retrieves all the possible semantic features for the input concept
>
> s3. the therapist selects the specifications
>
> s4. the system prints the concept-feature pairs matching the search terms
>
> s5. the therapist creates a list of concept-feature pairs composed by right (selected from the system output) and wrong (hand-crafted) pairings
>
> s6. the therapist asks the patient to mark the pairs of the list as true or false

3.3.3   UCS3 - GET CONCEPT FROM CONCEPT

In a third scenario, the key notion is that of similarity. From a computational point of view, the evaluation of the similarity between two concepts can lead to different results depending on the measure and on the nature of the semantic model adopted.

Many branches of research in NLP deal, more or less overtly, with the problem of finding a reliable measure for approximating human's ability to perceive similarity between concepts (for a review, see Jurafsky and Martin, 2009, chap 20). On the other hand, even in the psychological literature a precise characterization of how humans' similarity judgments work is matter of debate, and for long time the study of concepts itself has been tied up with the study of similarity (Murphy, 2002).

On the basis of (a) the kind of knowledge source (raw text or a semantic network) they use, (b) the nature of the similarity they compute (taxonomical, featural or associative), we can identify two broad families of measures:

Distributional-based Similarity measure. Following the idea that *"you shall know a word by the company it keeps"* (Firth, 1957), several methods for evaluating the distributional similarity of words have been proposed (for a review, see Mohammad and Hirst, 2012). Roughly speaking, the logic behind such measures is that the degree of similarity between words is a function of the number of the co-occurring words they share (Sahlgren, 2006; Turney et al., 2010). Notwithstanding such measures seem to be good approximation of the human's semantic memory (e.g. Lund and Burgess, 1996; Landauer and Dumais, 1997), the obtained similarity values (a) are difficult to interpret semantically, (b) are very sensitive to data-sparseness problems and (c) refer to words rather than to concepts. As a consequence, for the STaRS.sys project we chose to limit ourselves to the network-based model.

Network-based Similarity measure. A broad range of similarity measures exploiting the structure of semantic networks like WordNet (Fellbaum, 1998b; Miller et al., 1990) has been proposed in the literature (for a review, see Budanitsky and Hirst, 2001; Patwardhan et al., 2003; Budanitsky and Hirst, 2006). These measure can be further characterized on the basis of the nature of the similarity they compute. **Taxonomy-based** measures rely on the taxonomic structure of a network to compute the similarity of two concepts. According to Budanitsky and Hirst (2001) and to Patwardhan et al. (2003), the measure

of this kind performing better across a number of NLP tasks is the one by Jiang and Conrath (1997). Such measures can be exploited by the user working on concept categorization, or in any fashion interested in finding concepts that lay in the taxonomic neighbor of an input concept. The configuration will be quite straightforward, and the user will have to specify the semantic category of comparison and eventually an additional semantic category for restricting the search space. **Feature-based measures** evaluate the similarity between concepts from the amount of overlap in their featural descriptions, as it happens with the pattern-based relatedness measures proposed by Hirst and St-Onge (1998). In interacting with our tool, the user will have to indicate, in addiction to the eventual semantic category to investigate, the featural specifications on the basis of which the concepts similarity has to be evaluated.

Given the different rationales behind the several proposed in the literature, given the fact that their performance can vary notably from task to task and given the therapeutic purposes of our tool, we decided to implement a "Concept Similarity" functionality, that allows the user to select and configure both the similarity measure and the semantic space/network to exploit. For a given concept, the output of this functionality should be an ordered list of similar concepts, with a value of semantic similarity, whose meaning depends on the measure employed.

An additional option is available in the "Concept Similarity" output interface, that is the possibility to find concepts that are *dissimilar* from the input one, given the measure and parameters already specified in the first phase (i.e. in the search for similar concepts). Again, the therapist is allowed to browse and filter the group of concepts that have low scores of similarity with the input concept. The specific amount of semantic distance from the input concept will be set by the therapist through a graphical bar whose values range from "high" (for highly dissimilar concepts) to "slight" (for slightly dissimilar concepts), whose effective meaning is relative to the measure and parameters already specified.

The following two stories illustrate the utility of the "Concept Similarity" functionality in the preparation of a odd-one-out task, in which the patient is asked to select the incoherent concept out of a list of three. As already pointed out, according to the nature of the similarity exploited in the task, one measures can be more appropriate than the others. In the first story, a feature-based similarity method seems to be the best choice.

USE CASE: PREPARING AN ODD-ONE-OUT (FEATURE-BASED)	

Primary Actor: the therapist (EP)

Interests: to find concepts whose descriptions are similar to those associated to a given (set of) concept(s)

Needs: the retrieval process has to be quick and accurate

the therapist must be able to modulate the semantic distance

variables to be controlled have to be handled carefully

Time Span: before or during the therapeutic session

Interactions	
Step	**Action**
1	**Before the therapeutic session**. Therapist EP starts the organization of the next gL's therapeutic session by reading the patient's case history: gL's rehabilitation program schedules, among the others, odd-one-out task.
2	Given gL's disease, the therapist fills by hand a list of concepts to be used in the task. Such a list includes the concepts leone (*"lion"*), gazzella (*"gazelle"*), canarino (*"canary"*), luccio (*"pike"*) and other sixteen animals.
3	Given gL's disease, EP picks the concept leone from the list in order to build a triple composed by that concept and other two (1) animal concepts (2) living in a similar/a different natural habitat.
4	To find the other two concepts of the triple, EP opens up STaRS.sys and selects the "Concept Similarity" option in the initial panel three item-menu.
5	In the "Concept Similarity" modality, EP selects the feature-based similarity measure and submits both the probe concept (leone) and the complex query: 1. [semantic category = ANIMAL] **AND** 2. [^{comp}(feature type) = *location*]
6	STaRS.sys elaborates EP's request and supplies an ordered list of concepts referring to animals living in a similar natural habitat of leone. The highest positions includes the concepts leopardo (*"leopard"*) and ghepardo (*"cheetah"*).
7	By browsing the similarity output of STaRS.sys, EP chooses leopardo as the coherent element of the triple in preparation. He then select the "dissimilarity" option in the output interface.

8	STaRS.sys provides a list of (animal) concepts that are dissimilar from `leone`. This raking ranges from concepts like `foca` (*"seal"*, highly dissimilar) to concepts like `gorilla` (*"gorilla"*, slightly dissimilar).
9	EP chooses `gorilla` as the odd element of the triple, and prints the triple.
10	EP repeats steps 3-9 for the remaining nineteen concept of the initial list, so as to prepare a list of twenty odd-one-out trials.
11	**Therapeutic session**. During the therapeutic session, EP submits gL a list of twenty triples of animal concepts. Each triple is composed by one of the concepts chosen in step 2, one coherent concept obtained by exploiting the STaRS.sys similarity output and one dissimilar concept obtained through the dissimilarity output of the system.
12	Patient gL tries to find the odd element in each of the triples supplied by EP.

Variations

Step	Action
2	Given gL's disease, the therapist exploits STaRS.sys itself to identify the set of initial concepts starting from which similar concepts are further searched (INCLUDE: steps 1-5 of UCS1 - Get Concept from Specifications). Such a list includes the concepts `leone` (*"lion"*), `gazzella` (*"gazelle"*), `canarino` (*"canary"*), `luccio` (*"pike"*) and other sixteen animals.

Extensions

In the similarity and dissimilarity modalities (steps 7,9) EP can filter the results by exploiting the range of possible semantic characteristics to look for or to control:

- values of prototypicality for a given semantic category or for the semantic category of comparison: e.g. [semantic category = FURNITURE & prototypicality = "high"] for the category FURNITURE, [ᶜᵒᵐᴾ(semantic category) = PREDATOR & prototypicality = "high"] if the semantic class of comparison is PREDATOR;

- values of distinctiveness for given features, feature types or for the feature of comparison: e.g. [color = red WITH distinctiveness = "high"] for the feature red, [feature type = COLOR WITH distinctiveness = "high"] for the feature type COLOR, [ᶜᵒᵐᴾ(color) = red WITH distinctiveness = "high"] if the feature of comparison is red;

- values of mean feature distinctiveness: e.g. [mean feature distinctiveness = "high"] for the feature red,

- values of semantic relevance for given features or for the feature of comparison: e.g. [color = red WITH relevance = "high"] for the feature red, [comp(color) = red WITH relevance = "high"] if the feature of comparison is red;
- values of frequency: e.g. [frequency = "high"];
- any meaningful combination of these.

A taxonomy based similarity method is more useful when looking for triples of concepts whose similarity has to be evaluated on the basis of their position in the STaRS.sys *isA* hierarchy.

USE CASE: PREPARING AN ODD-ONE-OUT (TAXONOMY-BASED)
Primary Actor: the therapist (EP)
Interests, **Needs** and **Time Span**: as above
Interactions

Step	Action
1	**Before the therapeutic session**. Therapist EP starts the organization of the next gL's therapeutic session by reading the patient's case history: gL's rehabilitation program schedules, among the others, an odd-one-out task.
2	Given gL's disease, the therapist fills by hand a list of concepts to be used in the task. Such a list includes the concepts leone (*"lion"*), gazzella (*"gazelle"*), canarino (*"canary"*), luccio (*"pike"*) and other sixteen animals.
3	Given gL's disease, EP picks the concept leone from the list in order to build a triple composed by it and other two concepts referring to (1) living beings that are similar to leone (2) in their being or not predators.
4	To find the other two concepts of the triple, EP opens up STaRS.sys and selects the "Concept Similarity" option in the initial panel three item-menu.
5	In the "Concept Similarity" modality, EP selects the taxonomy-based similarity measure and submits both the probe concept (leone) and the complex query: 1. [semantic category = LIVING BEING] **AND** 2. [comp(semantic category) = PREDATOR]

6	STaRS.sys elaborates EP's request and supplies a set of concepts referring to living beings similar to `leone`, ordered by similarity. The highest positions includes concepts like `tigre` (*"tiger"*) and `giaguaro` (*"jaguar"*).
7	By browsing the similarity output of STaRS.sys, EP chooses `tigre` as the coherent element of the triple in preparation. He then selects the "dissimilarity" option in the system interface.
8	STaRS.sys provides a list of living beings that are dissimilar from `leone`. This raking ranges from slightly dissimilar concepts like `persico` (*"perch"*, to highly dissimilar concepts like `acero` (*"maple"*).
9	EP chooses `foca` as the odd element of the triple, and prints the triple.
10	EP repeats steps 3-9 for the remaining nineteen concept of the initial list, so as to prepare a list of twenty odd-one-out trials.
11	**Therapeutic session**. During the therapeutic session, EP submits gL a list of twenty triples of concepts. Each triple is composed by one of the concepts chosen in step 2, one coherent concept obtained by exploiting the STaRS.sys similarity output and one dissimilar concept obtained through the dissimilarity output of the system.
12	Patient gL tries to find the odd element in each of the triples supplied by the therapist EP.

recap

s1. the therapist specifies an input concept and a similarity measure

s2. the therapist further specifies the parameters of the similarity measure

s3. the system prints the input most similar concepts

s4. the therapist explores one of the two sets of dissimilar concepts or specifies the search parameters for the dissimilar concepts

s5. the system prints the dissimilar concepts it found

s6. the therapist creates a triple composed by: the input concept, one similar concept chosen from the list of similar concepts, one dissimilar concept chosen from the of dissimilar concepts

s7. the therapist asks the patient to select the incoherent (i.e. the dissimilar) concept from the triple

3.4 Another (possible) Context of Use

A possible extension of the basic STaRS.sys Use Cases concerns its possible usage in a research context. The kind of information exploited by the therapist, indeed, can be used also by a researcher for facing two well know problems affecting the research on category-specific semantic disorders. First of all, as pointed out by Capitani et al. (2003), many works in this tradition suffer from the lack of control for different nuisance variables. Even if the examples cited by these scholars, i.e. familiarity and visual complexity, clearly falls out the scope STaRS.sys, our tool allows for the control of many other equally relevant variables, such as prototypicality and frequency.

A second problem can be charged to the habit of testing patients by choosing a small sample of concepts from a large domain (McRae and Cree, 2002). As a consequence, not only it is often difficult to identify what category is most relevant for describing the pattern of a given patient, but it is impossible even to sketch out the plausible boundaries of such impairment.

As an example, at the end of the nineties patient EA (Laiacona et al., 1997) has been described as generically impaired in BIOLOGICAL categories (and MUSICAL INSTRUMENTS). However, in a 9-years later reexamination (unpublished data reported by Capitani et al. (2003, appendix E)) a different pattern emerged, with a clear dissociation between the more impaired category of FRUIT AND VEGETABLES and those of ANIMALS and MUSICAL INSTRUMENTS.

Cases like this show the importance of relying on a more structured and systematically accessible knowledge base than the (therapist's or researcher's) human lexical abilities. Moreover, it suggests that our tool, as a consequence of the way the semantic information is organized, can be used also for a comparison between some of the proposed theories about category-specific semantic deficits. This is undoubtedly a positive by-product.

4

building the STaRS.sys Lexical Database

Electronic Lexical resources like WordNet (Fellbaum, 1998b), Cyc (Lenat, 1995), ConceptNet (Liu and Singh, 2004) or FrameNet (Baker et al., 1998), are widely used for a great variety of tasks, ranging from query expansion to word sense disambiguation, from text classification to textual entailment. In this chapter we will propose a new use for the information encoded into these lexicons. We will argue, indeed, that such resources may be exploited for building therapeutic tools, in a way that shares some commonalities with the project ViVA (Nikolova et al., 2009a,b). We will explain why we think that the semantic resource that best fits our needs is a wordnet-like lexicon and in which directions the existing wordnets should be extended, by leaving a detailed description of the modifications we designed to the following chapters.

4.1 Semantic Requirements

On the light of the scenario depicted in chapter 3, it is easy to see how the major challenge in developing a system like STaRS.sys comes from its need to fit the needs of every patient. This is a different view of flexibility than that driving the design of AAC tools such as ViVA. Indeed, adopting Nikolova and colleagues' terminology, the fundamental requirement a communicative tool has to meet is to be at the same time "*adaptable*", i.e. flexible enough to be reconfigured by the user, and "*adaptive*", i.e. able to tailor itself automatically to the profile of its actual user.

In the developing of a CAT tool like ours, on the other side, the notion of flexibility that has to be adopted is strongly connected to that of cognitive plausibility. That is, the only way for STaRS.sys to be useful in a therapeutic context is to be able to cope with the major variables that influence the performance of the patients reported in chapter 2, and this is possible only if it leans on a cognitively modeled knowledge base. In particular, we believe that, for every concept in our knowledge base, at least five kinds of semantic information have to be encoded.

4.1.1 Featural Descriptions

With some remarkable exceptions (e.g. Fodor, 1998), there seem to be a wide agreement on the plausibility of a somehow featural/compositional nature of human conceptual knowledge. This consensus spreads through different fields ranging from Linguistics (e.g. Pustejovsky, 1995), to neuropsychology (e.g. the OUCH model by Caramazza et al. (1990), but see Capitani et al. (2003) for a review) and cognitive psychology (e.g. the FUSS hypothesis by Vigliocco et al. (2004), but see Murphy (2002) for a review).

Adopting a mild position on this ongoing debate, for our purposes it is sufficient to assume, with Cree and McRae (2003), that speakers-generated Featural Descriptions (FDs), i.e. concept-description pairs of the form <cat> is lazy or <camels> are found in the desert, provide a window into the human semantic memory. In chapter 2, moreover, we reported that the same semantic information is exploited in the preparation of semantic tasks for the treatment of anomic patients (see Nickels, 2002).

As a consequence, a necessary requirement the the STaRS.sys lexical database has to meet is the availability of featural descriptions associated to every concepts. Ex-

ploited in the simplest possible way, such information can be used for choosing stimuli to submit tasks like semantic questionnaires. In addition, featural specifications can be used for selecting different kinds of concept groups, such as those sharing a feature value (e.g. *"red objects"*) or those for which a type of feature is particularly relevant (e.g. *"animals with a peculiar fur"*).

4.1.2 CONCEPTUAL TAXONOMY

A kind of Featural Description that is particularly interesting is the *is-a* relation (Collins and Quillian, 1969). On the basis of such kinds of information, indeed, concepts can be organized in a conceptual taxonomy, that is another vital requirement of our tool, especially in the light of the existence of patients affected by category-specific semantic disorders.

A category-specific semantic deficit is a disruption of the semantic knowledge that appears to disproportionately or selectively affects one semantic category. Since the first informative study available in the modern literature by Warrington and Shallice (1984), more than one hundred cases have been presented and discussed, three quarters of which were affected by a disproportionate impairment for LIVING THINGS if compared to NONLIVING THINGS (Capitani et al., 2003).

Others, more fine-grained, deficits have been described, involving, in different ways, spurious categories among which FRUITS AND VEGETABLES (Hart et al., 1985; Hillis and Caramazza, 1991; Samson and Pillon, 2003), ANIMALS (Caramazza and Shelton, 1998; Blundo et al., 2006), BODY PARTS (Warrington and McCarthy, 1987; Hillis and Caramazza, 1991), MEDICAL TERMS (Crosson et al., 1997), CONSPECIFICS (Ellis et al., 1989; Miceli et al., 2000), MUSICAL INSTRUMENTS (Warrington and Shallice, 1984; Laiacona et al., 1993a), FOOD (Warrington and Shallice, 1984; Hart et al., 1985; Hillis and Caramazza, 1991), GEMSTONES (Warrington and Shallice, 1984), WEAPONS (Gonnerman et al., 1997), VEHICLES (Hillis and Caramazza, 1991; Warrington and McCarthy, 1994; Gonnerman et al., 1997), CLOTHING and FURNITURE (Warrington and Shallice, 1984; Hillis and Caramazza, 1991; Gonnerman et al., 1997), OCCUPATION and FABRICS (Warrington and McCarthy, 1987) and EDIBLE LIQUIDS and SOLID EDIBLE SUBSTANCES (Borgo and Shallice, 2001).

As underlined by Capitani et al. (2003), however, much of these fine-grained trends of impairment are not supported by enough experimental evidence to let us assume the existence of a relevant category-specific deficit, so that it appears safer

to characterize the major patterns of category-specific semantic disorders along the following lines:

1. The most reliable category-specific deficit involves the dissociation of LIVING THINGS from NONLIVING THINGS;

2. The impairment of biological objects can be better described as affecting the two subcategories ANIMALS and FRUITS AND VEGETABLES. These categories can be impaired together or separately, and FRUITS AND VEGETABLES can be impaired also with NONLIVING THINGS;

3. FOOD and MUSICAL INSTRUMENTS can be impaired along with LIVING THINGS, even if not necessarily;

4. BODY PARTS are most often impaired along with NONLIVING THINGS.

As far as our tool is concerned, a minimal requirement is the design of a taxonomic organization able to deal with such major patterns of impairment showed by these patients.

4.1.3   FEATURE TYPES CLASSIFICATION

An efficient classiffcation of the types of information that can be associated to a concept is vital for the functioning of our system. Such level of representation, again, is not only useful per se, e.g. as a mean for controlling for some variables or for selectively work on feature types of interest (as happens, e.g., for the feature type color in Connolly et al. (2007), but it also allows for the estimation of feature-derived semantic measures such as feature distinctiveness, semantic relevance, concept similarity and feature correlation (Cree and McRae, 2003; Sartori and Lombardi, 2004; Vinson et al., 2003). As we will argue in the next section, feature types can be mapped onto semantic resource in a computational lexicon.

4.1.4   PROTOTYPICALITY AND FREQUENCY

Following McRae and Cree (2002), the habit of testing patients choosing a small sample of concepts can be very dangerous. Let alone issues like the fuzziness of category boundaries, a concept can be more or less representative of its membership

category (Murphy, 2002). Choosing and working on concepts with different levels of Prototypicality can be very informative, for both therapeutic and diagnostic purposes.

Another variable that can affect the patients' performance in semantic tasks is word frequency. Thereby, a critical skill for our tool is the ability to discriminate between words used with different frequencies. This would allow the final user to control for such dimension of variation.

4.2   Searching for a Semantic Lexical Resource for STaRS.sys

On the practical side, prototypicality and word frequency are properties that, once collected from scratch or extracted from exiting resources such as prototypicality norms (e.g. Uyeda and Mandler, 1980; Arcuri and Girotto, 1985) or frequency lexicons (for Italian, e.g. De Mauro et al., 1993; Bertinetto et al., 2005), can be easily encoded in any existing electronic lexicon by simply adding the appropriate fields[1].

The encoding of the others kinds of information, on the other side, appear to be a more challenging issue for the building of a semantic database. In details, we're looking for a semantic model able to meet the following requirements:

R1 it should be cognitively motivated;

R2 it should be based on a fully-specified is-a hierarchy;

R3 it should be intuitive enough to be used by a therapist;

R4 it should be apt to encode FDs.

In the design of the STaRS.sys knowledge base we tested the hypothesis that the only model able to meet the these requirements could be the WordNet (WN) model (Miller, 1995; Fellbaum, 1998b). At a first glance, indeed, WN seems to easily meet three of the above criteria.

First, WN was initially conceived as a model of the human lexical memory. Many psycholinguistic assumptions lay at the basis of this model (e.g. Miller et al., 1990; Miller, 1998a), and its psychological validity has been tested explicitly or implicitly

---

[1]Note that we are not saying that this information is easy to collect. However, their brute encoding in a semantic lexicon may not require any fundamental restructuring, so that we won't focus on prototypicality and word frequency any more in this thesis.
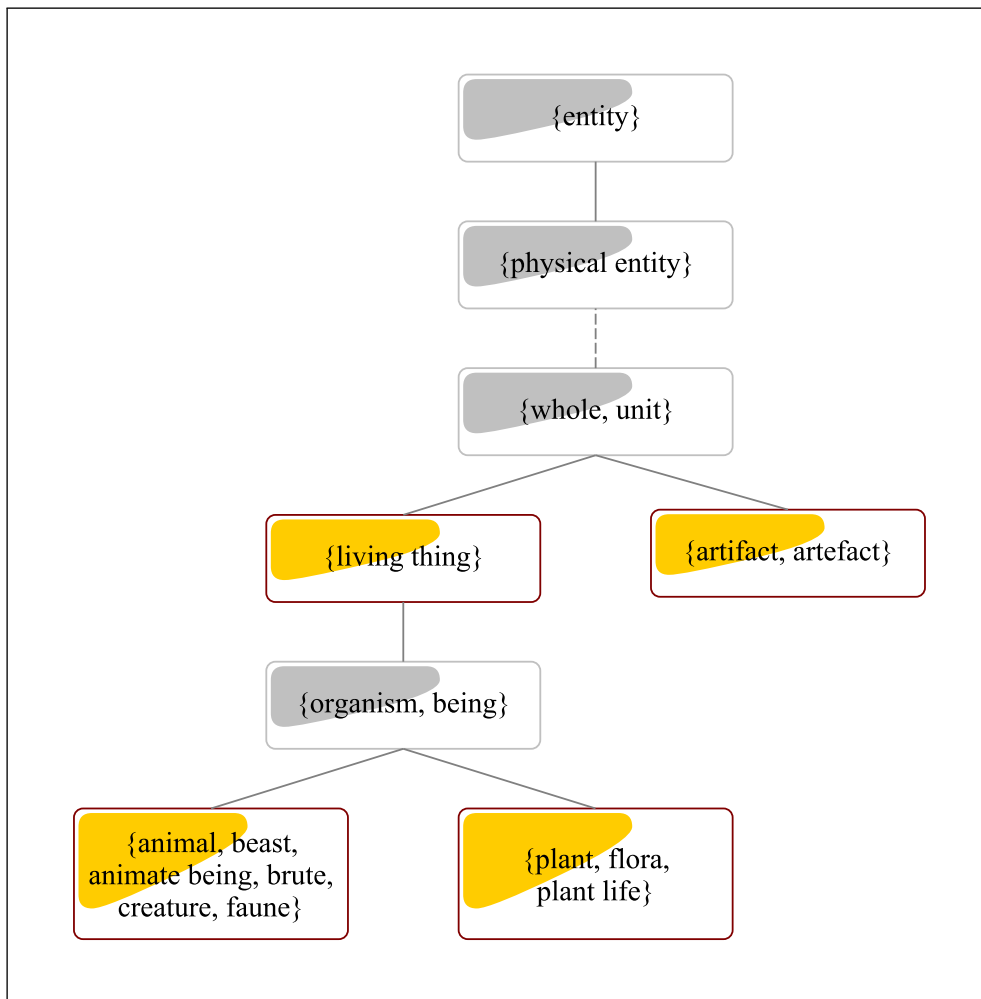
Figure 4.2.1: Mapping of WN 3.1 top nodes to therapeutic relevant categories (Dotted lines indicate jumps in the hierarchy)

by several scholars (e.g. Fellbaum, 1998a; Izquierdo et al., 2007; Barbu and Poesio, 2008).

Moreover, WN implements extensive and systematic noun hierarchies (Miller, 1998a). Tough not perfect from a strictly ontological point of view (see Oltramari et al., 2002; Gangemi et al., 2003; Miller and Hristea, 2006), figure 4.2.1 shows that the semantic categories which are relevant for rehabilitation purposes can be easily mapped onto WN 3.1 top level nodes (tools, animals, fruit and vegetables).

Third, WN is based on a conceptual model which is relatively simple and near to language use (as opposed to more sophisticated logics-based models). We expect

that this feature will facilitate the use of STaRS.sys by therapists, which may not have all the formal logics awareness that is needed to use formal ontologies.

4.3   A QUICK INTRODUCTION TO WORDNET

WordNet (WN) is the largest and most systematic electronic lexical database available to date. Its model is built around the key notion of "*synset*" (short form for "*synonyms set*") and of lexical and semantic relations connecting these minimal units.

4.3.1   THE NOTION OF SYNSET

The notion of synset is based on a weak, contextual dependent, definition of synonymy. According to this view, "*two expressions are synonymous in a linguistic context c if the substitution of one for the other in c does not alter the truth value*" (Miller et al., 1990, pag 240). Moving along these lines, it can be stated that it is possible to identify a concept with the set of words that can express it.

In WordNet such view is implemented by exploiting synsets like {morning, morn, morning time, forenoon} as pointers to minimal lexical semantic units, in this example to the notion of "*time period between dawn and noon*" as expressed in the sentence: "*I spent the morning running errands*". Polysemy, that is the fact that a word can and usually does have more meanings, is reflected in WN by multiple occurrences of the same lemma in multiple synsets, as it's the case for morning[2] in the following synsets:

> {morning, morn, morning time, forenoon}: the time period between dawn and noon
>
> {good morning, morning}: a conventional expression of greeting or farewell
>
> {dawn, dawning, morning, aurora, first light, daybreak, break of day, break of the day, dayspring, sunrise, sunup, cockcrow}: the first light of day
>
> {dawn, morning}: the earliest period

---

[2]all the material for the morning example comes from WordNet 3.1
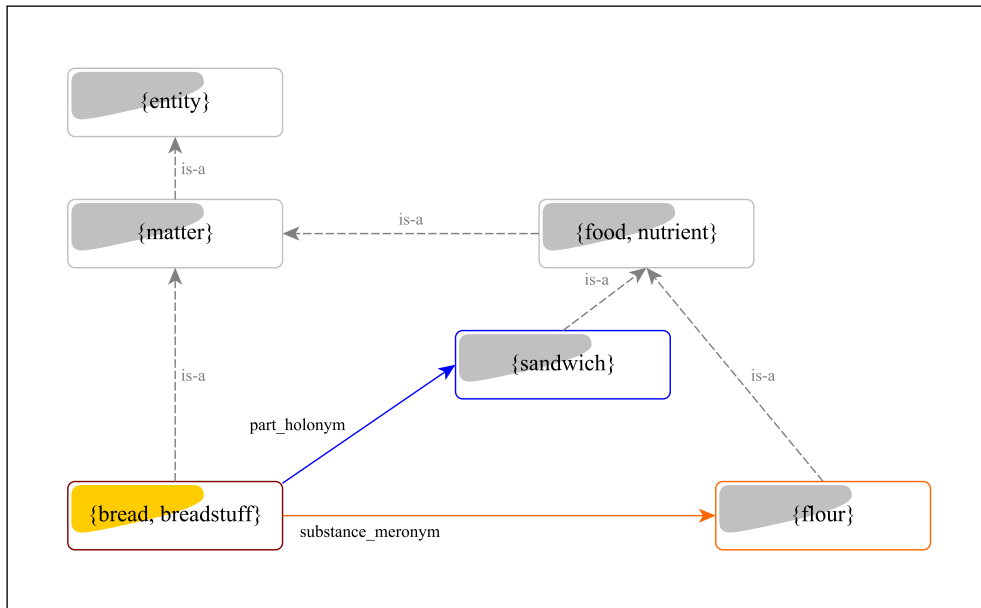
Figure 4.3.1: Sample of the WordNet 3.1 network

Synsets are connected each others via semantic and lexical relations, thus forming a network of the kind shows in figure 4.3.1. Lexical relations are defined as those holding between words, while semantic relations hold between whole synsets.

In WN, the different Parts of Speech (PoS) are organized around different semantic relations, so that the whole database is actually formed by four, poorly interconnected, semantic networks for the four major PoS: nouns, verbs, adjectives and adverbs.

The most important semantic relation for the noun synsets is the *hyperonymy/hyponymy* relation (Miller, 1998a), also dubbed *super/subordinate* or *is-A* relation, defined by Cruse (1986) in the following way:

> "X will be said to be a **hyponym** of Y (and, by the same token, Y a **superordinate** of X) if *A is f(X)* entails but is not entailed by *A is f(Y)* [...] where *f(X)* is an indefinite expression, and represents the minimum syntactic elaboration of a lexical item X for it to function as a complement of the verb *to be*."

Hyponymy is then the asymmetric relation holding between a more general con-

cept like {domestic animal, domesticated animal} and one or several more specific concepts, like {dog, domestic dog, Canis familiaris} and {domestic cat, house cat, Felis domesticus, Felis catus}.

In WordNet, every noun is bound to be part of the *hyponymy* chain by having a hypernym, so that from any point of the hierarchy it is possible to move up to the root node {entity}, as shown by the gray dotted lines in figure 4.3.1.

Other major semantic relations encoded for the noun synsets, though not as central as *hyperonymy/hyponymy*, are the *meronymic* or *part-whole* ones, exemplified by the colored arrows in figure 4.3.1. Inspired by the classification proposed by Winston et al. (1987) three such relations are implemented in WN: the *part meronym* relation holding between a concept and its components (e.g. {car, auto, automobile, machine, motorcar} → {car seat}), the *substance meronym* relation holding between a concepts and the substance it is made of (e.g. {steel} → {iron, Fe, atomic number 26}) and the *member meronym* relation holding between a group its members (e.g. {family, family unit} → {parent}).

The semantic relation central to the organization of the Verbs subnet is *troponymy* (Fellbaum, 1998a, 2002), that is a kind of entailment relation defining a manner elaboration. Two verbs synsets are said to stand in a *troponymic* relations if they express a particular manner of the other, as for {run} and {jog}.

Modifiers, on their side, are divided into descriptive adjectives , relational adjectives and adverbs (Miller, 1998b). Descriptive adjectives, the larger class, are organized in *antonym* clusters, each one centered on a direct antonym pair such as {beautiful} and {ugly}. Each adjectival synset that cannot form a *direct antonym* pair, belongs to a cluster as long as it is *semantically similar* to a pole of a cluster, as it is for {gorgeous} and {beautiful}. Such concepts are said to be *indirect antonym* of the direct antonym of their cluster central synset, in our case {gorgeous} is an *indirect antonym* of {ugly}.

Relational adjectives, like {atomic} and {musical}, are adjectives that are morphologically and semantically related to a noun. In most cases, these adjectives lack a direct antonym, so that these adjectives are encoded in WordNet by linking them to the noun from which they are derived from ({atom} and {music}). In a similar way, adverbial synsets are encoded by linking them to the adjective synset they are derived from (e.g. {highly, extremely} *pertainym* {high}).

4.3.3 Wordnets in the Outside World

In the years, the Princeton WordNet (PWN) has been successfully exploited in a wide range of Natural Language Processing tasks, from word sense disambiguation to query expansion, from machine translation to text summarization, from text categorization to multimedia retrieval (for a review, see Fellbaum, 1998b; Morato et al., 2004). Many extensions to the original model have been proposed, concerning the nature and number of encoded semantic relations (Alonge et al., 1998; Amaro et al., 2006; Boyd-Graber et al., 2006), a restructuring of its ontological status(Gangemi et al., 2003; Miller and Hristea, 2006), the encoding of syntagmatic information (Bentivogli and Pianta, 2004), the encoding of a domain hierarchy (Bentivogli et al., 2004b), its mapping to other resources such as Wikipedia (Wolf and Gurevych, 2010; Niemann and Gurevych, 2011) or Framenet (Laparra et al., 2010).

Furthermore, wordnets for specialized domains have been built, like Economic-WordNet for the economic and financial domain (Magnini and Speranza, 2001), Jur-WordNet for the legal domain (Sagri et al., 2004), ArchiWordNet for the architecture and construction domain (Bentivogli et al., 2004a) and Maritime-WordNet for the maritime domain (Roventini and Marinelli, 2004), WordNet-Affect for the representation of affective knowledge (Strapparava and Valitutti, 2004), sometimes with negative results (Poprat et al., 2008).

Bond and Paik (2012) report the existence of more than 40 ongoing projects to build wordnets in Languages other than English[3], aimed at the creation of resources for Languages as diverse as Albanian, Arabic, Bantu, Basque, Bulgarian, Catalan, Chinese, Czech, Danish, Dutch, Estonian, Finnish, French, German, Hebrew, Hindi, Indonesian, Irish Gaelic, Italian, Japanese, Korean, Latin, Macedonian, Malay, Nepali, Norwegian, Portuguese, Romanian, Russian, Slovene, Spanish and Thai.

Among these project, a prominent role is played by the multilingual projects EuroWordNet (EWN: Alonge et al., 1998) and MultiWordNet (MWN: Pianta et al., 2002). Apart from being the two most cited wordnets after PWN according to Bond and Paik (2012), these two projects exemplify the two main approaches for developing new wordnets identified by Vossen (1998): the "*merge*" and the "*expand*" models.

The key characteristic of the "*merge*" model is that the wordnet for every language is built independently, and linked to the others only in a second phase. In the case

---

[3]alternatively, a list can be found in the Wordnets in the World page maintained by the Global WordNet Association: `http://www.globalwordnet.org/gwa/wordnet_table.html`

of the EU-funded project that ended in 1999 with the creation of the seven new networks that compose the EWN database[4], this linking functions is accomplished by an unstructured version of the PWN 1.5 called Interlingual Lexical Index (ILI). ILI synsets are thus used as interlingual concepts to be used for moving from one language to another. The main advantage of exploiting a *"merge"* strategy is a relative freedom for the building of each wordnet. This freedom allowed the implementation in EWN of a series of modifications to the original wordnet model that have been exploited in subsequent works, such us the introduction of novel relations[5], also linking different PoS, the introduction of relation features (see chapter 6) and the exploitation of the notion of *"semantic order"* of entities formalized by Lyons (1977).

The MWN project, developed and maintained by Fondazione Bruno Kessler (formerly ITC-irst), is an instantiation of the *"expand"* model, the most widely exploited strategy for building new wordnets according to Bond and Paik (2012). In this approach new wordnets are built by creating the synsets of the new language in correspondence with the PWN synsets and importing the relevant English relations. The MWN database to date is composed by seven languages[6] and is still expanding. The main advantages of adopting an *"expand"* strategy is the minor complexity and the higher degree of compatibility between the aligned wordnets.

4.4   Shortcomings of the WordNet model

Given its widespread use and its long-life, the WordNet model has also been widely criticized. Some common criticisms are that:

- its sense distinctions are too fine-grained for some purposes like word sense disambiguation (but see Palmer et al., 2007);

- its content partly suffers from being tied to the lexicographers' intuitions (Fellbaum, 2006), so that sense identification may appear difficult even to human speakers (Fellbaum et al., 1997);

---

[4]i.e. wordnets for Dutch, Italian, Spanish, German, French, Czech, and Esto

[5]More than one hundred semantic and lexical relations are implemented in this multilingual resource, 47 of which involving nouns. An example are those reported in Appendix A.3.2. EWN, however, is based on the WN 1.5, so that the overlap with PWM is only partial (see Pazienza et al., 2008)

[6]i.e. PWN 1.6 and wordnets for Italian, Spanish, Portuguese, Hebrew, Romanian and Latin

- a better organization of the verbal subnet could be conceived, e.g. by encoding thematic roles (Alonge et al., 1998) or verbal alternations (Kohl et al., 1998).

However, the design of the STaRS.sys lexicon poses different issues from those faced by traditional NLP tasks. For our purposes, indeed, the critical limitations are those precluding the existing wordnets from satisfying the fourth criterion cited on page 35, that of being able to represent FDs like <cat> is a feline or <camel> is found in the desert.

In a brute approach, indeed, FDs could be represented in the synset glosses, currently composed of a definition and a list of sample sentences created by lexicographers. This solution is economical, in that no modification to the WN model should be implemented, and can even be useful in the context of the STaRS.sys for some tasks like retrieving all the semantic information associated to a concept. However, the most useful STaRS.sys functionalities, like comparing concepts, comparing descriptions or finding descriptions possessing certain characteristics, need a more explicit encoding of the semantic content of FDs.

What we're looking for is a procedure to map each FD as a semantic relation holding between a described "*source*" concept and the most prominent concept of the description, or "*target*" concept. Accordingly, then, FDs like <chair> has legs should be encoded into WN as a *meronymic* relation holding between the source synset {chair} and the target synset {leg}. This approach, moreover, is consistent with what is the common practice in both the NLP literature related to feature norms (e.g. Barbu and Poesio, 2008; Kelly et al., 2010, 2012) and the psychological literature (e.g. Vinson and Vigliocco, 2008; Andrews et al., 2009; Steyvers, 2010; Steyvers et al., 2010).

Barbu and Poesio (2008) have been the first to investigate the possibility of encoding FDs into WN, by measuring the overlap between the semantic knowledge encoded in PWN 2.1 and the descriptions collected by Garrard et al. (2001) and by McRae et al. (2005).

For every FD in the psychological norms, these scholars tested if the relevant focal concept was present in the lemmas composing the synsets and glosses of the "*semantic neighborhood*" of the source concept, where the notion of "*semantic neighborhood*" is defined as a graph <N,R> where N is a finite sets of synset nodes and R is a set of *hyperonymic* or *meronymic* relations linking the nodes. As an example, given FDs like <camel> is found in the desert and <cat> is a feline, the au-

thors mapped the source concepts `camel` and `cat` to the appropriate WN synsets and generate their semantic neighborhood, where they looked for the target concepts `feline` and `desert`.

The overlap found by Barbu and Poesio (2008) ranged from 22% to 40%, depending on the feature norm dataset and on the method (automatic vs. manual) used to calculate the overlap. Moreover, the same analysis showed that the WN coverage of the different kinds of FDs is highly skewed, with an overwhelming advantage of *categorical* and *meronymic* descriptions over the other information types, like the *functional* and the *evaluative* ones. Taken together, these results suggest that, in order to be usable in the context of the STaRS.sys project, WN should be enriched with novel kinds of semantic information.

In a somehow similar fashion, Boyd-Graber et al. (2006) describe the small number of semantic relations as one of the three fundamental limitations of the WN network. Other shortcomings are the absence of cross-PoS speech links (a problem already addressed in the context of the EWN project, see Alonge et al. (1998)) and the impossibility to encode the strength of semantic relations.

In order to enhance PWM, these authors proposed the introduction of a novel qualitative relation, *evocation*, representing how much a synset such as {`car, auto`} evokes another, e.g. {`road, route`}. To populate this relations, speaker generated judgments have been collected (Boyd-Graber et al., 2006; Nikolova et al., 2012), and these data proved to be very useful in the development of the project ViVA (see section 3.2), also due to the prominent role played by this kind of semantic information in the psychological literature (see Murphy, 2002).

Another problem of the WN model, marginally cited also by Barbu and Poesio (2008), is the impossibility to encode modification in the relation, as its the case for the quantifier in the FD <`car`> `has four wheels`[7].

Summing up, then, we can identify the followings four as the major shortcomings of exploiting an existing wordnet as the semantic base for a CAT tool:

- too few relations
- no possibility to encode the strength of a relation
- no possibility to encode quantification or logical operators
- no possibility to encode syntagmatic information

---

[7]Barbu and Poesio (2008), coherently with the common practice in the feature norms literature (see chapter 6), decided to neglect quantifiers.
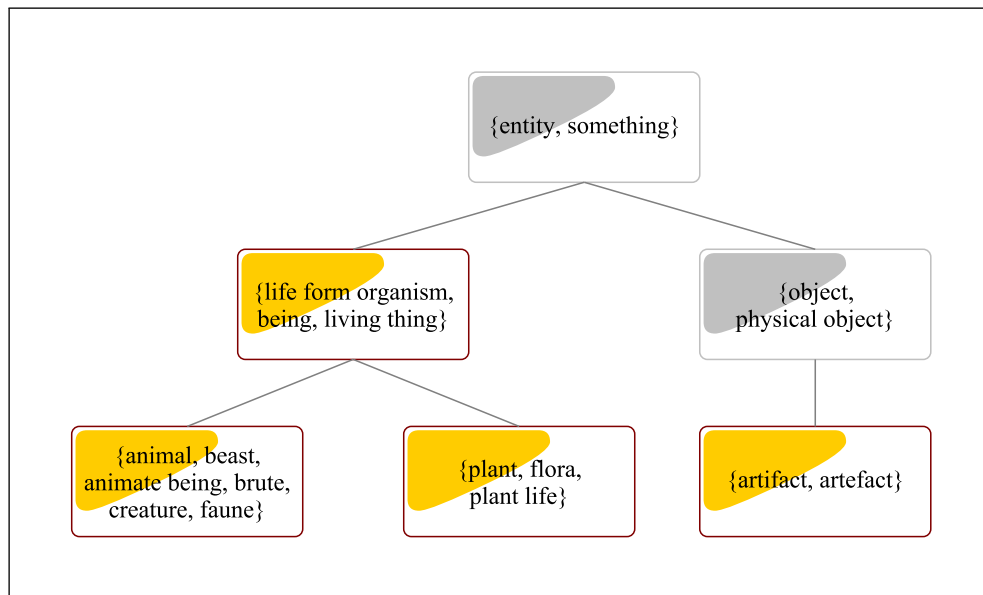
Figure 4.5.1: Mapping of WN 1.6 top nodes to therapeutic relevant categories

4.5 WHICH ITALIAN WORDNET?

In order to overcome these limitations, we designed and tested the modifications illustrated in the following chapters. The goal of the STaRS.sys project is the development of an Italian therapeutic tool, so that in implementing these modifications in an existing wordnet, we had to make a choice between two available networks: the Italian lexicon in the EWN database (IWN: Roventini et al., 2000) and the one in the MWN database (iMWN: Pianta et al., 2002).

Notwithstanding the major number of synsets encoded in IWM (48,529 vs. 38,877 synsets), we choose the iMWN lexicon for three main reasons. First, iMWN is still in development trough an on-line Web application. We expect that such application can be used by therapists using STaRS.sys for the shared and community-based development/maintenance of our lexical resource.

Furthermore, iMWN implements the notion of *"phraset"* introduced by Bentivogli and Pianta (2003, 2004) for coping with complex structures like `coltello da pane` recurrently used to express a concept, in our case a concept that corresponds to the English `breadknife`. As it will be shown in chapter 6, such device will turn to be essential in the encoding of complex FDs.

A third motivation in favor of the choice of iMWN is its stricter semantic align-

ment with the other Languages of the MWN project, obtained through the adoption of the "*expand*" model. A positive consequence of this characteristic, we think, is that at least part of the semantic information which is encoded for Italian can be ported to the aligned languages and used for similar purposes. This is a possibility that has yet to be explored, but whose plausibility is confirmed by psychological studies showing that speakers of different languages see similarities between objects in the same way, even in presence of very different patterns of naming (e.g. Malt et al., 1999).

As a post note, finally, figure 4.5.1 show that the semantic categories which are relevant for rehabilitation purposes can be easily mapped onto the WN 1.6 top level nodes as well, maybe in more intuitive way than what happens in WN 3.1 (see figure 4.2.1 on page 36).

5

a novel Feature Type Classification

One of the requirements that STaRS.sys has to meet is to implement a classification of the types of information that can be associated with a concrete concept. This classification, we argued section 4.1, can be exploited by the therapist in finding the needed stimuli. Given such a scenario, the classification we have in mind should meet at least the following three criteria:

R1 it must be built by moving from plausible hypothesis about the functioning and architecture of the human cognition;

R2 it must be powerful enough to account for any kind of linguistic description that a speaker associates to a concrete concept;

R3 it must intuitive and near–to–language–use enough for (1) being apt to represent the kind of information encoded by brief linguistic descriptions (i.e. FDs) and (2) being usable by therapists (that is, not by lexicographers or machines).

Notably, what is implied by these requirements if that the objects of our classification, FDs, are linguistic objects, as opposed to objectively and universally true concept properties (also see Cree and McRae, 2003). Different FDs somehow related like <crow> is black and <crow> has a black plumage have to be though as instances of different classes, even if they both imply that crows are black. In our case, they should be analyzed as belonging to the types *has Color* and *has Component*, respectively.

Moreover, from the decompositional approach described in section 4.4 (see page 42) it follows that assigning a FD to a feature type is equivalent to assigning it to a relation, given that is possible to mapping every type to a correspondent word-like relation (one type ~one relation type). That is, analyzing the description has legs in the FD <chair> has legs as of a *has Component* type, is equivalent to decomposing it into the *{source_concept, relation, target_concept}* triple: {chair} *has Component* {leg}.

It is easy to see the affinity between the information encoded in FDs and the relation-based conceptual representations common in NLP related fields such as ontological representations (e.g. Gerstl and Pribbenow, 1995; Guarino and Welty, 2000; Vieu and Aurnague, 2005), computational lexical resources (e.g. Lenat, 1995; Fellbaum, 1998b; Liu and Singh, 2004; Ulivieri et al., 2006), and theoretical linguistics (e.g. Cruse, 1986; Winston et al., 1987; Pustejovsky, 1995).

5.1 Background: Classifying Feature Types

The issue we faced in the design of our classification is a central issue for every work exploiting a FD-like representation:

> Is it possible to isolate a psychologically plausible set of description types (semantic relations) that efficiently represent the entire knowledge that can be associated to a concrete concept?

Before moving to a quick survey of the already existing compatible classifications, we feel the need to explain our view on the notion of *"cognitive plausibility"*. We intend as cognitive plausible any feature type opposition that can be motivated by referring to some fundamental aspects of the human cognition. From a procedural point of view, we fulfill this requirement by modeling our classification on three different kinds of evidence.

5.1.1 Psychological Evidence

First of all, we tried to account for well documented phenomena and psychological theories that receive some consensus. As an example, it seems that vision for perception and vision for action involve partially different neural mechanisms (Milner and Goodale, 1993). We interpreted this evidence as suggesting the existence of a clear distinction between perceptive FDs and those encoding the way actions are performed.

Moreover, it has been shown that the knowledge about how and why an object is used can be dissociated (Buxbaum et al., 2000). That is, that there are patients whose impairment affects just one of the two kinds of knowledge. Furthermore, anomia itself can affect just some semantic categories, leaving the others unimpaired (Capitani et al., 2003). We tried to find a hierarchy of feature types whose oppositions can be motivated by such existing psychological evidence.

5.1.2 Therapeutic Practice

In the so–called semantic therapeutic treatments (Nickels, 2002), anomic patients are often submitted tasks involving a feature-compliant representation of the semantic content of a concept. A vital component of this kind of tasks is an efficient classification of the types of knowledge to tap into.

It should be noticed that such classifications, being thought primarily for practical purposes, often appeal directly to the intuition and skills of the therapist, so that some of the oppositions they encode are vaguely defined. As an example, in their classification Laiacona et al. (1993b) mention a *contextual functional subordinate* class encompassing descriptions like sFeatsgrows underground, `is used by the carpenter` and `is played with a bow`. Even if the notion of "*context*" implied in the definition of this class can be intuitively interpreted, we think that a better definition of its meaning would be beneficial to the therapeutic practice. In spite of their limitations, we interpret the therapeutic effectiveness of such classifications as a proof of their psychological validity. This consideration persuaded us to look at them as useful sources of inspiration for the building of our classification.

Semantic Feature Analysis (Boyle and Coelho, 1995) is a technique in which the patient is encouraged to produce, for every target concept, six kinds of descriptions (*Group, Action, Use, Location, Properties* and *Association*). Apart from its well docu-

mented therapeutic effectiveness, a positive outcome of this protocol is that it provides the patient with a self-cueing strategy that is helpful in his/her everyday life (Davis and Stanton, 2005).

Laiacona et al. (1993b) presented a semantic questionnaire built for evaluating the naming abilities of the anomic patients. Such a questionnaire is completed with normative data from 60 normal old-age subjects. For each of the 80 concept stimuli of their questionnaire, such scholars propose six questions for investigating the following six kinds of semantic information: *general superordinate, superordinate within category, perceptual subordinate, comparative perceptual subordinate, functional subordinate* and *contextual functional subordinate.*

Finally, we asked therapists from CIMeC Center for Neurocognitive Rehabilitation (CeRiN) for a sample list of *{concept, description}* pairs they employed for therapeutic purposes. The classification adopted for organizing these (FD-like) attributes distinguishes between eight types, spanning along a main distinction between VISUAL (i.e. *color, dimension, matter* and *morphology* types) and NON VISUAL (i.e. *natural environment, taxonomic category, function* and *others non visual encyclopaedic* types) attributes.

5.1.3 OTHER PSYCHOLOGICALLY PLAUSIBLE CLASSIFICATIONS

The results of an accurate analysis and comparison of other available cognitively plausible classifications further helped us to isolate the individuation of other relevant oppositions.

A first classification that moves from a careful review of the psychological literature is the "*brain region taxonomy*" proposed by Cree and McRae (2003), one of the two classifications they exploited for marking the semantic content of their norms (McRae et al., 2005). In details, these scholars developed a taxonomy that "*can be linked to neural processing regions and [that] incorporates minimal assumptions*" (Cree and McRae, 2003). They distinguish between nine kinds of descriptions: three encoding various kinds of visual information, four encoding other kinds of sensory information, a functional type and a residual one. Because of the low number of oppositions implemented, we saw this proposal as a good starting point for our enquiry.

Probably the most exploited feature type classification [1] is the one proposed by

---

[1]modified versions of it are used also by McRae et al. (2005); Brainerd et al. (2008); Kremer and Baroni (2011)

Wu and Barsalou (2009). In their investigation concerning the role of perceptual simulation, these scholars developed a rather fined-grained two-level knowledge type taxonomy composed of 37 types partitioned into four major classes: TAXONOMIC CATEGORIES, ENTITY PROPERTIES, SITUATION PROPERTIES and INTROSPECTIVE PROPERTIES.

This classification, however, has been designed with a specific experimental goal in mind, so that many oppositions are of little use for our purposes, while others would even lead to miss some commonalities between FDs. As an example, the three FDs <tree> has lots of leaves, <tree> has leaves, <tree> has leaves depending on the type of tree would be assigned to three different types: *Quantity*, *External component* and *Contingency*, respectively. Instead for our purposes it would be more useful to encode all of them as (modified) meronymical descriptions.

As stated in chapter 4, WordNet itself is a cognitively modeled lexical resource. Even if *synonymy* and *hyperonymy* are its most important and populated relations, the third release of this semantic base encodes 46 different relations specified for every part of speech (20 for nouns).

5.1.4   RELATION SETS USED IN NLP ORIENTED RESOURCES (AND OTHERS)

We've already discussed (see section 4.4) about the scarcity of semantic relations encoded in WordNet and derived resources like MultiWordNet and EuroWordNet. Such a problem, however, is not a prerogative of WordNet. Similar considerations equally apply to other electronic resources like ConceptNet (Liu and Singh, 2004), the PAROLE-SIMPLE-CLIPS lexicon (Ulivieri et al., 2006) or to those theoretical works focusing on a specific subset of relations, like part-of relations (Winston et al., 1987; Gerstl and Pribbenow, 1995; Girju et al., 2006; Vieu and Aurnague, 2005), relations occurring between nominals (e.g. the SemEval tasks: Girju et al., 2007; Hendrickx et al., 2009) or what Morris and Hirst (2004) dub as "*classical relations*" (e.g. Cruse, 1986), that basically are the relations implemented in WordNet.

On the opposite side, classifications that implement an extensive number of relations, see the 15,000 types implemented in Cyc or the hundreds of the Roget's thesaurus (Mawson (1911); for an evaluation see Cassidy (2000)), are scarcely usable from our point of view. In such proposals, indeed, the encoding of very specific relations, like *ComputersFamiliarWith* referred to people (from Cyc, reported by Cassidy

(2000)), seems to suggest that the notion of relation involved is more near to that of "*linguistic predicate*" than to that of cognitively relevant property advocated here.

5.2   THE STaRS.SYS FEATURE TYPE CLASSIFICATION

The STaRS.sys feature type classification has been built in three distinct steps. In the first phase we isolated a set of candidate types moving from a critical analysis of the literature reviewed in the previous subsection. Given the current state of the art in therapeutic practice, we focused only on knowledge types that could describe what Lyons (1977) defined as "*first order entities*": concrete and physical entities that are publicly observable and that are located, in any point in time, in a three-dimensional space.

In a second phase, the candidate types have been exploited for annotating two different collections of FDs: the norms by McRae et al. (2005), that represents the most extensive resource of this kind to-date freely available, and those by Kremer and Baroni (2011), the only dataset available for Italian.

In a third step, all the points of inefficiency of the classification, such as overlapping between types, ill-defined types, types that appear to be motivated by specific needs, were fixed by removing or merging them, and the improved classification has been retested (and received further minor adjustments). Examples of types that have been discarded for these reasons are the *Quantity*, *Repetition* or *Meta-comment* ones proposed by Wu and Barsalou (2009).

The outcome has been the isolation of the 25 types reported in table 5.2.1 and defined in Appendix A.2. Inspiring ourselves from the Semantic Feature Analysis technique (Boyle and Coelho, 1995), we organized all of our types, apart from the residual *is Associated with* one, into six classes.

- TAXONOMIC PROPERTIES. This class include all those FDs describing a categorical relation between two concepts. in addition to the canonical *is-A* relation (Collins and Quillian, 1969), another such relation implemented in our classification is *Coordination*, holding between concepts that are similar in their belonging to the same category, like `cat` and `tiger`. Considerations leaded by the vital importance of categorization in human cognition and by the practical scope of STaRS.sys motivated our choice.

Relation	Description	Example	inverse
TAXONOMIC PROPERTIES			
is-A	Con_1 is a kind of Con_2	pear $\rightarrow$ fruit	*is the Category of*
Coordination	Con_1 and Con_2 share a common ancestor	dog $\leftrightarrow$ wolf	=
PART-OF			
has Component	Con_1 is an object that has Con_2 as a component	dog $\rightarrow$ tail	*is a Component of*
has Member	Con_1 is a collection or a group to which Con_2 belongs	wood $\rightarrow$ tree	*is a Member of*
has Portion	Con_1 is a mass of which Con_2 is a portion	bread $\rightarrow$ slice	*is a Portion of*
Made of	Con_1 is made of the substance Con_2	guitar $\rightarrow$ wood	*Composes*
has Geographical Part	Con_1 is a geographical area in which the location Con_2 can be found	sea $\rightarrow$ island	*is Geographical Part of*
PERCEPTUAL PROPERTIES			
has Size	Con_1 typically has size Con_2	rat $\rightarrow$ small	*is the Size of*
has Shape	Con_1 typically has shape Con_2	clock $\rightarrow$ round	*is the Shape of*
has Taste	Con_1 typically has taste Con_2	candy $\rightarrow$ sweet	*is the Taste of*
has Smell	Con_1 typically has smell Con_2	colonie water $\rightarrow$ rose	*is the Smell of*
has Sound	Con_1 typically produces the sound Con_2	dog $\rightarrow$ bark	*is the Sound of*

Table 5.2.1 (Continued from previous page)

Relation	Definition	Part of Speech	inverse
has Color	Con_1 typically has color Con_2	grass $\rightarrow$ green	*is the Color of*
has Texture	The surface of the substance that composesCon_1 typically has feel, appearance or consistency Con_2	eel $\rightarrow$ slimy	*is the Texture of*

USAGE PROPERTIES

Relation	Definition	Part of Speech	inverse
is Used for	Con_1 is typically used to attain the goal of to perform the action Con_2	cup $\rightarrow$ drink	*is a Use of*
is Used by	The tool Con_1 is typically used by the agent Con_2	hook $\rightarrow$ fisher	*Uses*
is Used with	Con_1 and Con_2 are typically used together to perform the same action	violin $\leftrightarrow$ bow	*=*

CONTEXTUAL PROPERTIES

Relation	Definition	Part of Speech	inverse
Situation Located	Con_1 is typically found in the situation Con_2	car $\rightarrow$ race	*is a Situation Location of*
Space Located	Con_1 is typically found in the location Con_2	fish $\rightarrow$ sea	*is a Space Location of*
Time Located	Con_1 is typically associated with the time period Con_2	grease $\rightarrow$ 50s	*is a Time Location of*
has Origin	Con_1 is produced by/is born in/grows in Con_2	apple $\rightarrow$ tree	*is the Origin of*

ASSOCIATED EVENTS AND ATTRIBUTES

Relation	Definition	Part of Speech	inverse
has Affective Property	Con_1 is associated with the emotional state Con_2	game $\rightarrow$ funny	*is an Affective Property of*

Table 5.2.1 (Continued from previous page)

Relation	Definition	Part of Speech	inverse
has Attribute	Con_1 has the property Con_2 which is not perceptual or affective	car $\rightarrow$ fast	*is an Attribute of*
is Involved in	Con_1 plays a non-instrumental role in the action or process expressed by the predicate Con_2	bird $\rightarrow$ fly	*Involves*
is Associated With	Con_1 is associated with Con_2 in a way that can't be described by any other relation	dog $\leftrightarrow$ man	=

Table 5.2.1: STaRS.sys feature-type classification (quick reference)

- PART-OF RELATIONS. Winston et al. (1987) proposed a six type classification of meronymic relations that can be expressed by an English speaker in talking about something being "*a part of*" something else. As in the WordNet model, we followed the work by these scholars in distinguishing five types describing a relation between a concrete concept and its part(s): *has Component, has Member, has Portion, Made-of* and *has Geographical Part*.

- PERCEPTUAL PROPERTIES. Moving from the "*brain region taxonomy*" by Cree and McRae (2003), we distinguished between six types of properties that can be perceived through the senses: *has Size, has Shape, has Taste, has Smell, has Sound, has Color* and *has Texture*. Note that, mainly because of the different purposes, our sub-classification is not perfectly isomorphic to the one proposed by these authors.

- USAGE PROPERTIES. This class is composed by three types of characteristics connected to the use of an object: *is Used for, is Used by* and *is Used with*. While the first two types have parallels in other classifications, we introduced the is Used with one for specifying a very common pattern found in a preliminary investigation.

- CONTEXTUAL PROPERTIES. FDs of this kind describe one of four different kinds of contexts in which an object can be found: *Situation Located, Space Located, Time Located* and *has Origin.* Even if the only classification parallel to our as for this class is the one by Wu and Barsalou (2009), all other relevant classifications have at least a type similar to one of ours.

- ASSOCIATED EVENTS AND ATTRIBUTES. This general class has been thought to encompass the opposition between features expressing an attribute of a concept and those that predicate about the role it plays in an action or in a process. FDs belonging to this class describe a permanent property of a concept (*has Attribute*) or the role it plays in an action or in a process (*is Involved in*). These two types are residuals in that each of the previous classes is a specification of one of them. In addition, this class includes a third type, *has Affective Property*, which express the emotional properties of an object.

The only type that falls out of our major classes is the residual *is Associated with*, used for classifying all those FDs that do not belong to any other type, like <dog> is a man's best friend. Its semantics can be roughly paraphrased as "*the two concepts are somehow related*".

A positive consequence of the methodology exploited for the creation of our classification is that it makes our proposal compatible with a number of well known theoretical and experimental frameworks. Our classification, then, may well serve as the common ground for the interplay of theories, insights and ideas originated from the above mentioned research areas.

To this purpose, the comparison tables A.3.1 and A.3.2 in Appendix A.3 illustrate the results of a first analysis of compatibility between our classification and the others. This tables show to what extent and in which cases it's possible to directly map the other proposals into ours. Even at a first glance, it is evident that no new relation or novel opposition has been introduced in our classification. That is, apart from the above cited case of the *is Used With relation*, every type of ours has a parallel type in at least one of the other classification. The most shared STaRS.sys types, on the other side, appear to be the *is-A*, the *has Component*, the *Made of* and the *is Used for* relations, all of which are present in all the other models.

5.3   STaRS.sys Classification Evaluation

The third main requirement that our classification has to meet is intuitiveness and ease of use. For testing if it is the case for the 25 types presented here, we asked to a group of members of the CIMeC CeRiN staff to exploit them for labeling a subset of the Kremer and Baroni (2011) norms and measured the inter-coder agreement between them (Artstein and Poesio, 2008).

In setting up the experiment, we followed all the Krippendorff (2004, 2008)'s recommendations: we employed easily reproducible coding instructions (see subsection 5.3.2), we relied on judgments from non expert coders working separately and we adopted a suitable agreement measure. The resulting agreement, measured using Fleiss' κ (a.k.a. multi–π) (Fleiss, 1971) because more than two coders were involved, has been interpreted both as a measure of reliability, and as a measure of how much intuitive and clearly defined are the distinctions in our classification.

In the literature there is no consensus on how absolute κ values should be interpreted. Some authors (e.g. Krippendorff, 2004, 2008) recommend 0.8 as a good reliability threshold, and 0.67 as a value that allows only minimal conclusions. Other scholars do consider reliable also κ values ranging from 0.67 to 0.8. The main significance of an agreement value, however, depends on the task, on the purpose of the study and on the methodology exploited (see Artstein and Poesio, 2008).

5.3.1   Preliminary Work

The evaluation reported here builds on the results of a preliminary study that tested a slightly different version (version 3.1) of our classification with six naive speakers not involved in speech therapy (Lebani and Pianta, 2010b). Subjects were submitted with a pen and paper questionnaire and a reference booklet. Results were promising (Fleiss' κ=0.73) and motivated the introduction of a new relation (*has Origin*), the removal of two relations (*has Domain*, *has Phase*) and the refinement of the definitions of other two relations (*has Sound*, *has Affective Property*).

Subsequently, we decided to check whether similar, or better, results could be obtained with speech therapists, which are expected to use the STaRS.sys classification in their daily work. To this purpose, we submitted the same task, with the modified classification but identical procedure, to three CeRiN therapists. We registered an overall agreement (Fleiss' κ=0.719) not substantially different from that obtained by

our naive subjects, but substantial disagreement on types (*has Smell, has Taste, Time Located*) that were well above the 0.8 threshold in the preliminary experiment.

Such a performance drop can be explained by two factors. A first possibility is that the experimental procedure we employed, even if adequate for an experiment in the laboratory (as it were the preliminary evaluation with the naive subjects), could be too demanding for a task to be completed by speech therapists in their spare time. Another possible hypothesis is that, because of the training they receive and because of their work, therapists are used to semantic distinctions that somehow collide with those implemented in our classification. We tested these two hypotheses by lightening the procedure and submitting an on-line version of the same task to other therapists and trainees. In this chapter we focus solely on such final evaluation.

5.3.2   Experimental Setup

Participants. Four Italian speakers from the CeRiN staff (1 therapist, 3 trainees) were recruited. All of them were about 25 year old (mean: 25.25, s.d. 0.96).

Materials. The test set was identical to the one used in Lebani and Pianta (2010b) and was composed by 300 FDs manually selected from a non-normalized version of the Kremer and Baroni (2011) dataset. We choose this dataset because: (1) it's a collection of descriptions generated by Italian speakers and (2) we wanted to avoid any bias that can be due to a normalization procedure (see chapter 6), so as to provide our subjects with FDs that were as plausible as possible.

The experimental stimuli have been chosen trying to balance the distribution of types and of concepts. As for the concepts, it's been easy to maintain an uniform distribution of FDs per source concept category (30 FDs for each of the 10 categories of the Kremer's dataset, see chapter 6 for details) and a fairly uniform distribution of FDs per concept (between 4 and 7 FDs for each of the 50 concepts).

It's been, however, impossible to balance the distribution of the feature types, mainly due to nature of the source concepts of the Kremer's dataset and to the skewness of its type distribution. Therefore, we decided to include in the test set 11 translated FDs from the dataset by McRae et al. (2005) and 12 FDs translated from the Leuven dataset by De Deyne et al. (2008).

Still, it has not been possible to reach the arbitrary threshold of ten FDs for the *has Portion* and *has Affective Property* types and no *has Member* and *has Geographical Part* FDs has been included in the test set.
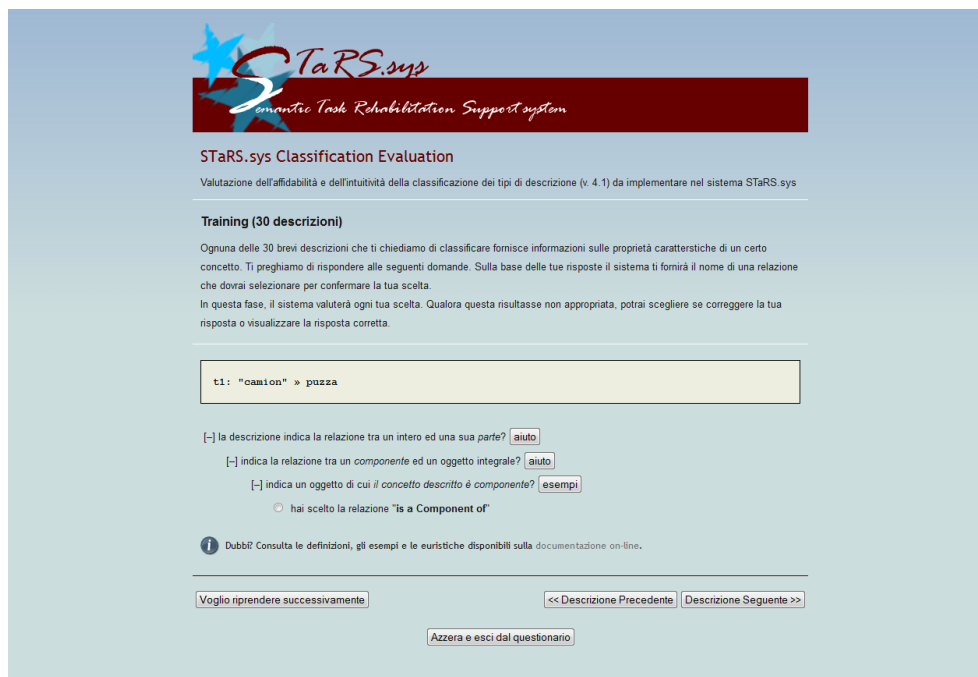
Figure 5.3.1: STaRS.sys classification evaluation: web interface

Procedure. The annotations have been collected through an on-line experiment, exploiting the web interface shown in figure 5.3.1. Participants were presented with a FD per web page, followed by a series of questions intended to guide and facilitate his/her choice.

Questions were prepared by rewriting the semantics of every type and every class in interrogative form like: "*does the description depict the typical shape of the concept?*" (for the type *has Shape*) or "*does the description depict a perceptive property of the concept?*" (for the PERCEPTUAL class). Questions were visualized on the basis of the partial answer of the subject, so that at every moment he/she had to consider just a limited amount of possibilities. For every choice, the relevant examples and definitions of the on-line documentation were accessible by clicking on help buttons located next to the question text.

Participants were asked to check the question that most accurately described the information conveyed by the description, and were allowed to make pauses and to perform the task wherever it pleased them. To get used with the task, they received a training set composed of 30 FDs, for which they received immediate feedback by the system and by the experimenter. On average, completing the task took 3 hours.
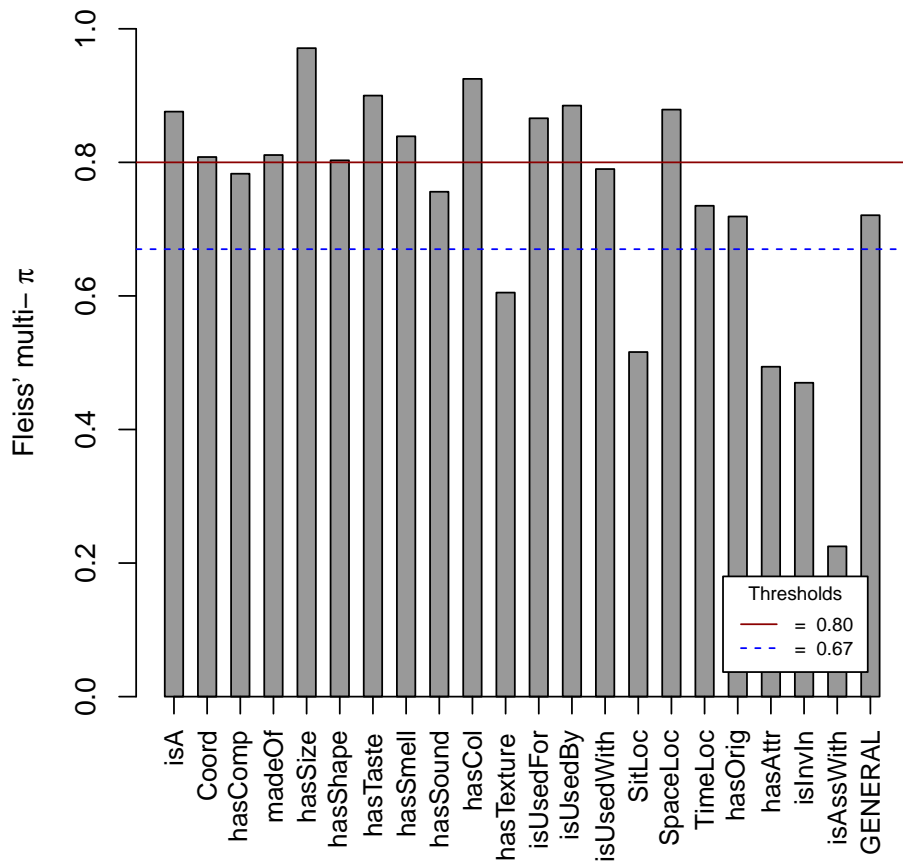
Figure 5.3.2: Annotators type-wise agreement (values reported in table A.4.1 in Appendix A.4)

5.3.3 RESULTS

The annotations collected from the participants have been normalized by conflating direct (e.g.*is-A*) and inverse (e.g. *is the Category of*) relation labels, and the agreement between their choice has been measured adopting Fleiss' κ. Figure 5.3.2 compare the type-wise agreement for each of our knowledge type, and compare it against the 0.67 and 0.8 thresholds commonly adopted in the literature. The middle column of Table A.4.1 in Appendix A.4 reports the actual κ scores if associated with $p < 0.001$.

Apart from the overall Fleiss' κ score of 0.721, the agreement is above the 0.8 threshold in 11 cases, above the 0.67 threshold in 4 cases and significant disagreement has been registered for 6 relations. The case of the *has Affective Property* relations is not puzzling, in that there were too few FDs of this kind in the test set to draw any conclusion. The semantic of the *has Affective Property* relation has indeed
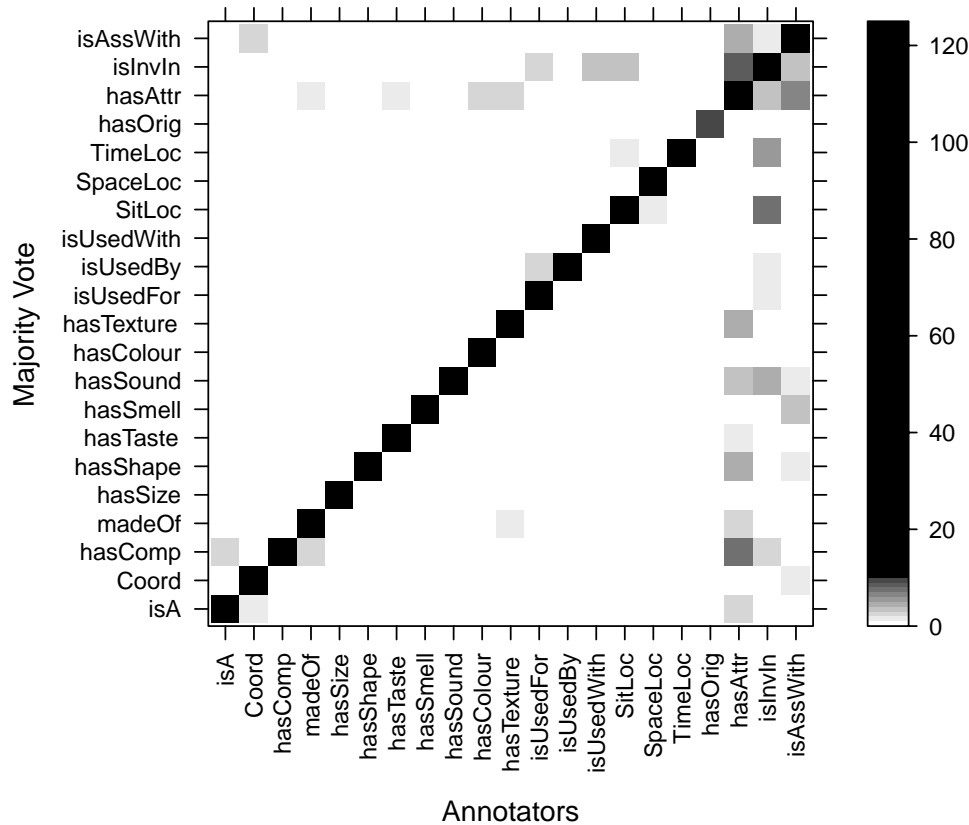
Figure 5.3.3: summed Annotators vs. Majority vote Confusion Matrix

changed from those of the version 3.1 of our classification, so that many of the FDs that were chosen in Lebani and Pianta (2010b) as candidate instances of this kind in this experiment should be thought as belonging to other kinds[2].

As for the other 5 relations on there is a significant disagreement, these are the most general, or "*residual*", ones: *has Texture, Situation Located, has Attribute, is Involved in* and *is Associated with*. This is endemic, in that, given the structure of our classification, the most plausible doubt is between a more general and more specific type, rather than within two different specific types. This interpretation is supported by the confusion matrix in figure 5.3.3,obtained by comparing the summed performance of the annotators against their majority vote. The majority vote is calculated by assigning to a FD the label chosen by the majority of the annotators [3] and rep-

---

[2]As suggested also by the non significativity of the relevant gold/majority κ score in the rightmost column in Table A.4.1. Note that we didn't represent this value in figure 5.3.2.

[3]We found 11 ambiguous cases, that we handled by referring to the annotations of similar FDs.
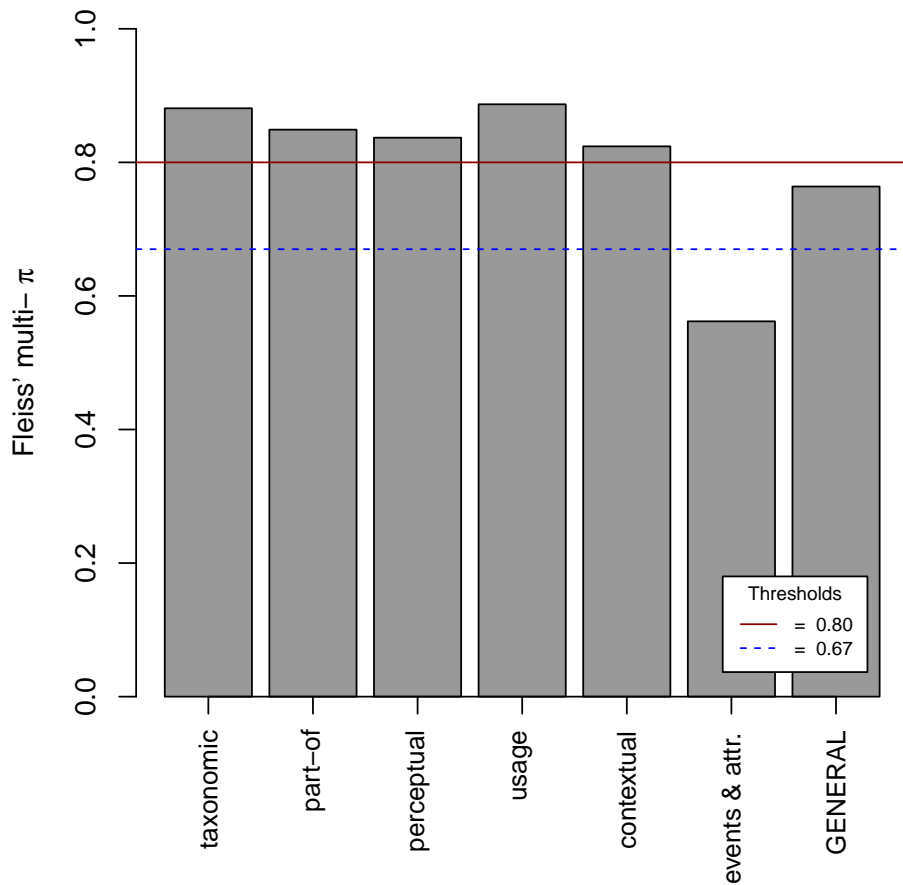
Figure 5.3.4: Annotators class-wise agreement (values reported in table A.4.2 in Appendix A.4)

resents what annotators should have chosen to perfectly agree. The distribution of false positives indicates that virtually all errors involve one of our five residual types.

The inter-coder agreement values referred to the classes of types, pictured in Figure 5.3.4 and reported in Table A.4.2, show a very similar pattern of low agreement on the residuals ASSOCIATED EVENTS AND ATTRIBUTES class as opposed to all the others, in which the κ score is well above the 0.8 threshold.

As pointed out by Artstein and Poesio (2008), the fact that annotators agree implies that they share a similar view, but not that they made the task in the right way. For evaluating the performance of our annotators, we compared their majority vote with the gold standard annotated by the two authors With some approximation, we see this last performance as the "*right*" one.

The agreement values, measured through exact κ (Conger, 1980), are represented
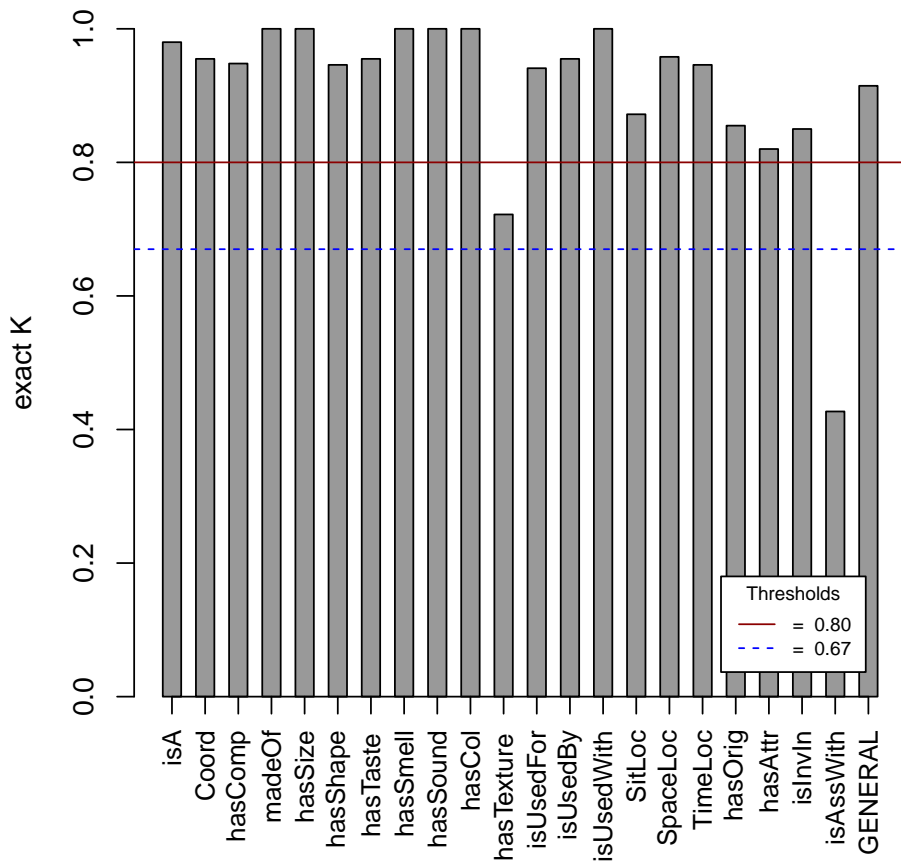
Figure 5.3.5: Gold vs. majority agreement (values reported in table A.4.1 in Appendix A.4)

in Figure 5.3.5 and reported in the rightmost column of table A.4.1 in Appendix A.4. The overall value is rather high (κ=0.915), and the only relations below the 0.8 thresholds are the residual *has Texture* and *is Associated with*, where the latter is the only one showing a significant disagreement.

These data further confirm the difficulties in handling residual types, but, more importantly, prove that our majority annotator has been able to learn the classification in a fairly correct way or that, at least, it did in a way similar to the two authors of the STaRS.sys classification.

5.3.4 DISCUSSION

Taken together, the performance of the therapists in the on-line task substantially replicates the results obtained by Lebani and Pianta (2010b) with naive subjects,

with slight improvements. The comparison reported in Figure A.4.1 (see Appendix A.4) shows a significant improvement of the agreement on the *has Sound* type (whose definition has been refined), and a significant agreement on the new *has Origin* relation (above the 0.67 threshold). More importantly, the pattern of out therapists' performance looks flatter, with important improvements on all the categories that shows a significant disagreement in the first evaluation.

Our results, moreover, look promising also when compared to analogous studies reported in the literature. For testing the McRae et al. (2005)'s derived coding scheme exploited in labeling their norms, Kremer and Baroni (2011) asked to an Italian speaker to label a random sample of 100 of their Italian FDs, reporting a Cohen's κ value of 0.676 in a task that is quite similar to ours. Moreover, they reported a much higher agreement ($\kappa = 0.844$) in an analogue evaluation of their German norms. Unfortunately, the authors do not report any other details about their validation, so that any deep comparison between our results and theirs is not possible.

LoBue and Yates (2011) reported a detailed evaluation of their classification of the kinds of knowledge involved in recognizing textual entailments. This classification has completely different goals from ours, even the linguistic form of the described entities is very different from our FDs. Examples of the categories identified in this works are *Definition*, as in the statement "*a seat is an object which holds one person*", or *Probabilistic Dependency*, as in "*Stocks on the Nikkei 225 exchange and Toyota's stock both fell, which independently suggest that Japan's economy might be struggling, but in combination they are stronger evidence that Japan's economy is floundering*".

These authors asked 5 annotators to label 221 statements of world knowledge with one of their 20 types, thus obtaining an overall Fleiss' κ of 0.678. However, just on 8 knowledge types the annotators reached the 0.67 agreement threshold, while on the remaining 12 they substantially disagreed.

What can be concluded by this quick review is that the development and evaluation of a knowledge type classification is a hard task, both for the experimenter and to the annotator. At the light of this difficulty, we interpret the results of our evaluation as a demonstration of the reliability of our coding scheme as well as of the usability of our classification. At the same time, these data suggest that, in the training phase of the final user of STaRS.sys, particular attention should be paid to the so-called "*residual*" relations in order to full satisfy our third requirement, that of (relative) intuitiveness and ease of use.

We are, however, aware of the limitations of our classification. Its main short-coming is that it handles only FDs that can be attributed to first order entities, i.e. concrete concepts. Another critical limitation, moreover, follows from the fact the our types have been modeled for representing semantic information that can be expressed by FDs, that are simple short linguistic descriptions. We leave to the future the investigation of whether and how the classification presented in this chapter could be extended to encode information associated to other kinds of concepts and expressed in more linguistically complex ways.

6

a WN-encoded set of Feature Norms

In chapter 4 we argued that the only usable approach for coding Featural Descriptions into a network-based semantic resource like WordNet is to encode them as a relation holding between two concepts. Accordingly, then, FDs like <cup> is used for drinking should be represented as a *is Used for* relation going from the *"source"* synset {cup}, representing the described concept, to the *"target"* synset {drink, imbibe}, representing the most prominent concept or concepts of the description.

Existing wordnet models and FDs collections, however, cannot be exploited for this task. Existing wordnets, indeed, appear to be able to encode only a small portion of the semantic information available in FDs collections. This is partly due to the lexical coverage and to the number and semantics of the semantic resources implemented in these resources. Other structural aspects contribute to this scenario, as well. Examples are the impossibility to encode the strength of a relation, the impossibility to encode quantifiers or the impossibility to encode syntagmatic information like selectional preferences or restrictions.

On the other side, also the existing FDs collections have to face some critical issues of data-sparseness. As an example, it is well known that the distribution of knowledge types in these datasets tends to be extremely skewed. As we will show at the end of the next section, it can be argued that the scientific questions that drive their building, i.e. characterizing the *"prominent"* semantic information in the human semantic memory, lay at the basis of the unexploitability of these resources for lexicographic purposes.

As a consequence, in order to identify the modification required to the wordnet model for encoding FD-like semantic information, we decided to build a new feature norms collection for 50 Italian concepts. We subsequently encoded these FDs in a dedicated version of iMWN dubbed StarsMultiWordNet (sMWN) to show the goodness of our proposal. This set of FDs, a the same time, constitute the core of the STaRS.sys semantic knowledge base described in chapter 4.

A project that share some commonalities with ours is the one centered around the *evocation* relation proposed by Boyd-Graber et al. (2006), that encodes how much a concept like {car} evokes another concept, e.g. {road}. In both works, indeed, the enrichment of a wordnet with speaker generated semantic information requires an adaptation of the wordnet model. In both works the resulting resources are apt to be used in the treatment of aphasic patients: ViVA (Nikolova et al., 2009a) and STaRS.sys. The crucial difference between these two works concerns the kind of semantic information encoded. A generic associative relation, such as *Evocation*, indeed, is not able to meet the requirements of the speech therapists, which need instead a more fine-grained classification of semantic relations, more similar to what can be obtained by exploiting a feature generation paradigm, than to what can be obtained through free associations.

6.1 FEATURE NORMS: THE STORY SO FAR

In psychology, it is common to investigate the nature of the human conceptual representation and computation by exploiting the so-called *"feature norms"*. Feature norms are collections of FDs elicited from speakers by asking them to describe a given set of concepts. For having an idea of the kind of linguistic material that is represented in such datasets, table 6.1.1 reports all the descriptions associated to the concept `airplane` in a subset of the freely available collections.

Dataset	Descriptions for the concept `airplane`
Garrard	`is a aircraft, is a vehicle, is large, is made of metal, is fast, can make a noise, has cockpit, has fuselage, has propeller, has seat, has tailplane, has rudder, has controls, has flap, can fly, carry passengers, can land, can taxi, can crash, can take off, is useful, can drop bombs, can fight, is powerful, has engine, is expensive, has pilot, is fuel-driven`
McRae	`flies, found in airports, has a propeller, has engines, has wings, crashes, is fast, is large, made of metal, requires pilots, used for passengers, used for transportation, used for travel`
Vinson	`deadly, destination, distance, hold, hollow, lift, mail, move, press, private, propeller, seat, sky, war, big, cargo, change-location, crash, device, ground, huge, object, passenger, safe, take-off, engine, fast, high, large, manmade, metal, wing, expensive, go, machine, travel, up, carry, vehicle, air, humans, transport, fly`
Kremer [ITA]	`a vehicle, used for transportation, flies, used for long travel, travels in the sky, has 2 wings, has wings, used by passengers, has an engine, is fast, is large, used for cargo`

Table 6.1.1: FDs associated with `airplane` in a group of freely available datasets

These descriptions, being the result of an explicit linguistic production task, should be thought as devices providing a window into a mental concept representation, rather than as faithful records of the semantic memory (Cree and McRae, 2003).

Nonetheless, similar collections has been employed in many different paradigms for over 40 years. The notion of *"family resemblance"*, first proposed by Wittgenstein (1953), has been tested by Rosch and Mervis (1975) by collecting feature norms, along with other ratings and judgments. These authors asked 400 speakers to describe 120 items from 6 categories in order to test the hypothesis that *"members of a category come to be viewed as prototypical of the category as a whole in proportion to the extent to which they bear a family resemblance to (have attributes which overlap those of) other members of the category. Conversely, items viewed as most prototypical of one category will be those with least family resemblance to or membership in other categories".*

Following this route, feature norms have been employed to design experiments (e.g. Rosch and Mervis, 1975; Ashcraft, 1978; Vigliocco et al., 2006), to build connectionist, computational or other formal models (e.g Collins and Loftus, 1975; Hinton and Shallice, 1991; McRae et al., 1997; Plaut, 2002; Vigliocco et al., 2004; Storms et al., 2010) or to account for empirical phenomena such as semantic priming, categorization and conceptual combination (see McRae et al., 2005).

Probably the theoretical issue where feature norms has been most widely employed is the investigation of category-specific semantic deficits. Many authors (e.g Garrard et al., 2001; Moss et al., 2002; McRae and Cree, 2002; Vinson et al., 2003; Sartori and Lombardi, 2004) ascribe these disorders to characteristics of the impaired concepts other than their categorical status, e.g. to a disruption of some featural properties or feature types (for a review, see Mahon and Caramazza, 2009).

Several statistics and regularities can be derived from the distribution of FDs in a collection, among which:

- *cue validity*: introduced by Rosch and Mervis (1975), it's the conditional probability of a concept, given a feature: $P(C_j \mid F_i)$;

- *distinctiveness*: following Devlin et al. (1998), it's the inverse of the number of concepts in which a feature or a feature class appears;

- *mean feature distinctiveness*: referred to a concept, it is a measure of the distinctiveness of the whole set of features derivable from its semantic representation (Cree and McRae, 2003);

- *feature correlation*: this notion has been described by McRae et al. (1999) as referring to the tendency of some description pairs, like `has feathers` and `has a beak`, to appear together;

- *semantic relevance*: proposed by Sartori and Lombardi (2004), the semantic relevance of a feature is a measure of how much it contributes in distinguishing a concept from other similar ones.

In this context, it is worth remarking the study by Wu and Barsalou (2009), that collected FDs for testing predictions following from Barsalou's *"situated simulation"* hypothesis (Barsalou, 2008). More influential than their conclusions, however, it's been the development of a feature type classification that became de facto a standard

for other similar collections built in the last decade, such as McRae et al. (2005); Brainerd et al. (2008); Kremer and Baroni (2011); Frassinelli and Lenci (2012).

In recent years, there has been some interest also in the NLP community towards the feature norm tradition. This interest focused both on the development of automatic methods for extracting feature-like representations (e.g. Poesio et al., 2008; Baroni et al., 2010; Kelly et al., 2010, 2012), and on the exploitation of this methodological device (e.g. Barbu and Poesio, 2008; Andrews et al., 2009; Steyvers, 2010; Steyvers et al., 2010).

Despite this wide use, however, the only freely available resources of this kind are:

- the Garrard dataset: Garrard et al. (2001);
- the McRae dataset: McRae et al. (2005);
- the Vinson dataset: Vinson and Vigliocco (2008);
- the Leuven dataset for Dutch: De Deyne et al. (2008);
- the Italian and the German datasets by Kremer and Baroni (2011);
- the contextualized norms by Frassinelli and Lenci (2012).

6.1.1 COLLECTING FDs

Feature norms are strongly influenced by the goals and theoretical framework of the connected studies, so that they differ substantially on the quantity and kind of described concepts, on the procedure adopted for collecting and normalizing FDs and on the classification adopted for classifying them.

As for the collection method, raw descriptions are typically collected by asking to a group of speakers to freely describe concepts, often explicitly stating that unwanted information like free associations of dictionary-like definitions should be avoided. A notable exception to this trend is represented by Garrard et al. (2001), that employed a sentence completion paradigm. That is, these authors asked their participants to complete sentence of the form [the concept] *"is......"*, *"has......"* or *"can......"*.

Another dimension of variation concerns the number and types of concepts described by the subjects. Most datasets, indeed, contain only descriptions of concrete objects, with the exception of the Leuven dataset, that contains 30 professions and 30 sports out of 425 concepts, and the dataset by Vinson and Vigliocco (2008), that contains 216 verbs and 71 nouns referring to events out of a total of 456 described concept. The freely available dataset with the highest number of described concepts is the one by McRae et al. (2005), that counts 725 concrete objects.

dataset	language	subjects no	concepts no	concept types	normalized FDs no	frequency filter	types classification
Garrard	ENG	20	62	concrete objects	1656	$f \geq 2$	ad-hoc (4 classes)
McRae	ENG	725	541	concrete objects	7259	$f \geq 5$	modWB, CM[br]
Vinson	ENG	280	456	concrete objects, events	12919	none	ad-hoc (4 classes)
Leuven	DUTCH	1003	425	concrete objects, professions, sports	*	*	none
Kremer [ita]	ITA	69	50	concrete objects	476	$f \geq 5$	modWB
Kremer [ger]	GER	73	50	concrete objects	486	$f \geq 5$	modWB
Frassinelli	ENG	25	8	concrete objects	2907	none	Wu and Barsalou (2009)

Table 6.1.2: Comparison of the freely available FDs collections. Values are referred to the available norms, not to the descriptive papers. Legenda: fields marked by * cannot be filled in because inappropriate; **modWB** stands for the modified Wu and Barsalou (2009) taxonomy proposed by Cree and McRae (2003); **CM[br]** stands for the the "brain region taxonomy" proposed by Cree and McRae (2003).

The number of subjects involved is another important dimension of variation, that range for the quite low 20 subjects recruited by Garrard et al. (2001) to the impressive 1,003 participants of the Leuven dataset. Some of these differences are highlighted in table 6.1.2.

For being exploited, however, raw descriptions must be processed in order to isolate the clusters of information they convey. The whole process can be divided into two phases. In a first normalization phase, the raw linguistic phrases are split, merged and conformed to a linguistic template in order to isolate the minimal chucks of information they convey.

From this perspective, we can oppose three approaches: (1) only a "minimal stemming" and removal of coordinations, subordinations and modifiers is performed (De Deyne et al., 2008); (2) the linguistic form is conformed to a phrase template according to a top down process (Garrard et al., 2001; McRae et al., 2005; Kremer and Baroni, 2011; Frassinelli and Lenci, 2012); (3) the linguistic form of the description is reduced to its focal concept (Vinson and Vigliocco, 2008).

In a subsequent processing phase, it is common to classify the normalized features according to the type of information they convey. By exploiting this procedure, more general descriptions clusters are identified, so as to formalize the intuitive resemblance between features such as is used for eating and is used for art work, and their distance with features such as is red. Classifications vary in the coarseness of the distinctions, and the specificity of the FDs they produce. While both Garrard et al. (2001) and Vinson and Vigliocco (2008) exploited two simple ad-hoc created four-category classifications, McRae et al. (2005), Kremer and Baroni (2011) and Frassinelli and Lenci (2012) adopted the 37-types classification proposed by Wu and Barsalou (2009), or a modified version of it[1]. It should be noted that McRae et al. (2005) further annotated their data with the 9-classes knowledge type classification by Cree and McRae (2003) and that no classification is employed in the Leuven dataset.

6.1.2 Shortcomings of the existing FDs collections

Recognized limitations of the available FDs collections concern the low number of described concepts and the fact that their building requires a very heavy and time-

---

[1] the McRae et al. (2005) modified Wu and Barsalou (2009) classification is composed by 27 types, that only partially overlaps with those of the original taxonomy.

consuming work. To have an idea of this, it's enough to consider that the building of the McRae dataset started in 1990 and has been carried out in three different universities.

The most widely cited analysis of the "*feature norms*" paradigm is the one by McRae et al. (2005). Two main shortcomings are highlighted. First, FDs are linguistically based, so that there's a clear advantage of some kinds of information, such as *parts* or *object locations*, over others that are equally psychologically relevant but more difficult, if not even impossible, to express in short linguistic phrases, as can be the case for "*spatial relations*" or "*kinds of movement*". Notice that some information types included in the STaRS.sys classification, like *has Smell* or *has Taste*, suffer from this same problem. After-all, we can describe the taste of an orange as being `sweet-sour citric, juicy with bumpy texture, soft...,` or `almost sour, but sweet, even like citrus that's sweet depending on the ripeness of the orange`[2], but definitively, an `orange` tastes like an `orange`.

Another issue concerns the behavior of the speakers. In analyzing norms, indeed, these authors noticed that participants tend to be biased towards those descriptions that "*enable people to distinguish a concept from other, similar concepts*". As a consequence, FDs collections tend to lack descriptions that are true for most concepts, like `has a heart`. McRae et al. (2005) does not give an explanation of this bias, but propose two possible explanation: either these are the most salient concepts in the speakers' mind or that's the way participants understood the experimental task.

A consequence of these two phenomena is that the distribution of knowledge types tend to very skewed in these collections. As an example, see how 75.44 % of the McRae FDs belongs to just 7 types out of 27. Because of the sparseness of property types, it turned out that none of the available collections can be efficiently exploited for our purposes, as we need to collect FDs that are as varied as possible.

Another issue is raised by Frassinelli and Lenci (2012), that stressed how participants are asked, in this tasks, to describe concepts in isolation, out of context, a rather unnatural condition. These authors collected 6922 FDs for 8 concrete concepts in 5 different contexts (1 isolation, 2 linguistic contexts and 2 visual contexts) from 125 English speakers[3]. Even if no effects of the context has been found in the number

---

[2]Retrieved from Yahoo!Answers in reply to the question "*What does an orange taste like?*": `http://answers.yahoo.com/question/index?qid=20080722230737AAhl7m8`

[3]note that table 6.1.2 reports different values. This is probably due to the fact that the one publicly available is an old version of their database.

of features produced by the speakers, these authors report a great context sensibility for some types, namely all the Introspective properties and some of the Situation Properties, following Wu and Barsalou (2009)'s classification.

Another problematic issue of the existing collections concerns the normalization of raw descriptions. Even if this practice is claimed to be as much conservative as possible, the ways in which it is usually carried out leads, from our point of view, to a loss of knowledge. Furthermore, our feeling is that too much is left to the interpretation of the persons in charge of the normalization. As an example, in the Kremer and Baroni (2011) norms, the description of the FD <garage> `can be used as a utility room` is paraphrased as `used for storing`. However in this way we miss the information that `garage` and `utility room` are similar concepts, encoded by the *Coordination* relation in our relation scheme.

6.2   Yet Another FDs Collection?

Given the limitation of the existing collections, we decided to conduct an elicitation experiment adopting the stimulus set by Kremer and Baroni (2011) (from here on: Kremer dataset) and a comparable number of participants, with a slightly different methodology. This allows us to to compare our dataset with the only freely available Italian FDs collection.

Our choice to adopt the Kremer norms has a positive by-product. In collecting and encoding the FDs, indeed, these authors accurately followed the methodology by McRae et al. (2005), and their subsequent comparison failed to highlight any remarkable difference between their (German and Italian) dataset and the English norms. For our purposes, we took advantage of these parallelisms as an indication that the conclusions drawn from our comparison with the Kremer norms could indirectly extend to the McRae et al. (2005) dataset, that is, to the biggest available FDs collection to date available.

6.2.1   Experimental Setup

Participants. 60 native Italian speakers participated to the experiment (10 males, 50 females). All of them were recruited in environment of the University of Trento or of the Fondazione Bruno Kessler and their age ranged from 19 to 55 years (mean: 28.9, s.d. 9.27). Undergraduate students received credits for their participation.

Concept Class	Concepts
BIRD	oca (*"goose"*), gufo (*"owl"*), gabbiano (*"seagull"*), passero (*"sparrow"*), picchio (*"woodpecker"*)
BODY PART	occhio (*"eye"*), dito (*"finger"*), mano (*"hand"*), testa (*"head"*), gamba (*"leg"*)
BUILDING	ponte (*"bridge"*), chiesa (*"church"*), garage (*"garage"*), grattacielo (*"skyscraper"*), torre (*"tower"*)
CLOTHING	camicia (*"chemise"*), giacca (*"jacket"*), scarpe (*"shoes"*), calzini (*"socks"*), pullover (*"sweater"*)
FRUIT	mela (*"apple"*), ciliegia (*"cherry"*), arancia (*"orange"*), pera (*"pear"*), ananas (*"pineapple"*)
FURNITURE	poltrona (*"armchair"*), letto (*"bed"*), sedia (*"chair"*), armadio (*"closet"*), tavolo (*"table"*)
IMPLEMENT	scopa (*"broom"*), pettine (*"comb"*), pennello (*"paintbrush"*), spada (*"sword"*), pinze (*"tongs"*)
MAMMAL	orso (*"bear"*), cane (*"dog"*), cavallo (*"horse"*), scimmia (*"monkey"*), coniglio (*"rabbit"*)
VEGETABLE	mais (*"corn"*), cipolla (*"onion"*), piselli (*"peas"*), patata (*"potato"*), spinaci (*"spinach"*)
VEHICLE	aeroplano (*"airplane"*), autobus (*"bus"*), nave (*"ship"*), treno (*"train"*), camion (*"truck"*)

Table 6.2.1: Stimuli used in the elicitation experiment (translated)

Materials. The stimulus set was composed by 5 concepts for each of 10 classes, resulting in a total test of 50 concrete concepts reported in Table 6.2.1. Kremer and Baroni (2011) selected these concepts from the sets exploited by McRae et al. (2005) and by Garrard et al. (2001), because their Italian (and German) translations are reasonably unambiguous and monosemic.

Procedure. The descriptions have been collected through an on-line experiment by exploiting the web interface shown in figures A.5.1 and A.5.2, see Appendix A.5.

12 groups of 5 tasks were prepared, each task composed by 10 random selected concept, one for each category, presented in casual order. In this way, every concept has been submitted to 12 subjects, and no participant received a questionnaire previously assigned to someone else.

The semantics of each relation has been paraphrased as a question of the form: *"what are the portions of a [concept]?"* for the *has Portion* relation, see figures A.5.1 and A.5.2 for the complete list of questions. This allowed us to populate as much as possible all feature types, and to reduce the need for interpretation in the normalization process.

Every participant has been presented a concept per web page, followed by a set of relevant questions. For each question, examples were available in the online documentation, accessible by clicking on the question text. Subjects were instructed not to report any biographic or technical knowledge, and they were allowed to leave a field empty if they didn't come up with any answer.

To get used with the task, they were trained on two example concepts (`cat`, `knife`) for which some suggestions were supplied in different ways (pre-filled fields, auto-completion). On average, completing the task took 1 hour, and participants were allowed to make pauses and to perform the task wherever it pleased them.

6.2.2   PRELIMINARY RESULTS: RAW DESCRIPTIONS

We collected 18,884 raw descriptions. Raw counts are reported in Table A.5.1 in Appendix A.5. For every concept, we collected on average 377.68 descriptions (s.d. 60.71). Every subject, on average, produced 314.73 (s.d. 115.68) descriptions over 10 concepts and 31.47 descriptions per concept (s.d. 13.71).

A chi-square analysis conducted by using the R toolkit[4], showed that the number of descriptions per concept category is significantly different (χ^2 = 208.79, df = 9, p < 0.001). The residuals analysis revealed that the less described categories were BIRD, BODY PART, IMPLEMENT and VEGETABLE, while the most described ones were MAMMAL, VEHICLE, CLOTHING, FRUIT and FURNITURE.

In a pre-processing phase every FD has been labeled according to the STaRS.sys classification. For this task, we took advantage of the fact that the descriptions have been produced have been produced in answer to a specific question that was formulated on the basis of one these feature types. That is, our classification phase has been driven by the the question-answering paradigm employed for the elicitation.

On the basis of their semantic, most of the properties of our classification can be grouped into types classes, along the pattern suggested by the gray shadings in table 5.2.1. A chi-square analysis highlighted a major difference in the distribution of

---

[4]`http://www.r-project.org/`
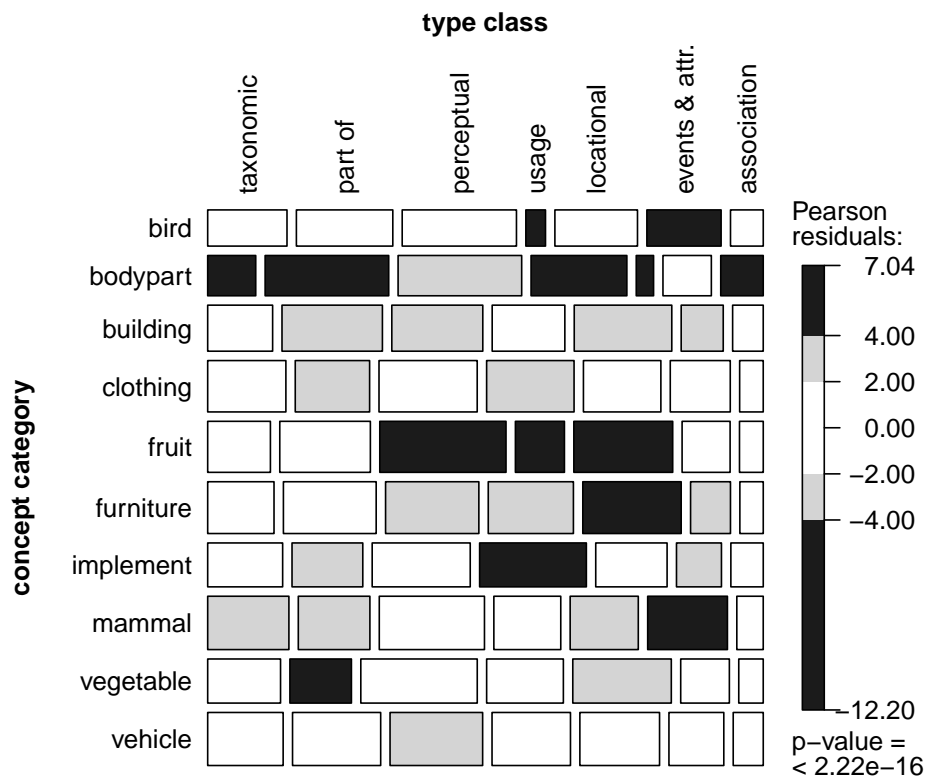
Figure 6.2.1: Cross-category distribution of types in the raw FDs

description classes among the different categories ($\chi^2 = 1030.82$, df $= 81$, p < 0.001). The details are shown by the mosaic plot in figure 6.2.1 (Meyer et al., 2006), where the width of every rectangle shows the proportion, for every concept category, of the relevant kind of description; while the height is proportional to the total number of features produced for a certain category. The shadings represent the results of a Pearson residual test, for which black shadings indicate a larger, more significant deviance from the expected values, and the gray shadings represent a medium-sized, still significant, deviance.

The appropriateness of the descriptions was manually checked by one of the authors. This led to the deletion of 1,023 raw descriptions, because they were conveying technical, autobiographical, information referred to a different meaning of the source concept or no information at all, e.g. aborted descriptions, typing errors. For the remaining descriptions, in 2,247 cases we recategorized the FD, and associated it to a feature type different from that implied by the subject. This process involved

mainly the Part of relations and the residual relations. Summing up, a total of 3270 features (17.3% of the total) underwent some change in this phase.

Comparison with the Kremer norms. A quantitative comparison of our dataset and the Kremer norms shows a significant difference in the proportion of raw description per category ($\chi^2 = 52.18$, df = 9, p < 0.001), but the analysis of residuals shows that this difference reaches a medium size significant deviation only for the bird and clothing concepts.

However, the biggest difference lays in the number of descriptions collected. Indeed, beside the fact that we collected more than twice the number of raw description the compose the Kremer norms (18,884 vs. 8,250), the important comparisons concern the number of descriptions per subjects (314.73 vs. 123.48), the number of descriptions per concept (377.68 vs. 170.4) and the average of feature per concept produced by every subject (31.47 vs. 4.96). These differences reach statistical significance at p < 0.001 on several Wilcoxon tests. Taken together, these data suggest that our strategy paid off, by providing a richer and more systematic set of feature descriptions for each concept.

6.3 sMWN: Normalizing is Encoding

The main goal of the normalization process, in the standard feature norm paradigm, is to group those raw FDs that somehow describe the same properties of an object, so to "*make sense*" of them. In the approach we're proposing, such a process coincides with the encoding of FDs into a wordnet-derived semantic lexicon. We're already discussed in chapter 4 how the STaRS.sys application scenario motivates choice to encode FDs as semantic relations holding between two synsets.

Accordingly, then, normalizing a pair such as `<cup> is used for drinking` is equivalent to encoding it as a *is Used for* relation holding between `{cup}` and `{drink}`.

Available existing wordnets, however, cannot be exploited for such task. The encoding is not a problematic for this resources, given that it is usually expressed as an isolated word that has to be described, e.g. `scimmia` (*"monkey"*). The only problematic aspect in this step may be the resolution of possibles ambiguities, an issue that in some case is not trivial at all. As an example, see the case of `cipolla` (*"onion"*), that in MWN has to intended either as a vegetable or as a food, among the others.

However, the main issue for the existing wordnets is the encoding of the descrip-

tion, i.e. `is used for drinking`, which is a free and possibly complex linguistic description. This cannot be achieved by the existing WordNet model because of (1) the scarcity of semantic relations defined and (2) the impossibility to represent composite semantic information.

We decided then to create a new wordnet, derived from the Italian MultiWord-Net lexicon, called StarsMultiWordNet (sMWN), where we encoded the raw descriptions elicited from our participants. The first extension to the wordnet model introduced in our semantic resource has been the implementation of our set of 25 semantic relations (~feature type).

This resulted in the creation of 21 new word relations, given the existence in the standard WN of semantic relation analogous to our *is-A*, *has Component*, *has Member* and *Made of*. Note that the identification of every semantic relation holding between the source and the target concepts of every FDs has been implicitly done in the preprocessing phase of our norms, see 6.2.2.

Other improvements of the wordnet models have been implemented in sMWN for allowing the encoding of the target concepts. These concern structural aspects and have been introduced for coping with recurrent syntactic and semantic patterns produced by our subjects. We will review these modifications in the next section.

6.4   Encoding Target Concepts into sMWN

The manual encoding of the FDs content in sMWN has been based on the following two main criteria:

1. the annotator should have minimum space for interpreting the data;

2. the simplification of the informative content of a description should be used only as a "*last resort*" strategy.

Normalization. In works belonging to the feature generation paradigm, the collection of the descriptions is always followed by a normalization step, in which semantically equivalent FDs are merged.

However, often a clear explanation of how equivalent descriptions are identified is missing. As an example, raw descriptions like `is a quadruped` and `has four legs` can be seen as exemplars of the same feature, say `has four legs` and merged (see Vinson and Vigliocco, 2008). It is questionable, however, the conclusion that these

expressions convey the same information. A quadruped is *"an animal that moves by using four legs"*, and reducing its definition to *"having four legs* is reductive.

In our approach equivalent descriptions are defined as descriptions sharing the same semantic relation and the same source and target synsets. Accordingly, then, we consider the two FDs <wheel> is a component of a car and <wheel> is an auto part equivalent because they can be both mapped into a *meronymic* relation linking {wheel} and {car, auto}.

Ambiguity. Polysemy is an inner property of language. We encountered a number of cases in which FD contained ambiguous words, i.e. lemmas that were present in different sMWN synsets. We identified two variants of this situation.

If the concurrent synsets are in a *hyponymy* relation, and the property is possessed by all the hyponyms of the more general synset, the latter has been selected. As an example, the target concept of the FD <coltello> è usato dal cuoco (*"knife is used by the cook/chef"* can be represented in sMWN as the Italian equivalent of either {cook} or {chef}, where the first is a hypernym of the second. In this situation, given that the property of using a knife is possessed by all hyponyms of {cook}, our choice fall on the more general synset.

On the opposite situation, that is when the property cannot be predicated for all the hyponyms of the more general synset, we opt for the more specific one. Consider the pair <ciliegia> cresce in giardino (*"cherry grows in gardens/grounds"*). The target concept, in this case, can be encoded with the Italian translations of both {grounds} and {garden}. However, since cherry trees usually grow in a {parvis} or in other hyponyms of {grounds} according to sMWN, we encoded this feature as a relation holding between {cherry} and {garden}.

In most cases the synsets corresponding to the ambiguous words are not one the hyponym of the other. As an example, given the FD <corn> can be found in a cellar, the target concept cellar can be encoded as either {basement, cellar} or {root_cellar, cellar}. Given that both synsets look plausible, we chose to double the concept-description pair in the database.

Loose Talk. Speakers are not dictionaries, so they may ignore some terms or they simply may not recall them in a certain moment. As a consequence, some raw phrases express concepts that could be expressed by an existing term, such as is used by people who cook.

In the standard feature generation paradigm, descriptions like these can be interpreted in many ways. They may even be re-phrased as features of a different kind, such as `is used for cooking`. In our approach, the rephrasing is guided by the synsets and glosses available in sMWN. In our case, then, the gloss associated with `{cook}` is "*someone who cooks food*", so that the choice has been easy.

6.4.1   MODIFYING THE WORDNET MODEL

Even if the bulk of the design of sMWN is the WN model as implemented in the iMWN lexicon, some minor modifications to it has been necessary to cope with the problematic recur-rent kinds of descriptions described in this section. In details, we propose to encode complex concepts by mean of the "*phraset*", extending the scope of the proposal by Bentivogli and Pianta (2003, 2004), and we implemented a set of relation feature for refining the semantic of a specific relation-concept pair, in so doing following the proposal advanced in the context of the EuroWordNet project (EWN: Alonge et al., 1998).

Compositionality. One of the most complex issues faced in the encoding of FDs into sMWN is given by complex linguistic descriptions. Whereas in WN synsets are bound to contain only lexical units (with the few exceptions of the so called "*artificial nodes*"), the target of a featural description can be a free combination of words, for instance a noun modified by an adjective, like `has a long neck`, an adjective modified by an adverb, like `is very big` or a verb with an argument, like `is used to cut bread`.

Our solution has been to exploit the notion of "*phraset*" defined byBentivogli and Pianta (2003, 2004) as a data structure used for encoding "*sets of synonymous free combination of words (as opposed to lexical units) which are recurrently used to express a concept*". In the original works, the authors introduced such a data structure to cope with lexical gaps in multilingual resources or to encode complex ways of expressing an existing concept. Phrasets can be associated to existing synsets to represent alternative (non lexical) ways of expressing lexicalized concepts, like the Italian translations for `dishcloth`:

> Synset: `{canovaccio, strofinaccio}`
> Phraset: `{strofinaccio_per_i_piatti, straccio_per_i_piatti}`
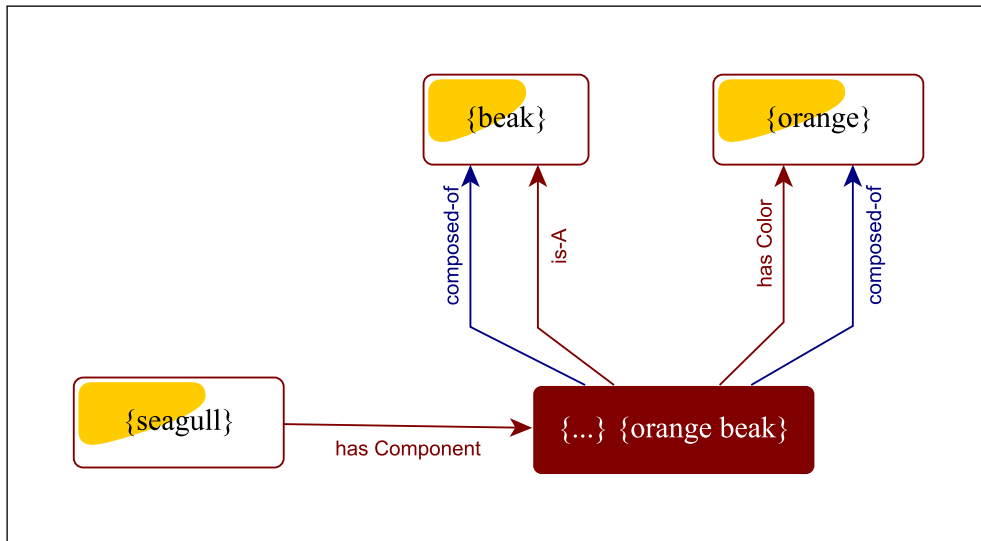
Figure 6.4.1: Representation of the FD `<seagull> has an orange beak`

where `strofinaccio per i piatti` and `straccio per i piatti` and are free combinations of words. Alternatively, they can be used to represent lexical gaps, such as the Italian translation equivalent of `breadknife`:

Synset: `{GAP}`

Phraset: `{coltello_da_pane, coltello_per_il_pane}`

Phrasets can be annotated by exploiting the *composes/composed-of* lexical relation linking phrasets with the synsets corresponding to the concepts that compose it. For instance the expression in the above phraset is linked by a hypernym and by a composed-of relation with the synset `{coltello}` (*"knife"*) and `{pane}` (*"bread"*).

Figure 6.4.1 shows how phrasets can be exploited for representing complex FD like `<seagull> has an orange beak` in sMWN. It also highlight a crucial difference with the original proposal, namely the fact that we represented also the semantic of the modifier (in our example `orange`), by linking the phraset to the modifying synset also by means of a semantic relation, in our case *has Color*. This has the positive outcome that allows us to draw inferences on properties of the described concept that would be otherwise lost.

Negation. In some norms collections, like the Kremer and the McRae et al. (2005) datasets, negative statements are treated as a separate class. In this collections FDs like `<bike> doesn't have an engine` and `<chicken> cannot fly` are treated

as conveying the same type of information. Moreover, the parallelism between FDs like <bike> doesn't have an engine and <bike> has 2 wheels is lost.

For our purposes, however, it is important to encode not only the properties that a concept possesses, but also those that it does not possess, if considered relevant by our speakers. Our solution is the exploitation, in sMWN, of a *negative* operator analogue to that implemented in the EWN database. In this way, a FD like <chicken> cannot fly is encoded as a *negation* marked sRelis Involved in relation holding from {chicken} to {fly}.

In accordance with the rationale behind the implementation of the *negation* operator in EWN, we noticed that the properties negated by our speakers can be seen as blocking "*expected*" undesired implications. In our example, indeed, the negated property {fly} is a distinctive property possessed by BIRDS, the category of the described concept.

Cardinality. An issue that every available collection has to face is the encoding of quantified properties like <human beings> have two legs. Many different, sometimes inconsistent, solutions have been proposed, none of which can be adopted for our purposes.

As a example, in the Vinson and Vigliocco (2008) dataset such quantified descriptions are analyzed as conveying separable complex meaning and split into different normalized descriptions like two and leg. However, what is predicated in the pair <human beings> have two legs cannot be equivalent to what is encoded by associating the concepts {two} and {leg} to the concept {human being}. Also McRae et al. (2005) treated these FDs as complex, but the strategy these authors adopted consisted in encoding, for any such quantified FD, two different kinds of normalized FDs: one simple, like <human beings> have legs, the other quantified, like <human beings> have two legs. The main drawback of this approach, however, is the introduction of a certain degree of redundancy in the data.

Our proposal is to encode cardinality by means of a *has cardinality* relation feature that specifies the number, numbers or range of numbers of the elements of the set referred to in the description. Accordingly, our example would be encoded as a *has Cardinality:2* marked *has Component* relation connecting the source synset {homo, man, human_being, human} to the target synset {leg}. Cases of FDs involving the same concepts but with different cardinalities, like <truck> has wheels, may have 4 wheels, may have 6 wheels, have been clustered by marking the range

or set of cardinalities encountered. In our example this would resulting in the encoding of a *has Cardinality:4,6* marked *has Component* relation connecting {truck, motortruck} to {wheel}.

Certainty features. Another common problem for the building of norms collections is the treatment of modifiers like generally, sometimes and most of the times. In the normalization phase such expressions are typically removed. Also standard WN encoding of semantic relations ignores any kind of qualification of the probability or strength of semantic relations between concepts. However we think that by ignoring this kind of information an important aspect of lexical meaning gets lost. In the same vein, Boyd-Graber et al. (2006) argue for the usefulness of adding to the WN model a characterization of the strength of the relation holding between synsets.

In sMWN we added a relation feature, called *Certainty*, representing the intuition of the speaker about how strong is his/her expectation that a certain relation holds between the instances of two concepts. We distinguish four levels of expectation:

- *True by definition*: the relation between two concept instances holds because of how the concepts are conventionally defined; no exceptions are admitted: <cat> is a feline;

- *Certain*: the relation to hold unless an anomaly occurs, which needs a causal explanation: <man> has two arms, <airplane> has wings;

- *Probable*: the relation is expected to hold most of the times; however if this does not occur it is not perceived as an anomaly: <wardrobe> is typically made of wood;

- *Possible*: the relation occurs sometimes, but not most of the times: <wardrobe> can be made of plastic.

These relation features represent a subjective notion of possibility/probability instead of a formally oriented notions defined in modal logic (Cresswell and Hughes, 2012). Note also that when a FD does not include any type of modifier, it is impossible to decide which of the four classes above it belongs to. Because of this, we represent the *Certainty* feature only when an explicit linguistic clue allows us to infer a value for it. In all other cases the value of the feature is undefined. We reserve for

the future the design of further experiments aiming at systematically collecting the value of the certainty feature for all relations, see Nikolova et al. (2012).

Conjunction and Disjunction. The last relation feature implemented in sMWN is a reimplementation of the *conjunction/disjunction* mechanism introduced in EWN for marking the relation holding between relations of the same type that have been predicated of a certain concept. For our purposes, however, the dichotomy *conjunction/disjunction* appears to be too restrictive. According to the mechanism introduced by Alonge et al. (1998), every concept instance is bounded to posses either all the properties of a kind that have been collected for it (*conjunction*), or just one of them (*disjunction*). This is not an issue for the EWN model, given the range of relations implemented in the lexical resource.

However, the semantic of some STaRS.sys relations cannot be fully captured by any of these two operators. This is particularly true for relations admitting some kind of optionality and for relations whose semantics is somehow underspecified or too wide. An example of the former kind is the is *a Space Location of* relation. Its semantics can be roughly paraphrased as "*the described concept typically is a place where the target concept can be found*". Given several relation instances of this kind, like `is a Space Location of trousers` and `is a Space Location of pullovers`, holding for the same source concept, e.g. `wardrobe`, it's easy to see how their mutual relation is nor of conjunction, nor of exclusive disjunction: in a wardrobe it is possible to find trousers, pullovers or both.

The same goes for "*residual*" relations like is Involved in. Given two instances of this kind, say `flies` and `is piloted`, referred to the same concept, e.g. `plane`, their mutual relationship can be described only as of inclusive disjunction: a plane can be piloted and can fly at the same time, but none of the two condition is necessary, e.g. it may well be driven by an autopilot or it is piloted also during the non-flying phases.

We therefore chose to implement in the STaRS.sys knowledge base three features labels for marking the relationship between concurrent feature: AND, analogous to the EWN *conjunction* label; XOR, analogous to the EWN *disjunction* label; and OR, inclusive disjunction. As illustrated in the last column of the table in Appendix A.2, every STaRS.sys relation is marked with a default feature value of *conjunction/disjunction*.

Relation instances that do not conform to the default behavior of their type can be marked by adding relevant labels to the semantic relations. As an example, the *has*

color relation is by default exclusively disjunctive. Accordingly, then, an `apple` can have just one color, say `green` or `red`. More complex color patterns, like `<apple>` `is red and yellow at the same time`, can be encoded as a group of relation instances standing in a conjunctive relation between them, like `<apple>` `is red` AND `<apple>` `is yellow`, and in a disjunctive relation with the rest of the *has Color* relation instances involving the same source concept.

6.4.2   RESULTS

Another characteristic of the WN-based normalization process described in this section is the fact that it allows for a systematic and consistent identification of synonyms. In our procedure, two FDs are synonymous if the lemmas or lemmas of their target concepts are members of the same synset.

Accordingly, then, the two FDs `<poltrona>` `è usata dalle persone` (*"armchair is used by the persons)"* and `<poltrona>` `è usata dagli essere umani` (*"armchair is used by the human beings)"* are analyzed as conveying the same semantic content, represented in sMWN as a *is Used by* relation from {`poltrona`} (PWN1.6: {`armchair`}) to {`persona, individuo, essere_umano, umano, mortale, anima`} (PWM1.6: {`person, individual, someone, somebody, mortal, human, soul`}).

When this situation was encountered in the FDs produced by the same subject, we decided to discard one of the two descriptions. We furthermore decided to apply a frequency threshold of 2, thus discarding all those FDs produced only by 1 speaker. This is coherent with the common practice both in the feature norms tradition, see table 6.1.2 for a comparison. Note that our frequency threshold is lower than the other adopted in the literature, but this is due to the lower number of subjects that described every single concept in our experiment.

Selected Descriptions. These two filtering processes reduced to 12,027 the actual number of speaker generated FDs that has been subsequently clustered and encoded into sMWN. We will refer to these descriptions as to the *"selected"* ones, to distinguish them from the *"raw"* FDs analyzed in the subsection 6.2.2, obtained by simply merging the lists produced by our speakers and discarding obvious errors.

Raw counts are reported in Table A.5.2 in Appendix A.5. For every concept we selected an average of 240.54 description (s.d. 47.89), and we retained an average
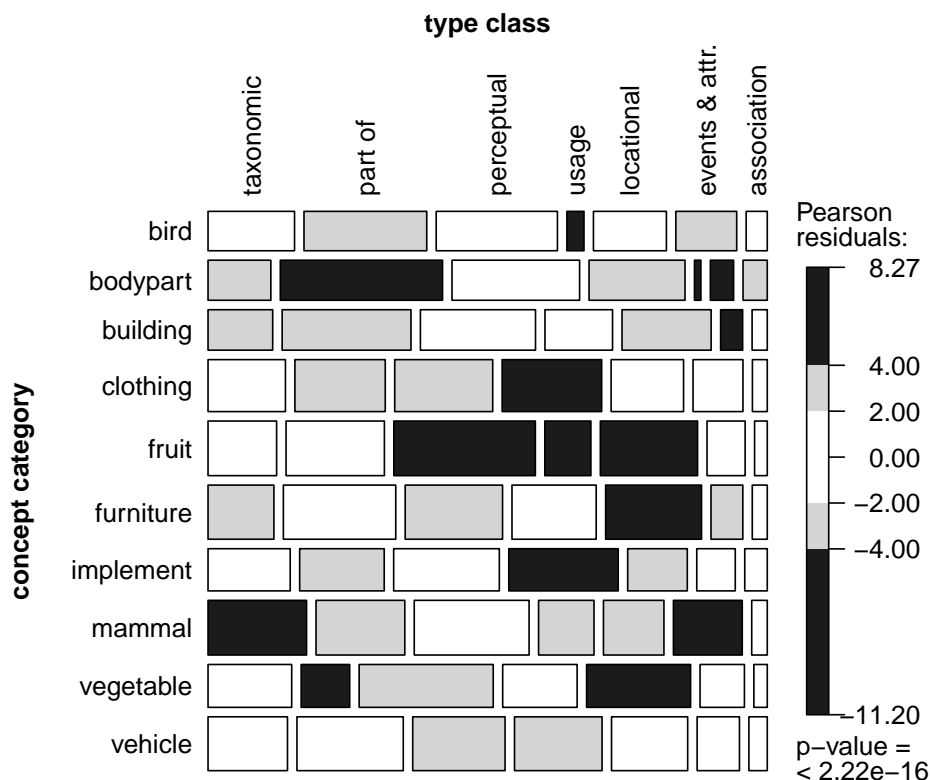
Figure 6.4.2: Cross-category distribution of types in the selected FDs

200.45 phrases (s.d. 54.58) from those produced by each participant.

Nevertheless, the main tendencies highlighted by the analyses on the raw FDs are still confirmed: the number of descriptions per concept category is significantly different ($\chi^2 = 228.43$, df = 9, p < 0.001), and the pattern shown by the mosaic plot in figure 6.4.2 shows that the distribution of property types among the different categories is significantly different.

Normalized Descriptions. The outcome of the encoding phase has been the insertion in sMWN of 3343 normalized FDs, resulting in a type/token ratio of 0.28. On average, every concept has been associated with 68.86 descriptions (s.d. 13.85). Raw frequencies are reported in Table A.5.3 in Appendix A.5.

The simplest normalizing procedure has been, as a matter of fact, powerful enough for encoding the vast majority of the description we collected. Indeed, 3116 of the normalized features (93.2%) have been efficiently encoded as a triple (source concept, relation, target concept). Of these, in 212 cases (6.34%) a synset for represent-
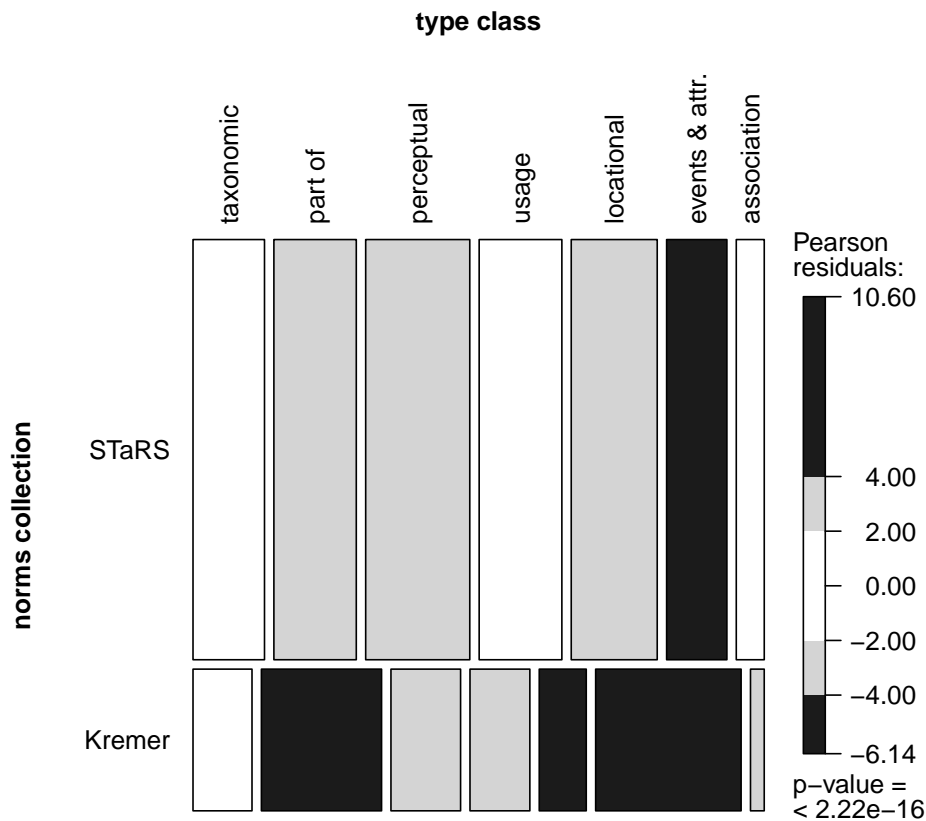
Figure 6.4.3: Cross-category distribution in our norms compared to that in the Kremer collection

ing the target concept was missing. The encoding of 112 normalized descriptions (3.26% of the total) required the creation of one o more phrasets.

In the manual identification of the involved synsets we had to face an average ambiguity of 3.27 synsets per lemma (s.d. 3), and 288 descriptions (8.39% of the total) are actually a doubling of descriptions for which more than one synset was appropriate in the context.

Overall, we overtly simplified 86 raw descriptions to encode them in 10 normalized descriptions (0.72% of the selected FDs). Only 7 selected FDs, corresponding to 3 normalized FDs, were discarded because it's not been possible to find an efficient way to encode them in sMWN.

Comparison with the Kremer norms. As a final analysis, we compared our norms with a version of the Kremer dataset normalized by exploiting the same procedure described in this chapter. By means of this re-encoding, we obtained a new set of
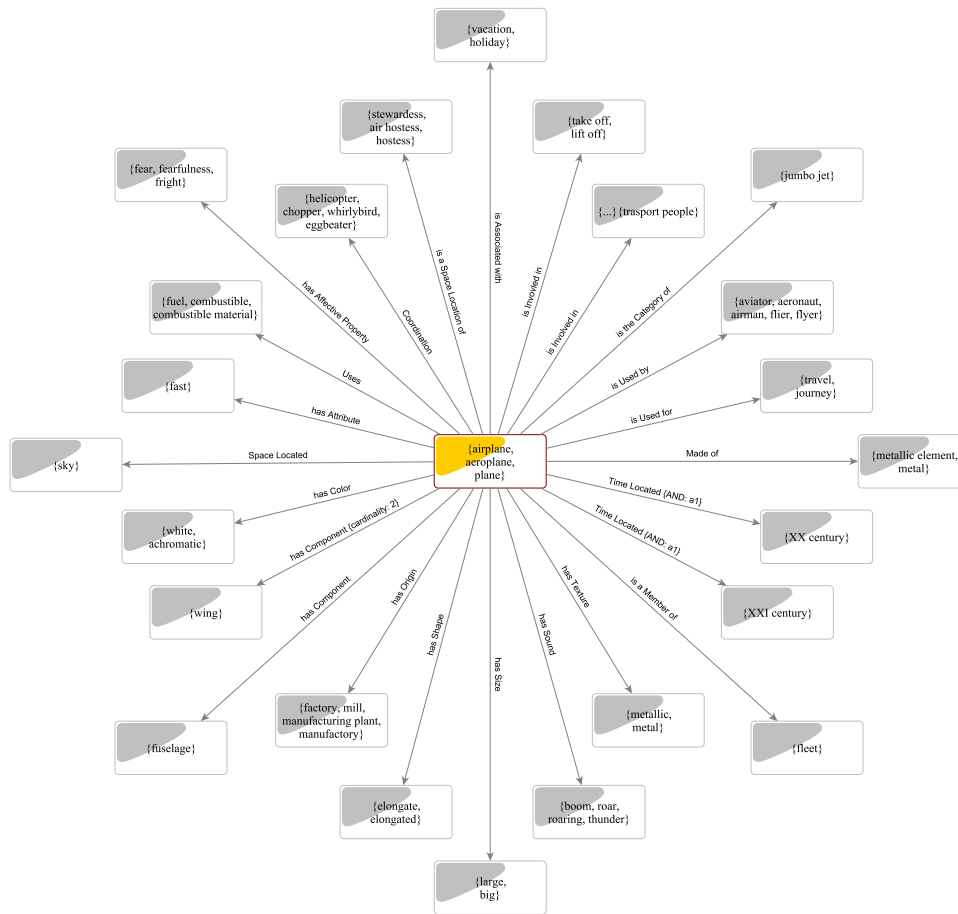
Figure 6.4.4: Portion of the semantic neighborhood of `Aeroplano` in sMWN (Italian synset lemmas are translated to the corresponding PWM 1.6 lemmas)

1,162 FDs[5], that is, a mean of 23.24 descriptions per concept (s.d. 4.75). The difference between this quantity and the average number of normalized FDs in our collection reaches statistical significance (w = 100, p < 0.01).

The mosaic plot in Figure 6.4.3, the two datasets differ significantly also in the distributions of FDs in the feature type classes (χ^2 = 259.79, df = 6, p < 0.001). While in our collection there are on average 118.72 descriptions for each feature types (s.d. 22.24), in the STaRS-normalized Kremer sample this value lower to 40.07 descriptions (s.d. 54.7).

Our feature norms collection, finally, seems to suffer a little less from the problem

---

[5]Note that in the Kremer dataset FDs produced by less than 5 subjects are discarded. By applying our frequency filter the number of FDs in the original dataset raises from 476 to 1,100.

of disproportionate representation of certain feature types over others discussed, among the others, by Kremer and Baroni (2010) and McRae et al. (2005). In their re-tagged dataset, indeed, the 6 most frequent relations account for the 67.99% of the whole set of features, while in our sample the 6 most frequent relations account for the 41.1% of the total amount of FDs. The most represented feature type in both datasets is the *has Component* relation, that involves the 17.64% of the Kremer FDs, and the 10.13% of ours.

6.4.3   PRELIMINARY CONCLUSIONS

Taken together, our results confirm our hypothesis that, with minor modifications, the WN model is apt to represent the kind of commonsense knowledge available in FD collections. Moreover, we interpreted these resulting as pointing to the effectiveness of sMWN as a semantic resource encoding kind of semantic information needed by our therapists, that is, every kind of knowledge that can be associated to a concrete concept. Figure 6.4.4 shows how a portion of the FDs collected for the source concepts `Aeroplano` are encoded into sMWN, thus shaping its semantic neighborhood.

However, we've already discussed how the collection of subject-generated Featural Descriptions requires a very time consuming and costly work. As a consequence, planning to populate the whole STaRS.sys semantic knowledge base with semantic information collected with this paradigm looks fairly unfeasible. That's why we decided to investigate the possibility to develop an automatic method for extracting FD-like statements to include in our wordnet.

7

the Automatic Extraction of Featural Descriptions

In recent years, the issue of automatic extraction of FD-like semantic information gained some interest in the NLP community. This task can be defined as the extraction of concept-description pairs that are though to hold for most of the instances of a concept by the majority of the speakers of a given language.

It is clearly related to the traditional semantic relation extraction task, for which many approaches have been proposed, from kernel methods (e.g. Zelenko et al., 2003; Bunescu and Mooney, 2005) to pattern-based methods (e.g. Hearst, 1992, 1998; Girju et al., 2006). Nevertheless, the automatic extraction of FDs poses a set of additional issues that makes it a deeply different task. In details, we identified the following core differences between the issue we're facing in these chapter and those usually addressed in the relation extraction literature:

1. there's no limitation in the number or kinds of semantic relations holding between the two concepts;

2. there's no ontological constraint on the involved concepts, other than the fact that the source concepts have to be concrete objects;

3. technical knowledge, that is the knowledge possessed just by a specialized minority of the speakers of a language, is ruled out;

4. the kind of semantic information extracted is not bound to be factually true, as long as the majority of the speakers agrees on it.

7.1 THE STATE OF THE ART

Only recently scholars begun exploring the possibility of exploiting state of the art extraction techniques to collect short linguistic description homologous to FDs. To the best of our knowledge, the only such works are those by Almuhareb and Poesio (2004, 2005), Barbu (2008), the Strudel model by Baroni et al. (2010) and those by Devereux, Kelly and colleagues (Devereux et al., 2009; Kelly et al., 2010, 2012).

All works but those by Devereux, Kelly and colleagues are based on the notion of *"lexico-syntactic pattern"*, first employed by Hearst (1992) to extract hyponyms from corpora. In her pioneering work, this author described a technique for identifying patterns of the kind:

NP_0 *such as* $\{NP_1, NP_2 \dots, (and \mid or) NP_n\}$

that can be exploited as unambiguous pointers for the existence of an *hyperonymic* relation holding between NP_0 and NP_i, where i > 0. An exemplar instance of this pattern could be the sentence *"Animals like dogs and cats are cute"*, and it can be taken as a clue to the existence of an *hyperonymyc* relation holding between `cat` and `animal` and between `dog` and `animal`. Hearst showed how such clues can be exploited for designing light-weight techniques that extract semantic knowledge from corpora in order to populate Machine Readable Dictionaries.

Almuhareb and Poesio (2004, 2005) exploited manually built lexico-syntactic patterns to extract concept descriptions from the Web. In their first work, they compared two different methods for building descriptions, one based on *"attributes"* like `color`, the other based on more general modifiers, called *"values"*, like `red`. The authors then compared the different kinds of representation by exploiting them in

a clustering task, and found that attributes-based models outperformed the value-based ones. Subsequently, Almuhareb and Poesio (2005) compared their simple pattern-based extraction method with another based on parser-generated grammatical relations. Again exploiting clustering for their evaluation, the authors reported a better performance for the simple pattern-based method. The indirect evaluation of their models, based on a clustering task, doesn't allow us to compare the results obtained by these author with those in the following literature. A main criticism to these results, however, has been raised by Devereux et al. (2009), according to which *"the pattern-based model performs well because the scope is restricted: the method of Almuhareb and Poesio is applicable to (and evaluated for) two types of relations between concepts and features only (is-a and part-of) and the patterns for each relation are developed manually."*.

A similar problem affects the work by Barbu (2008), which focused on six property types derived from the classification by Wu and Barsalou (2009): *superordinate, part, stuff, location, action, quality*. In his exploration, Barbu decided to exploit different methods for the different relations: while *action* and *quality* are extracted by means of a co-occurence based approach, the others are learned by means of lexico-syntactic patterns. By collecting the descriptions from the British National Corpus and from ukWac (Baroni et al., 2009), and by evaluating them against the WN-extended McRae dataset (ESSLLI dataset: Baroni et al., 2008), tha author reported quite high Precision and Recall values for the *superordinate* relation, i.e. 0.87 and 0.85 respectively, while obtaining low results on the others.

Strudel (Baroni et al., 2010) has been the first model to face the issue of unconstrained FDs extraction, that is, the extraction of conceptual properties of any kind, not restricted to a predefined list of target relations. The aim of this model is to identify *"the most distinctive properties for each concept"* by analyzing the distribution of more general, Part-of-Speech based, patterns than the Hearst (1992)-derived ones. Exemplar *"type sketches"* can be:

$C\_is\_P, C\_is\_ADV\_P$

where "*C*" is the described concept and "*P*" the candidate property. Examples of sentences instantiating such patterns are *"the grass is green"* or *"the grass is really green"*.

The Strudel model implements three leading intuitions, namely that (1) it is possible to isolate a group of general patterns over those that connecting a concept and

		Precision	**Recall**	**F–measure**
PAIRS	Baroni et al. (2010)	0.239	—	—
	Kelly et al. (2010)	0.1943	0.3896	0.2592
	Kelly et al. (2012)	0.2417	0.4847	0.3225
TRIPLES	Kelly et al. (2010)	0.1102	0.221	0.1471
	Kelly et al. (2012)	0.1238	0.2493	0.1654

Table 7.1.1: State of the Art: performance of the best performing models for the extraction of FDs semantic information (evaluated against the ESLLII dataset)

a property in text; (2) that the presence of a semantic link, not necessarily semantic relation, can be cued by the variety of pattern connecting a concept and a candidate property; (3) that the distribution of connecting patterns is less ambiguous than single patterns à la Hearst for characterizing the type of relation holding between concepts. Accordingly, then, in Strudel a strong cue for the existence of a semantic link between a concept and a property is the fact that they are connected by several distinct patterns.

Training Strudel on ukWac and evaluating it against the ESSLLI test set, the authors reported a precision score of 23,9%, to date the highest value obtained by a model not focused on a specific subset of relations, together with the model by Kelly et al. (2012), as shown in Table 7.1.1. Unfortunately, Strudel characterizes the semantic link between a concept and its property in an implicit way, as distributions of type sketches. Coherently with Devereux et al. (2009), then, we see the Strudel model as the best one to-date available to extract *{concept,property}* pairs, but that cannot be exploited for mining the *{concept,relation,property}* we need to encoded into sMWN.

Devereux, Kelly and colleagues (Devereux et al., 2009; Kelly et al., 2010, 2012) have been the first scholars exploring the possibility to automatically extract from corpora FDs-derived *{concept,relation,property}* triples of the kind `turtle` *has* `shell`, without limiting their searching space to a finite set of semantic relations.

In Devereux et al. (2009); Kelly et al. (2010), the authors proposed a two-stage model that they applied to parsed versions of the British National Corpus and of Wikipedia and that they evaluated against the ESSLLI test set. In the first stage *{concept,relation,property}* triples for a given sets of concepts were extracted by exploiting

manually generated syntax-based rules. In the second phase, these candidates triples were re-ranked and filtered on the basis of the conditional probabilities of concept and feature classes derived from the McRae dataset. Results for their best models are reported in table 7.1.1.

Other then the feasibility of this approach, these initial works highlighted the importance of exploiting syntactic information, the advantages of searching concepts in different kinds of corpora, BNC and Wikipedia, and how the knowledge extracted from FDs collections can be useful for shaping the selection of corpus extracted triple candidates.

7.1.1 THE SEMI-SUPERVISED APPROACH BY KELLY ET AL. (2012)

Their best performing model is, however, the semi-supervised one proposed in Kelly et al. (2012). The methodology proposed in this work articulates itself into three phases: training, extraction and filtering of candidate triples.

Training. In an initial phase, all sentences containing an instance of a *{source_concept, target_concept}* pair from an ad-hoc normalized British version of the McRae norms (Taylor et al., 2011) are extracted from Wikipedia and from the ukWac corpus.

In order to be exploited for this process, however, the modified McRae norms had to be previously recoded from their linguistic form to a *{source_concept, relation, target_concept}* form. The goal of this process was to reduce the description part of the FDs to pairs like *{relation, target_concept}*, where the target concept was bound to be a single lemmatized word. This process yielded a total set of 7,518 triples for 510 concepts[1], linked by one of the 254 distinct relation reported in Appendix A.6.1. It should be noted that the notion of relation employed here is distant from that of semantic relation employed in our FDs collection. It should be probably better understood as "*the concatenation of the text which lies after a given concept but before any noun/adjectival/adverbial features*" (from a personal communication with C. Kelly).

Accordingly, assuming the existence of the FD <airplane> has wings in their dataset, this should have been reformulated as the triple *{airplane, has, wing}* and all the sentences containing airplane and wing extracted from Wikipedia and ukWac.

The sentences extracted from the corpora are subsequently dependency parsed with the C&C parser (Clark and Curran, 2007) and for each of them the path con-

---

[1] The authors reported that some source concepts of the original McRae dataset have been discarded for interlingual incompatibility.

necting the two concepts is stored and labeled as an instance of a given relation. The format chosen by these authors for representing the linking path is dubbed GR-POS graph, where GR stands for "*Grammatical Relation*" and POS for "*Part of Speech*", and it is an acyclic graph whose nodes are the words of the sentence additionally labeled with their POS, and the edges are the grammatical relations connecting the nodes. For every single path, then, a relation-labeled flat vector of the kind in Table 7.1.2 is built and all vectors are feed to SVM*multiclass* (Tsochantaridis et al., 2006) for generating a learned SVM model.

Two such kinds of vectors were tested by these authors: one encoding the presence in the path of a "*relation verb*", that is a verb present in the training set of relations; the other kinds of vector simply ignores this information.

Extraction. In a second phase all the sentences containing one of the 44 source concepts of the test set are extracted from the above cited corpora, parsed and the GR-POS paths identified. Subsequently, the learned model is used to classify these GR-POS paths.

In order to ignore paths that unlikely contains useful information, the searching scope of the algorithm is limited to those paths whose target concepts is either a noun or an adjective. In this way, however, verbal target concept composing FDs like <cup> is used for drinking are discarded.

Filtering of candidate triples. The candidate triples obtained by the classifier are further filtered by discarding target concepts not belonging to WordNet (Fellbaum, 1998b) or belonging to the NLTK list of corpus stop-words (Bird and Loper, 2006). Subsequently, the remaining triples are ranked on the basis of the following weighted combination of the classifier score (SVM), pointwise mutual information (PMI: Church and Hanks, 1990) and log-likelihood (LL: Dunning, 1993):

$$ score(t) = \beta_{PMI} \cdot PMI(t) + \beta_{LL} \cdot LL(t) + \beta_{SVM} \cdot SVM(t) \qquad (7.1) $$

where all the scores are scaled to the interval $[0,1]$, and the classifier score is obtained by summing up the absolute values of the confidence scores of the single binary classifiers. Finally, from the resulting ranked list of *{relation, target_concept}* pairs the top 1,000 entries for every source concept are selected.

For identifying the best combinations of β values, the authors employed a ten-fold cross-validation on the 466 source concepts of their training set. Specifically,

Attribute Category	Example Attribute(s)
path-length	0.2
lemmatised source node	`turtle`
PoS of source node	NN
path labels from source (indexed)	GR1=dobjR GR2=ncmodR GR3=dobjR GR4=ncsubjN
path labels from target (indexed)	GR1=ncsubjR GR2=dobjN GR3=ncmodN GR4=dobjN
PoS of path nodes from source (indexed)	POS1=IN POS2=NNS POS3=VBP POS4=NNS
PoS of path nodes from target (indexed)	POS1=NNS POS2=VBP POS3=NNS POS4=IN
lemmatised path nodes (set)	`include`, `species`, `of`
PoS of all path nodes (set)	IN, NNS, VBP
Relation verbs	N/A
path labels (set)	ncsubjR, dobjN, ncmodN
lemmatised target node	`reptile`
PoS of target node	NNS

Table 7.1.2: An example vector for the path linking the concepts `turtle` and `reptile` in the sentence `Marine reptiles include five species of turtle`. Adapted from Kelly et al. (2012)

for each of the ten folds the authors applied the training steps to the triples in the development folds and the extraction phase on the concepts of the held-out fold. Subsequently, they calculated the F-measures of the performance obtained by vary-

ing the β values by 0.05 steps in the range $[0,1]$ and comparing the top 20 triples with those in the held-out set. The authors reported the best performance with high values for β_{SVM}, medium-low values for β_{LL} and very low values for β_{PMI}. Moreover, they reported significant lower β values for the association measures for the retrieval of *{source_concept, relation, target_concept}* triples than for the retrieval of *{source_concept, target_concept}* pairs.

Results. Kelly et al. (2012) employed their method for extracting both triples and pairs for three possible corpora settings: ukWac, Wikipedia and both. For comparing the performance of their model with that of their previous system (Devereux et al., 2009; Kelly et al., 2010), these authors extracted the top 20 triples and pairs for every concepts and evaluate the model against the ESSLLI dataset.

Table 7.1.1 reports the performance of their best configuration, that was obtained by exploiting *relation verb*-augmented vectors on a joint ukWac-Wikipedia corpus. Paired *t*-tests highlight a significant improvement of the performance obtained with this new model over their older proposal. By recognizing the limits of using the McRae norms as a gold standard, Kelly and colleagues further performed a manual evaluation on the triples extracted for 15 of their test concepts. According to their subjects, about 51.1% of their triples were either correct or plausible, as it was the case for 76.8% of the pairs.

7.2 EXTRACTING FDs FROM ITWAC

In order to test the feasibility of extracting Italian FDs-like semantic information for populating sMWN, we decided to implement several, slightly modified, versions of the method proposed by Kelly et al. (2012). Three main reasons motivate out choice to adopt this State-of-the-Art model as our starting point:

- it does not require manual encoding of lexico-syntactic patterns or rules;

- it is the best performing model on *{concept,relation,property}* triples, and the best on *{concept,property}* pairs, tied with Strudel;

- the corpora employed for its testing, i.e. ukWac and Wikipedia, have Italian counterparts, so to leverage the inevitable comparability issues due to the minor availability of resources and tools in languages other than English.

Our inquiry differs from the original study on several points. First, due to comparability issues, we decided to run our model solely on itWac, thus ignoring the linguistic material that can be extracted from the Italian Wikipedia. While the two WaCKy corpora have similar sizes, both about 2 billion words, the difference between the Italian and the English Wikipedias appears to be rather dramatic, [2], and the benefits of employing also the Italian Wikipedia dubious. We keep however this issue as a possible future extension.

Moreover, we exploited our norms both as a gold standard for the evaluation, both as a reference for extracting the training sentences from our corpora. This choice have two main consequences. First, by discarding complex target concepts from our normalized FDs collection, we obtained an average of 59.34 adjectival or noun descriptions per source concept, as opposed to the 14.7 triples of Kelly et al. (2012). This values raises up to 66.56 if verbal descriptions are included. Second, we collected FDs only for 50 concepts, so that we chose not to employ a 10-fold cross validation for setting the β values of the final score. Rather, we divided out full set of concepts in the following way:

- **Training Set**: 40 concepts used in the training phase and further split into:

  - *Development Set*. 30 concepts used for training the classifiers used for setting the β values.

  - *Held-out Set*. 10 concepts, 1 for each category, used for comparing models with different β values.

- **Test Set**. 10 concepts, 1 for each category, used for measuring the performance of the different models.

Finally, we employed the STaRS.sys classification of semantic relations described in Chapter 5. Our classification differs not only in the number of opposition encoded, 25 vs. 254, but also in the ontological nature of this distinctions. While our classes identify a type of information, those in the classification exploited by Kelly et al. (2012) encode a notion that is more similar to that of lexico-syntactic pattern.

---

[2]Apart from the number of articles, where the proportion is 4:1 ratio may appear not that significant, the most important difference concerns to the length of the articles. We weren't able to find relevant statistics on `http://stats.wikimedia.org/`, but we were impressed by the 8.5:1 ratio between the sizes of the relevant dumps.

The exact procedure we followed in the preparation and testing of our models articulates along the following steps. Where not differently stated, we followed the original proposal by Kelly et al. (2012), at least in what we understood to be the intentions of these authors.

7.2.1   TRAINING PHASE

Step 1: Training Sentences Extraction. From our normalized FDs dataset we collected every entry whose source concept was in our training set and whose target concept was linguistically simple. Every entry had been rewritten as a "*RELATION-couple*" of the kind RELATION*{source_concept, target_concept}*. For each relation we extracted from itWac all the sentences containing all the RELATION-couples from our norms. Let's call any such sentence a "*RELATION-sentence*"

Accordingly, then, we rewrote the FD <airplane> has wings as the RELATION-couple *HASCOMPONENT{airplane, wing}* and extracted from itWac all the sentences containing the lemmas airplane and wing.

Step 2: Preprocessing. The extracted RELATION-sentences has been re-lemmatized, PoS-tagged and morphologically analyzed with TextPro 1.5 (Pianta et al., 2008), a suite of NLP tools for the analysis of Italian and English texts, and parsed with deSR 1.2.6 (Attardi, 2006), a dependency parser that we trained on a manually created TextPro-compliant version of the dataset used in the Dependency Track of the Evalita 2011 Parsing Task (Bosco and Mazzei, 2011). Even if PoS tags were available in itWac, we re-tagged our sentences for obtaining the other levels of analysis (e.g. morphological information) useful for increasing the accuracy of the parser. Moreover, the re-tagging of the RELATION-sentences allowed us to cross-check for wrong lemmatizations and wrong sentence boundaries that have been removed.

Step 3: Path Extraction. For each RELATION-sentence we extracted the shorter path connecting the source and the target concepts in the dependency tree. Figure 7.2.1 compare the dependency tree generated by deSR for the sentence Il coniglio nano si trova difficilmente in libertà (*"it is unlikely for the dwarf rabbit to be found in the wild"*) with the nodes and relations retained by removing all the material not involved in the syntactic path linking the concepts coniglio and libertà.

Step 4: Classifier Training. From each dependency path we build a vector of attributes coding the lexico-syntactic properties of the path, for details on the imple-
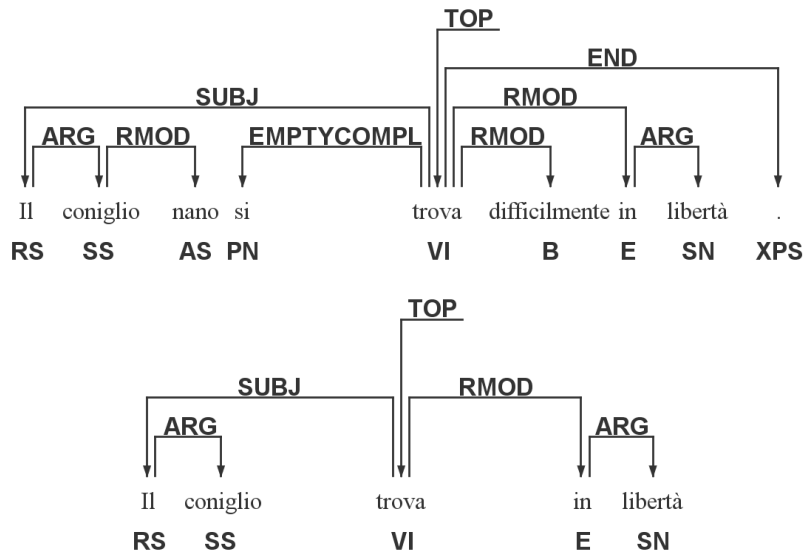
Figure 7.2.1: dependency trees generated by deSR for the sentence `Il coniglio nano si trova difficilmente in libertà` (above), and for the path linking the concept `coniglio` and `libertà` in the same sentence (below). Visualized trough DGAnnotator.

mented types of vectors see subsection 7.2.3. We then labeled each vector with the name of the semantic relation it is supposed to instantiate and exploited the whole set of collected vectors to train a Support Vector Machine (SVM: Vapnik, 1979) classifier by using SVM^{multiclass} 2.20 (Tsochantaridis et al., 2006), by setting all the parameters to default and the regularization parameter to 1.0.

7.2.2 Test Phase

Step 5: Test Sentences Extraction and Preprocessing. We extract from itWac all the sentences containing one of the 10 concepts in our test set. Let's call these sentences "*candidate-sentences*". Subsequently these candidate-sentences are preprocessed and parsed an in step 2.

Step 6: Candidate Paths Extraction. From each parsed candidate-sentence we extract a set of "*candidate-paths*" connecting the known source concepts with all the plausible target concepts in the sentence. In this context, we consider plausible target concept that is a noun or an adjective, or even a verb depending on the settings (see 7.2.3), that is part of a MWN synset and that is not present in the NLTK corpus stop-words for Italian (Bird and Loper, 2006).

In this step we decided to discard those paths that were too long to plausibly contain useful information. The maximum number of nodes for every path has been arbitrarily fixed to 10 by manually checking a sample of our CANDIDATE-sentences. Note that Kelly et al. (2012) didn't employ any such path length filter.

Step 7: Candidate Paths Classification. The CANDIDATE-paths are submitted to SVM^{multiclass} and classified according to the models in step 4. The output of the classifier is the labeling of each CANDIDATE-path with a relation label and a set of discriminant values for each of our 25 semantic relations. As in the original proposal, we summed over all the absolute values of these decision values and interpret the resulting value as a confidence score.

Step 8: Candidate Paths Clustering. The classified CANDIDATE-paths are rearranged as triples of the kind RELATION*{source_concept, target_concept}*. Identical triples obtained from different CANDIDATE-paths are clustered together, thus obtaining "CANDIDATE-*triples*" associated with a confidence score SVM(t), obtained by summing over all the confidence scores of the single paths. All summed confidence scored are then linearly scaled to the $[0,1]$ interval.

For speeding up the following passages, Kelly et al. (2012) ordered the CANDIDATE-triples on the basis of their SVM(t) value and selected the top ranked 1,000. We filter our CANDIDATE-triples in a different way in the next step.

Step 9: Association Scores Calculation. In this step we calculated from itWac the PMI and LL scores for every *{source_concept, target_concept}* pairs in our CANDIDATE-triples and normalize their values to the $[0,1]$ interval. Kelly et al. (2012) do not specify how they normalize their association scores. We calculated our normalized PMI by dividing the PMI value by the negative logarithm of the joint probability of the two items, as suggested by Bouma (2009), and rounding negative values to 0. LL scores have been normalized by rounding negative values to 0 and subtracting from 1 the corresponding *p*-value in a χ^2 distribution[3].

In calculating the association scores we apply a frequency threshold of $f \geq 10$, thus filtering out triples differently from Kelly et al. (2012). Considerations tied to the common practice in corpus linguistics and to the nature of the exploited association measures motivate our choice (see Evert, 2008).

Step 10: Total Scores Calculation and Final Selection. Every CANDIDATE-triple is associated with a total scores *score(t)* calculated by using the formula by Kelly et al.

---

[3]Thanks to Marco Baroni for the suggestion.

(2012), here reported in Equation 7.1. The list of CANDIDATE-triples is then ordered by *score(t)* and the top *n* ranked triples selected. When comparing this procedure against the results by Kelly et al. (2012), we set n = 20, while for the general analysis of the system performance we varied this value in the range [5,100], by steps of 5.

For estimating the βs exploited in the calculation of the total score *score(t)*, we employ a simpler strategy than the 10-folds cross-validation used by Kelly et al. (2012). We apply the training steps to the 30 concepts of the development set and the test phase to the remaining 10 concepts of the held-out set. We then run step 10 for all combinations of any β value in the range [0,1] (interval 0.05) and evaluated the top 20 pairs and triples per concepts against our norms. The βs of the configurations obtaining the highest F-measure values are saved and subsequently used for the final evaluation of the model.

7.2.3 THE IMPLEMENTED MODELS

Kelly et al. (2012) tested two kinds of attribute vector for coding source-target paths. "*Verb augmented*" vectors, encoding the presence of a relations verb, and "*non augmented*" vectors, lacking this information. In the light of the performance reported by these authors, we decided to implement only "*verb augmented*" vectors, by retrieving from our norms the set of 170 linking verbs reported in Appendix A.6.2.

Moreover, these scholar tested their model only on noun or adjectival target concepts. Given that the aim of our our experiment is the exploration of the possibility to automatically populate sMWN, and for keeping our results comparable with those reported for the original proposal, we decided to train and test all of our models both on sentences containing only noun or adjectival target concepts, both on sentences containing also verbal target concepts. In what follows, when relevant, the former models will be marked with "*[–v]*" and the latter with "*[+v]*".

We also explored the effects of exploiting also taxonomic and encyclopedic information in the recognition of semantic relations by encoding additional kinds of attributes in the vectorial representation of the linking path submitted to the classifier. The additional vector attribute we've tested are reported in Table 7.2.1.

These attributes can be divided into two broad classes. A first group of attributes, to with we will refer by means of the label "*+M*", has been designed to encode cate-

Attribute Category	Type	Example Attribute(s)
source norms category	**+M**	FRUIT
source MWN hypernyms	**+M**	{n#00009457}, {n#00010572}
source is hypernym of target	**+M**	false
target is hypernym of source	**+M**	true
target MWN hypernyms	**+M**	{n#00009457}
target MWN adj cluster	**+M**	—
target is in the 1st sentence of the source concept wikipedia article	**+W**	false
target is in the 2nd sentence of the source concept wikipedia article	**+W**	false
target is in the intro of the source concept wikipedia article	**+W**	true
link type if the target is a link in the source concept wikipedia article	**+W**	—

Table 7.2.1: Additional vector attributes tested in our models. Exemplar values refer to the path linking `ananas` and `frutto` in `L'ananas è il secondo frutto tropicale più consumato in Europa` (*"Pineapple is the second most popular tropical fruit"*). Taxonomic attributes are marked by "**+M**" (short for MultiWordNet), encyclopedic attributed by "**+W**" (short for Wikipedia).

gorical information about the source and the target concepts. The categorical nature of the involved concepts is a crucial aspect in the definition of some of our relations. As an example, the *Made of* relation holds between an object and a substance, e.g. `guitar` and `wood`. Accordingly, we expect this information to facilitate the discrimination between some of our relations.

For encoding taxonomic information we decided to rely on MWN. The only exception are the attribute types "*source norms category*", that associate the source concepts with its category following the organization designed by Kremer and Baroni (2011) and exploited also in our norms, and - indirectly - in all the experiments reported in this thesis.

The attributes related to the target concept receive a values only if it is either a Noun or an Adjective. Verbs are discarded because it is impossible to discriminate the kind of verbs that participate in our relevant feature types on the basis of their taxonomical properties. For each noun and adjectival target concept, we retrieve all the MWN synsets containing it, discarding those belonging to a different PoS. Let's call these retrieve synsets *"target synsets"*.

In the case of the adjectival target concepts, the attribute *"target MWN adj cluster"* encode the belonging of any of the target synsets to the adjectival clusters whose polar adjectives are reported in Table A.6.1, see Appendix A.6.3. This table lists the polar adjectives of all the clusters for which the summed frequency of their member in our norms is $f \geq 2$.

When the target concept is a noun, also the *"source synsets"*, i.e. the synsets containing the source concept lemma, are retieved. For all the source and the target synsets the hyperonym chain up to the root node is built and its properties encoded in the vector. If any of the target synsets is an hypernym of any of the source synsets, the attribute *"target is hypernym of source"* is marked as true. The feature *"source is hypernym of target"* encodes the specular configuration.

The *"source MWN hypernyms"* attributes, instead, encode which of the *"Relations Least Common Subsumers"* (RLCS) reported in Table A.6.2, see Appendix A.6.3, is present in the hyponymy chain of any target concept, if any. The RLCS list have been buil by collecting, for every relation in our norms, the least common subsumer, i.e. the most specific ancestor node, of the majority of the nominal target synsets.

"+W" attributes, in Table 7.2.1, have been developed to encode if the target concept is present in a meaningful position of the wikipedia article describing the source concept. Given the exploratory nature of the present study, we focused solely on the lead section and on the links.

Appendix A.6.4 shows the lead section for the Italian article *"Aeroplano"* and for the parallel article from the English Wikipedia. Many useful information can be extracted from this section, some of which are also present in our norms. As an example, è un mezzo di trasporto (*"it is a means of transport"*), sono dei velivoli (*"they are aircrafts"*), some hyponyms like bombardieri (*"bombers"*) or aerei da caccia (*"fighter"*) or some meronyms like motori (*"engines"*) or ala (*"wing"*). Note, moreover, how this information is represented as links pointing to other pages.

In order to evaluate if, and how much, the classifier would benefit from knowing if and were a target concept is represented in a Wikipedia article, we encoded in the vector also features specifying:

- if the target concept is in the very first sentence of the article describing the source concept. this may be a strong cue pointing to a *is-A* relation;

- if the target concept is in the second sentence of the article;

- if the target is present in the lead section at all;

- if a link in the source concept article points to an article describing the target concept. If so, the link type if specified, by choosing among five classes:

  - *intro links*: links appearing in the very first sentence;

  - *lead links*: links appearing in the lead of the article;

  - *textual links*: links appearing in the rest of the textual material of the link;

  - *list links*: links appearing in a list;

  - *appendix links*: links appearing in the "*See Also*" section of the article.

In the original Kelly vector, as shown in Table 7.1.2, the dependency relations and the PoS of the path nodes are encoded both in unordered way, by exploiting a bag-of-PoS and a bag-of-relations representation, both in ordered way, by keeping track of the linear order of the path. Note that the linear order of the path is obtained by moving from the source concept to the target, so that it could be different from the linear order of the words. We wanted to test the relative usefulness of these representation, so that we implemented two different kinds of vectors, the "*Bag of stuff*" and the "*Ordered*", obtained by employing just one of these two kinds of information.

Table 7.2.2 reports the differences between the different vector models we've tested. Each model have been tested both in a [−v] and in a [+v] setting, and its performance have been calculate by comparing the resulting *{source_concept, relation, target_concept}* triples and *{source_concept, target_concept}* against our norms.

Given the highest number of feature per concept in our norms than in those exploited by Kelly et al. (2012), a lower recall value is expected. We remain agnostic as far as the precision scores are concerned. While indeed the higher number of feature per concept should increase precision, this effect may be counterbalanced by

	Kelly	Kelly +W	Kelly +M	Kelly +WM	Bug of Stuff	Ordered
path labels from source	✓	✓	✓	✓		✓
path labels from target	✓	✓	✓	✓		
PoS of path nodes from source	✓	✓	✓	✓		✓
PoS of path nodes from target	✓	✓	✓	✓		
lemmatised path nodes	✓	✓	✓	✓	✓	
PoS of all path nodes	✓	✓	✓	✓	✓	
path labels	✓	✓	✓	✓	✓	
other Kelly attributes	✓	✓	✓	✓	✓	✓
Taxonomic attributes			✓	✓	✓	✓
Encyclopedic attributes		✓		✓	✓	✓

Table 7.2.2: Composition of the vector types employed in the different models. Blue colored ticks highlight the meaningful differences between the models. Attributes on the upper rows belong to the original Kelly model, see Table 7.1.2. Wiki and MWN attributes are described in Table 7.2.1.

the joint effects of (1) our smaller training dataset, (2) of our smaller number of described concepts and (3) of the different nature of the classification employed for labeling our norms.

7.3 RESULTS AND DISCUSSION

For evaluating the performance of the models in the different settings and comparing the results against Kelly et al. (2012), we selected the top ranked 20 *{source_concept, relation, target_concept}* triples and top ranked 20 *{source_concept, target_concept}* pairs for each model and calculated precision, recall and F-measure against our norms. Resulting values for triples are reported in Table 7.3.1, while Table 7.3.2 shows the performance obtained by ignoring the relation.

To get a more general view on the performance of the systems, we modulated the number of top concepts selected for each concept in the interval $[5, 100]$, by steps of 5. The general performance of the models is summarized by the Interpolated Precision/Recall curves in Appendix A.6.5, while the trends showed by the single mea-

Model	Setting	Precision	Recall	F-measure
Kelly	[−v]	**0.04**	**0.0149**	**0.0217**
	[+v]	0.03	0.0103	0.0154
Kelly +W	[−v]	0.03	0.0112	0.0163
	[+v]	0.03	0.0103	0.0154
Kelly +M	[−v]	0.03	0.0112	0.0163
	[+v]	0.03	0.0103	0.0154
Kelly +WM	[−v]	0.035	0.0131	0.019
	[+v]	0.02	0.0069	0.0103
Bug of Stuff	[−v]	**0.04**	**0.0149**	**0.0217**
	[+v]	0.01	0.0034	0.0051
Ordered	[−v]	0.0149	0.0056	0.0082
	[+v]	**0.035**	**0.0121**	**0.0179**

Table 7.3.1: Triples extraction performance of the models with the highest f-measure on triples, evaluated selecting the top 20 triples and comparing them against our norms. In the "*[−v]*" setting the model is trained and tested with paths containing only nominal and adjectival target concepts; it the "*[+v]*" setting also verbal target concepts are involved.

sures have been plotted in the figures of the Appendix A.6.6.

It is hard to state, by looking at these data, that some models have a clear advantage over the others, or that the encoding of some kind of information resulted in an marked improvement of the performance. What is striking, on the other side, is the low-performance of all models, in both settings. From a practical perspective, what can be concluded from this experiment is that, given the state of the art, the models presented here cannot be exploited as they are for automatically populating STaRS.sys without any supervision.

More interesting would be a discussion on the possible causes of these results. One possibility, already discussed by Baroni et al. (2008), Barbu (2008) and Kelly et al. (2010, 2012), pertains to the appropriateness of exploiting speaker-generated FDs collections as gold standards.

Properties can be expressed linguistically in many different ways. Aware of this

Model	Setting	Precision	Recall	F-measure
Kelly	[−v]	0.14	0.0563	0.0803
	[+v]	0.155	0.0579	0.0844
Kelly +W	[−v]	0.14	0.0563	0.0803
	[+v]	0.155	0.0579	0.0844
Kelly +M	[−v]	0.16	0.0644	0.0918
	[+v]	**0.17**	**0.0636**	**0.0925**
Kelly +WM	[−v]	0.165	0.0664	0.0947
	[+v]	0.165	0.0617	0.0898
Bug of Stuff	[−v]	**0.17**	**0.0684**	**0.0976**
	[+v]	0.155	0.0579	0.0844
Ordered	[−v]	**0.17**	**0.0684**	**0.0976**
	[+v]	**0.17**	**0.0636**	**0.0925**

Table 7.3.2: Pairs extraction performance of the models with the highest f-measure on triples, evaluated selecting the top 20 pairs and comparing them against our norms. In the "*[−v]*" setting the model is trained and tested with paths containing only nominal and adjectival target concepts; it the "*[+v]*" setting also verbal target concepts are involved.

problem, Baroni et al. (2008) created the ESSLLI dataset, an extended version of the McRae norms used for evaluating retrieval systems. This set has been built by generating the synonyms of the top 10 target concept for every concept in the McRae norms. In this context, it is worthwhile noticing that our norms are not and expanded set like the ESSLLI one, so that part of our lower performance with respect to the one reported by Kelly et al. (2012) can be accounted by this difference.

This problem is not restricted to the lexical forms of the target concepts of the norm. In this experiment we worked with sentences from corpora, that is, from sentences produced in a communicative context. As discussed by McRae et al. (2005), norms tend to represent distinctive properties of concepts, while in communicating we refer to any kind of property that is relevant for our communicative purposes.

Moving from similar positions, Kelly et al. (2012) manually evaluated a subset of their selected concepts. Their two subjects judged as at least plausible 51.1% of the

Model	Setting	Micro Average Precision	Micro Average Recall	Micro Average F-measure	Macro Average F-measure
Kelly	[−v]	0.2295	0.1198	0.1612	0.2869
	[+v]	0.2953	0.1039	0.1313	0.116
Kelly +W	[−v]	0.3112	0.1221	0.2061	0.2222
	[+v]	0.2492	0.1104	0.1579	0.1580
Kelly +M	[−v]	0.253	0.1703	0.1934	0.289
	[+v]	0.4387	0.1257	0.1929	0.1887
Kelly +WM	[−v]	0.35	0.1524	0.2598	0.2512
	[+v]	0.3255	0.135	0.1993	0.1891
Bug of Stuff	[−v]	0.4135	0.1476	0.2543	0.3319
	[+v]	0.3266	0.1368	0.159	0.2462
Ordered	[−v]	0.3761	0.1381	0.1651	0.2394
	[+v]	0.3337	0.1126	0.1581	0.133

Table 7.3.3: Performance of the different classifiers in the discrimination of the sentences containing the *{test_source_concept, target_concept}* pairs available in our norms.

triples and 76.8% of the pairs, thus suggesting that the human judgment may be a more suitable assessment method. We leave to the future a similar evaluation, but the impression that the evaluation conducted against our norms may have underestimated the real performance of our models can be gained also from a quick analysis of the sample pairs reported in Appendix A.6.7 for the "*[−v] Kelly*" model and in Appendix A.6.9 for the "*[−v] Kelly +WM*" model.

The sample triples reported in Appendix A.6.8 for the "*[−v] Kelly*" model and in Appendix A.6.10 for the "*[−v] Kelly -WM*" model suggest a rather different picture, instead. The main problem seems to be the performance of the classifier itself.

For testing this hypothesis we evaluated, for every model, the performance of the classifier alone in the discrimination of the itWac sentences containing all the *{test_source_concept, target_concept}* pairs from our norms.

The results of this evaluation, here reported in Table 7.3.3, seem to support our

concerns about the performance of the classifier. Part of the wrong classifications, obviously, can be the consequence of independent errors made in the preprocessing or in the parsing phase. The task itself, the identification of semantic relations holding between two concepts, is a difficult task for a computer, as has been discussed in the first section of this chapter. However, we ascribe these results also to two critical limitations affecting the whole procedure described here.

We see the first problem as somehow connected to the different ontological status of our relations, as opposed to the "*quasi-predicative*" nature of the notion of relation employed by Kelly et al. (2012). Adopting the perspective of these authors, indeed, it is a reasonable to treat as instantiation of the same relation all the sentences containing a *{source_concept, target_concept}* pair extracted from the norms. Note that these are the sentences on which the classifier is trained.

This assumption, however, is more problematic if we adopt a notion of semantic relation that is less linguistically based, as we have done in STaRS.sys. We observed that about 1/10 of the training pairs used for training the classifier were repeated across different relations, as a consequence of the fact that 399 of the 3433 concepts pairs in our norms have an homologous pairs labeled with a different relation. It means that the classifier has no possibility to distinguish between the paths for the 10% of the training pairs. Note that this number is a low estimate, because it is based on the assumption that the rest of the training paths are perfect examples.

The second crucial limitation of our procedure is the fact that it does not possess an efficient way to filter out path that do not instantiate any relation. As an example, in the model proposed by Poesio and Almuhareb (2005) a binary classifier is responsible to the identification of the candidates attributes that should be later labeled by another 5-way classifier in charge of categorizing the attribute.

Association measures are employed in this procedure "*to assess the relative saliency of each extracted feature*". However, these are, as a matter of fact, more apt "*to quantify the attraction between cooccurring words*" Evert (2008).

Many ameliorations to improve the performance of the procedure described in this chapter are conceivable. Apart from a more accurate selection of the training paths and the implementation of a filter, preliminary results suggest that the classification of the paths may benefit from the exploitation of a kernel like the tree kernel by Moschitti (2004). Moreover, information from Wikipedia can be extracted in a more systematic and aware fashion, checking the content of the article more care-

fully and even the content of the articles pointing to the source concept article.

The automatic extraction of FD-like semantic information is a difficult task. After all, it is a demanding task even for a human speaker. Even if many improvements are needed, by looking at the results reported here from a practical perspective, and especially at the nature of the *{source_concept, target_concept}* pairs extracted, there's the feeling that the output of an automatic system like the one described in these pages can be used for populating STaRS.sys. A heavy manual check will be needed, but this would be a less demanding and more time-saving process than the collection of a huge dataset of speaker-generated Featural Descriptions.

8

Conclusions

THE WORK PRESENTED HERE EXPLORED THE POSSIBILITY OF INTERPLAY BETWEEN NEUROPSYCHOLOGICAL NEEDS, PSYCHOLOGICAL THEORIES OF SEMANTIC MEMORY AND NATURAL LANGUAGE PROCESSING TECHNIQUES. We moved from a specific practical need, that is the development a software system for assisting the therapist during the preparation of a semantic task for an anomic patient. This allowed us to adopt a different, somehow privileged, perspective on two classical NLP issues: the identification of a motivated set of semantic relations and the encoding of commonsense knowledge in a machine readable format. A third NLP-related issue explored in this thesis has been the automatic extraction of commonsense knowledge.

8.1 SUMMARY AND ACHIEVEMENTS

Chapter 2 reviews the needs and the most common therapeutic practices for the treatment of naming disorders, while chapter 3 discusses the possible uses of com-

puters in the this context. Taken together, these two chapters have been thought to give the reader a complete view on the needs of our final user, i.e the therapist, and for illustrating the innovative aspects of our work.

To the best of our knowledge, indeed, no system like ours is to-date available. Most of the existing therapeutic softwares rely on collections of prepared tasks, some of which have been proven to be very effective for the therapeutic treatment. STaRS.sys, on the other side, has been conceived to fulfill another need of the therapist: the need to have access to a cognitively-modeled extensive knowledge base where commonsense semantic knowledge is explicitly represented and can be accessed in a fairly intuitive, near-to-language-use, way.

Taken together, chapters 4, 5 and 6 describe what are the requirements of such a knowledge base and how we developed "*StarsMultiWordNet*" (sMWN) in order to meet them. Our work is centered on the key notion of "*speaker generated Featural Description*", that is, short linguistic description produced by a speaker to describe a concept.

In chapter 5 we propose and evaluate a classification of all the kinds of properties that are typically associated with concrete objects in producing FDs. In chapter 6 we exploit this classification for structuring an elicitation experiment that tries to overcome some of the structural limitations of the traditional "*feature norms*" paradigm.

As discussed and motivated throughout these chapters, the sMWN database has been developed as an extension of the traditional WordNet model, improved by implementing (1) an extended set of semantic relations and (2) several structural modifications motivated by the linguistic properties of the FDs produced by our speakers.

In the psychological tradition, feature norms are represented as lists of pairs or as matrices. Some scholars openly criticized this practice and its implicit assumptions (e.g. Barbu, 2008). By implementing a network-like structure, our knowledge base can be seen also as a new model for representing the semantic information that can be collected by exploiting this experimental paradigm.

In the last chapter we investigate the possibility to automatically populate sMWN by collecting evidence from an Italian corpus. As expected, this revealed to be a rather difficult task, and the performance were quite low. Some reasons behind this difficulty were identified and analyzed.

8.2 Conceivable Improvements

Given its exploratory nature, the project STaRS.sys has many limitations. The most obvious is that it focuses solely on concrete objects. It is self evident that the classification described in Chapter 5 is not apt to describe the kind of properties that can be associated with predicates like "*to see*", or with abstract concepts like "*abstraction*" or "*concept*". It has been thought for, and limited to, properties possessed by concrete concepts. In a similar fashion, we collected FDs by asking our speakers to describe only concrete concepts, and the linguistic forms employed in these descriptions motivated some of the modifications to the WordNet model implemented in sMWN.

Accordingly, an easily conceivable improvement is the extension to cover other kinds of concepts. This extension, we foresee, would probably require a novel restructuring of the whole STaRS.sys model, and probably of the whole work done so far.

As frequently reported in these pages, the feature norm paradigm has been developed in psychology for collecting salient properties. The lexicographic and NLP exploitation of these resources, however, has to face many issues as a consequence of their lack of completeness. Our idea to elicit descriptions by submitting questions has been thought to increase the number and variety of descriptions per concept generated by our speakers. As such, it can be seen as an amelioration of the feature norm paradigm, at least for lexicographic purposes.

Frassinelli and Lenci (2012) showed that some property types are more likely to be produced if the source concept is presented to the speaker in a certain context. Such finding can be exploited to manipulate this variable to collect a wider range of properties.

Often FDs are filtered on the basis of their frequency of production. Again, this may not be optimal from a lexicographic perspective. Maybe better data can be obtained from other methods for selecting the appropriate descriptions. An example can be the employment of a feature generation task, i.e. a task in which participants are asked to judge the descriptions generated by other speakers. This task can be used also for creating different kinds pf semantic data, like those in the "*exemplar by feature applicability matrices*" by De Deyne et al. (2008), whose usability has been tested and proved in the works collected by Storms et al. (2010).

Many aspects of the work presented here move from the psycholinguistic evi-

dence the is available today. A clear example is the literature reviewed in order to identify the kinds of information that our classification discriminates. As such, the development of STaRS.sys has been based on assumptions and hypotheses that may well reveal to be inaccurate in the future, because of the normal evolution of ideas, models and methods in science. Certainly refinements, maybe even radical changes, will be needed in the future. For the author, however, the central achievement of this project has been to illustrate the benefits that can arise from the interplay between Natural Language Processing and psycholinguistic.

A
Appendix

A.1 Sample of the CeRiN target-attribute pairs

Category	Type	v/n	Target	Attributes
ANIMAL	amb	n	cammello	vive nel deserto
ANIMAL	amb	n	canguro	vive in Australia
ANIMAL	amb	n	cervo	vive nei boschi
ANIMAL	amb	n	gatto	è un animale da appartamento
ANIMAL	amb	n	maiale	è un animale da fattoria
ANIMAL	cat	n	cane	è un mammifero
ANIMAL	cat	n	coccodrillo	è un rettile
ANIMAL	col	v	aragosta	ha un corpo di colore rosso vivo
ANIMAL	col	v	cervo	ha un mantello marrone
ANIMAL	col	v	leone	ha un mantello giallo-marrone
ANIMAL	col	v	leopardo	è giallo con macchie nere
ANIMAL	col	v	maiale	ha la pelle rosea
ANIMAL	dim	v	cavallo	è un animale grande
ANIMAL	dim	v	elefante	è un animale molto grande
ANIMAL	dim	v	lumaca	è un animale piccolo
ANIMAL	enc	n	aragosta	la sua carne è costosa
ANIMAL	enc	n	asino	è testardo
ANIMAL	enc	n	cammello	conserva l'acqua nella gobba
ANIMAL	enc	n	cane	è fedele all'uomo
ANIMAL	enc	n	foca	mangia i pesci
ANIMAL	enc	n	gatto	miagola
ANIMAL	enc	n	lumaca	si muove molto lentamente
ANIMAL	enc	n	mucca	è allevato per la carne ed il latte
ANIMAL	enc	n	orso	ama il miele
ANIMAL	enc	n	tigre	è un carnivoro
ANIMAL	enc	n	topo	è cacciato dai gatti
ANIMAL	enc	n	volpe	è molto furba
ANIMAL	morf	v	aragosta	ha una corazza molto dura
ANIMAL	morf	v	asino	ha orecchie lunghe e dritte
ANIMAL	morf	v	cammello	ha due gobbe sulla schiena
ANIMAL	morf	v	canguro	ha un marsupio per i piccoli
ANIMAL	morf	v	capra	ha la barba
ANIMAL	morf	v	cavallo	ha la criniera
ANIMAL	morf	v	cervo	ha grandi corna
ANIMAL	morf	v	coccodrillo	ha una grande bocca e denti aguzzi
ANIMAL	morf	v	foca	ha le zampe palmate
ANIMAL	morf	v	giraffa	ha un collo molto lungo e sottile

Continued on next page

Appendix A.1 *(Continued from previous page)*

Category	Type	v/n	Target	Attributes
CLOTHING	cat	n	calzino	è un capo di vestiario
CLOTHING	cat	n	sandalo	è un tipo di scarpa
CLOTHING	enc	n	berretto	si indossa in testa
CLOTHING	enc	n	calzino	copre il piede
CLOTHING	enc	n	cinta	si indossa intorno alla vita
CLOTHING	mat	v	berretto	è di stoffa
CLOTHING	mat	v	cinta	è di pelle
CLOTHING	morf	v	berretto	spesso ha una visiera
CLOTHING	morf	v	camicia	ha le maniche
CLOTHING	morf	v	cinta	ha una fibbia
FOOD	amb	n	mela	cresce sugli alberi
FOOD	cat	n	limone	è un frutto
FOOD	cat	n	mela	è un frutto
FOOD	col	v	banana	è gialla
FOOD	col	v	carota	è arancione
FOOD	col	v	limone	è giallo
FOOD	enc	n	banana	si sbuccia prima di mangiarla
FOOD	enc	n	caramella	piace molto ai bambini
FOOD	enc	n	carota	fa bene agli occhi
FOOD	enc	n	cipolla	fa lacrimare quando si taglia
FOOD	morf	v	banana	è di forma allungata e curva
FOOD	morf	v	carota	finisce a punta
FOOD	morf	v	cipolla	è tondeggiante
FOOD	morf	v	noccioline	hanno un guscio rugoso
FOOD	morf	v	pera	ha una parte tonda e una appuntita
TOOL	amb	n	spillatrice	si usa in ufficio
TOOL	amb	n	spugna	si può trovare in bagno
TOOL	cat	n	martello	è un utensile
TOOL	cat	n	mestolo	è un utensile da cucina
TOOL	cat	n	spago	è una specie di corda
TOOL	col	v	colla	può essere bianca
TOOL	col	v	scotch	è trasparente
TOOL	enc	n	asciugamano	è morbido
TOOL	enc	n	catena	si può usare per chiudere un cancello
TOOL	enc	n	chiodo	si pianta col martello
TOOL	enc	n	lampadina	si illumina quando si accende
TOOL	enc	n	penna	contiene inchiostro
TOOL	funz	n	asciugamano	si usa per asciugarsi
TOOL	funz	n	attache	si usa per tenere insieme fogli di carta

Category	Type	v/n	Target	Attributes
TOOL	funz	n	barattolo	si usa per conservare il cibo
TOOL	funz	n	bottiglia	contiene i liquidi
TOOL	funz	n	cucchiaio	si usa per mangiare la minestra
TOOL	mat	v	asciugamano	è di spugna
TOOL	mat	v	attache	è di metallo
TOOL	mat	v	barattolo	è di vetro
TOOL	morf	v	barattolo	ha un coperchio
TOOL	morf	v	bottiglia	si chiude con un tappo
TOOL	morf	v	catena	ha anelli collegati l'uno all'altro
TOOL	morf	v	chiodo	è appuntito
TOOL	morf	v	chiodo	ha una testa appiattita

Table A.1.1: The v/n opposition stands for the dichotomy between visual and non-visual (kinds of) attributes. Associated type codes are to be understood as follows: col → color, dim → dimension, mat → matter, morf → morphology, amb → natural environment, cat → taxonomic category, funz → fuction, enc → other non-visual encyclopedic features

A.2 DEFINITION OF THE STARS.SYS DIRECT FEATURE TYPES

Relation	Definition	Part of Speech	default Conj/Disj
TAXONOMIC PROPERTIES			
is-A	This relation holds between a concept and a class to which it belongs	N ↔ N	AND
Coordination	This relation holds between two concepts that are similar in belonging to the same category	N ↔ N	AND
PART-OF			
has Component	This relation encodes the relationship between an integral object and one of its components	N ↔ N	AND [inv XOR]
has Member	This relation holds between a concept denoting a group or a collection and one of its members. Members are entities that keep their individuality with respects to the other member of the group to which they belong	N ↔ N	AND [inv XOR]
has Portion	This relation holds between a mass object and one of its portions	N ↔ N	AND [inv XOR]
Made of	This relation holds between an object and the substance that composes it	N ↔ N	AND [inv XOR]
has Geographical Part	This relation encodes a part-of relationship between two geographical areas	N ↔ N	AND [inv XOR]
PERCEPTUAL PROPERTIES			
has Size	This relations holds between a concept and its typical size	N ↔ Adj *N ↔ N	XOR

Relation	Definition	Part of Speech	default Conj/Disj
has Shape	This relations holds between a concept and its typical shape	N ↔ Adj *N ↔ N	XOR
has Taste	This relations holds between a concept and its typical taste	N ↔ N *N ↔ Adj	XOR
has Smell	This relations holds between a concept and its typical smell	N ↔ Adj *N ↔ N	XOR
has Sound	This relations holds between a concept and the typical sound it emits	*N ↔ N N ↔ N *N ↔ Adj	XOR
has Color	This relations holds between a concept and its typical color	N ↔ Adj *N ↔ N	XOR
has Texture	This relations holds between a concept and the typical appearance or consistence of its surface or of the substance that composes it	N ↔ Adj	AND

USAGE PROPERTIES

Relation	Definition	Part of Speech	default Conj/Disj
is Used for	This relation holds between a concepts and an action that is accomplished by using it (also metaphorically)	N ↔ V *N ↔ N	OR
is Used by	This relation holds between a concept and an actor that typically uses it	N ↔ N	OR
is Used with	This relation holds between two concepts that are typically used together	N ↔ N	OR

CONTEXTUAL PROPERTIES

Relation	Definition	Part of Speech	default Conj/Disj
Situation Located	This relation holds between a concept and a typical situation in which it is usually found	N ↔ N	OR
Space Located	This relation holds between a concept and a typical place in which it is usually found	*N ↔ V N ↔ N	OR

Relation	Definition	Part of Speech	default Conj/Disj
Time Located	This relation holds between a concept and a time period, that is any portion of the time line, that is strongly associated with it	N ↮ N	OR
has Origin	This relation holds between a concept and the agent that typically produces it, the typical place where it is born or where it grows or another concepts that originates it	N ↮ N	OR

ASSOCIATED EVENTS AND ATTRIBUTES

Relation	Definition	Part of Speech	default Conj/Disj
has Affective Property	This relation holds between a concept and an emotional (affective) state aroused by it	N ↮ N *N ↮ Adj	AND
has Attribute	This type encodes a generic relationship holding between a concept and one of its permanent properties other than the perceptive, taxonomic, part-of or affective ones	N ↮ Adj *N ↮ N	OR
is Involved in	This type encodes a relation holding between a concept and an action or process in which the concept is involved with are role other than those encoded by means of the contextual or the usage relations	N ↮ V *N ↮ N	OR
is Associated With	This type encodes a generic association between two concepts. It is used to encode a relationship that cannot be described by any of the other STaRS.sys types	N ↮ N	AND

Table A.2.1: Rare and exceptionally acceptable **Part of Speech** patterns are marked by an "*asterisk". Typical exceptional cases are those involving metaphoric language (e.g. it has the color of the sea). When the **default Conjunction/Disjunction** relation feature value of the inverse relation (e.g. XOR for *is a Component of*) is different from that of the direct relation (e.g. AND for *has Component*), it is reported between [inv square brackets]

A.3 Comparison tables between STaRS.sys and other classifications

STaRS.sys	CeRiN	LTBC	SFA	WB	CMᵇʳ
Taxonomic Properties					
is-A	Taxonomic Category	Superordinate across/within categories	Group	Cʜ	Taxonomic
Coordination				Cᴄ	Taxonomic
Part-of					
has Component	Morphology	Structural subordinate		(-)Eᴡ, Eᴄᴇ, Eᴄɪ	Visual Form and Surface Prop.
has Member					
has Portion	Morphology				
Made of	Matter			(-)[cm]Made of	Visual Form and Surface Prop.
has Geographical Part					
Perceptual Properties					
has Size	Dimension		Properties	Eꜱᴇ, Eꜱɪ	Visual Form and Surface Prop.
has Shape	Morphology		Properties	Eꜱᴇ, Eꜱɪ	Visual Form and Surface Prop.
has Taste			Properties	Eꜱᴇ, Eꜱɪ	Taste

Appendix A.3 *(Continued from previous page)*

STaRS.sys	CeRiN	LTBC	SFA	WB	CMᵇʳ
has Smell			Properties	Ese, Esi	Smell
has Sound			Properties	Ese, Esi	Sound
has Color	Color		Properties	Ese, Esi	Visual Color
has Texture	Morphology		Properties	Ese, Esi	Visual Form and Surface Prop., Tactile
Usage Properties					
is Used for	Function	Functional subordinate -use	Use	Sꜰ	Function
is Used by				∼Sᴘ	
is Used with					
Contextual Properties					
Situation Located	Natural Environment			∼Sᴇᴠ	
Space Located	Natural Environment	Functional subordinate -place	Location	Sʟ	
Time Located				Sᴛ	
has Origin	Natural Environment	Functional subordinate -place		Sᴏʀ	
Associated Events and Attributes					
has Affective Property				Iᴀ	

Continued on next page

128

Appendix A.3 *(Continued from previous page)*

STaRS.sys	CeRiN	LTBC	SFA	WB	CM^br
has Attribute				E$_{SYS}$	
is Involved in			∼Action	E$_B$, S$_A$	
is Associated with			Association	E$_{AE}$, S$_E$	

Table A.3.1: Comparison table between STaRS.sys and (neuro)cognitive feature type classifications. Inverse relations are marked by "(-)", while "∼" marks similar, but not identical, feature types. Header labels are to be understood as follows: **CeRiN**: the attribute type classification adopted in the concept-attribute pairs used in the CIMeC Center for Neurocognitive Rehabilitation (see Appendix A.1); **LTBS**: the classification exploited in the Laiacona et al. (1993b)'s semantic questionnaire; **SFA**: the classification of description types exploited in Semantic Feature Analysis (Boyle and Coelho, 1995); **WB**: ?WuBarsalou09)'s knowledge-type taxonomy (the superscript " [cm]" marking the types introduced by Cree and McRae (2003), if relevant); **CM^br**: Cree and McRae (2003)'s brain knowledge type taxonomy (not to be confused with the modified WB taxonomy).

STaRS.sys	EWN/WN	WCH	EQS
Taxonomic Properties			
is-A	Has Hyperonym		Isa
Coordination			
Part-of			
has Component	Has Mero Part	(-) Component/Integral Obj.	(-) is a part of
has Member	Has Mero Member	(-) Member/Integral Obj.	(-) is a member of
			(-) is a follower of
has Portion	Has Mero Portion	(-) Portion/Mass	
Made of	[ewn]Has Mero Madeof	(-) Stuff/Object	(-) Made of
has Geographical Part	[ewn]Has Mero Location	(-) Place/Area	
Perceptual Properties			
has Size, has Shape, has Taste, has Smell, has Sound			
has Color			Has as Colour
has Texture			
Usage Properties			
is Used for	[ewn]Role Instrument		Used for, Used against, Used as
is Used by			Used by

130

Appendix A.3 *(Continued from previous page)*

STaRS.sys	EWN/WN	WCH	EQS
is Used with			
CONTEXTUAL PROPERTIES			
Situation Located			
Space Located			LOCATION
Time Located			
has Origin			Produced by
ASSOCIATED EVENTS AND ATTRIBUTES			
has Affective Property			
has Attribute			
is Involved in	[ewn]ROLE AGENT, [ewn]ROLE PATIENT		Constitutive Activity, Produces, Object of the Activity, Activity, (-)Concerns
is Associated with	[ewn]FUZZYNYM, [wn]EVOCATION		Related to

Table A.3.2: Comparison table between STaRS.sys and other semantic relations classifications. Inverse relations are marked by "(-)", while "∼" marks similar, but not identical, types. Header labels are to be understood as follows: **EWN/WN**: Semantic and lexical relations implemented in EuroWordNet (Alonge et al., 1998) and in WorNet (Fellbaum, 1998b). The superscripts " [ewn]" and " [wn]" labels mark the relations available solely in one of the two resources. Note furthermore that the "Evocation" relation (Boyd-Graber et al., 2006) is not yet implemented in the default WN distribution; **WHC**: taxonomy of part-whole relation proposed by Winston et al. (1987); **EQS**: SIMPLE/PAROLE Extended Qualia Structure (Lenci et al., 2000).

Feature Type	Agreement	
	Annotators (Fleiss' κ)	Gold/Majority (κ)
is-A	0.876	0.98
Coordination	0.808	0.955
has Component	0.783	0.948
Made of	0.811	1
has Size	0.971	1
has Shape	0.803	0.946
has Taste	0.9	0.955
has Smell	0.839	1
has Sound	0.756	1
has Color	0.925	1
has Texture	0.605	0.722
is Used for	0.866	0.941
is Used by	0.885	0.955
is Used with	0.79	1
Situation Located	0.516	0.872
Space Located	0.879	0.958
Time Located	0.735	0.946
has Origin	0.719	0.855
has Affective Property	0.193	$p > 0.05$
has Attribute	0.494	0.82
is Involved in	0.470	0.85
is Associated with	0.225	0.427
general	0.721	0.915

Table A.4.1: STaRS.sys Classification Evaluation: Type-wise agreement values

Feature Type Class	Annotators' Agreement (Fleiss' κ)
Taxonomic Properties	0.881
Part-of	0.849
Perceptual Properties	0.837
Usage Properties	0.887
Contextual Properties	0.824
Associated Events & Attributes	0.562
general	0.764

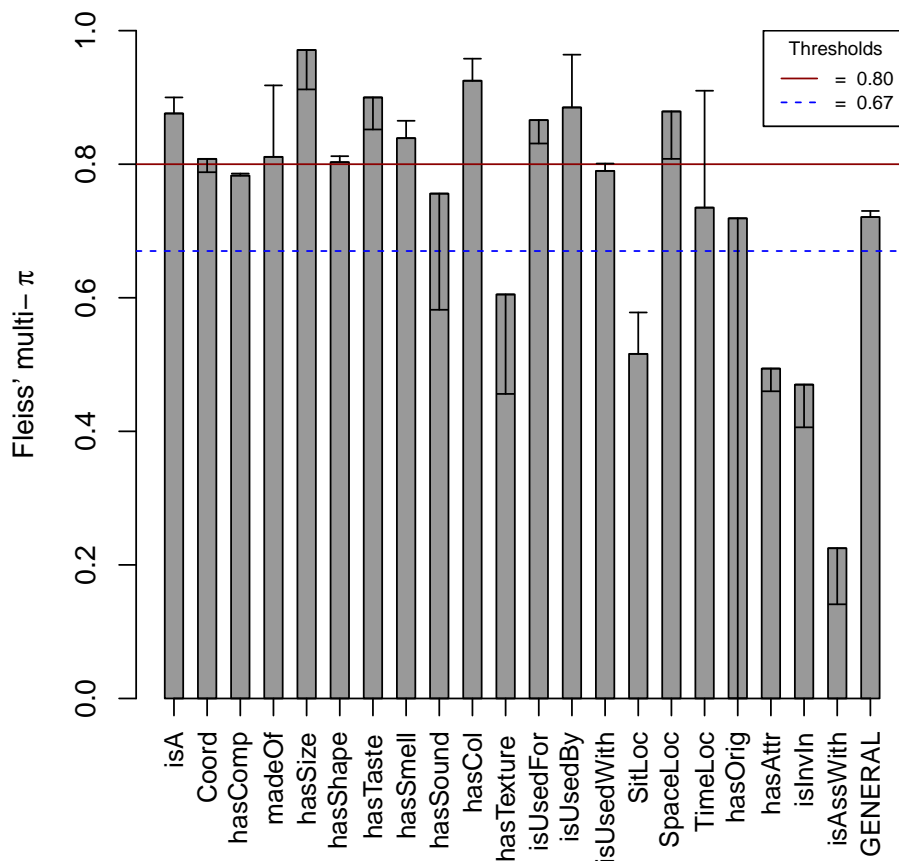**Table A.4.2:** STaRS.sys Classification Evaluation: Class-wise agreement values

Figure A.4.1: Evaluation: Annotators Type-Wise Agreement (bars indicate agreement from Lebani and Pianta (2010b)

A.5 Supplementary material for chapter 6

Figure A.5.1: Collection: web interface [original]

STaRS.sys

Semantic Task Rehabilitation Support system

STaRS.sys Feature Generation

STaRS.sys project: feature norms collection

0% ▣▣▣▣▣▣▣▣ 100%

Task

We are asking you to describe a set of concrete concepts by answering to some questions. Your answers have to be as simple and concise as possible. Separate your descriptions by "; " (*semicolon + blank space*), and do not report any biographic info or technical knowledge. If a question does not sound appropriate for a given concept, or if you don't come up with an appropriate answer, you're free to leave the field empty.

At any time, you can access the examples available in the online documentation by clicking on the text of the questions.

CHERRY

which *Category* does it belong to?

what *Kinds* of cherries are there?

what are the *Parts* of a cherry?

what is it a *Part of*?

what is it a *Member of*?

of what is it a *Portion*?

what *Substance* is it made of?

what does it *Compose*?

what's its *Size*?

what's its *Shape*?

what's its *Taste*?

what's its *Sound*?

what's its *Smell*?

what's its *Colour*?

what's its *Texture* ?

what is it *Used for* ?

by *Whom* is it typically used?

With What is it used?

Where can it be found?

in what *Situation* can it be found?

When can/could it be found?

what are its *Origins*?

what *is Produced By/From*
or *Grows From* it?

what *Emotion* is associated with it?

to what is it *Similar*?

in what *Events* is it involved in?

what are its other *Attributes*?

what can be *Associated with* it?

Others...

[Resume later]   [<< Previous Concept] [Next Concept >>]

[Exit and clear survey]

Figure A.5.2: Collection: web interface [translated]

135

	TAXONOMIC	PART OF	PERCEPTUAL	USAGE	LOCATIONAL	EVENTS & ATTR.	ASSOCIATION	total
BIRD	218	265	314	54	227	204	90	1372
BODY PART	147	378	378	293	53	148	130	1527
BUILDING	227	351	318	254	341	147	107	1745
CLOTHING	305	290	383	341	302	234	92	1947
FRUIT	243	351	489	190	383	186	92	1934
FURNITURE	261	365	367	335	387	157	92	1964
IMPLEMENT	250	236	327	356	238	150	108	1665
MAMMAL	329	290	426	271	276	323	107	2022
VEGETABLE	244	206	390	257	330	163	82	1672
VEHICLE	304	355	373	316	321	221	121	2011
total	2528	3087	3765	2667	2858	1933	1021	17859

Table A.5.1: Collection: frequency of type classes in the raw FDs

	TAXONOMIC	PART OF	PERCEPTUAL	USAGE	LOCATIONAL	EVENTS & ATTR.	ASSOCIATION	total
BIRD	172	244	241	34	145	121	42	999
BODY PART	127	328	258	194	13	47	49	1016
BUILDING	132	263	235	138	182	45	31	1026
CLOTHING	203	237	257	261	190	131	39	1318
FRUIT	189	271	390	127	268	105	35	1385
FURNITURE	180	307	266	230	262	87	41	1373
IMPLEMENT	175	180	225	233	127	82	48	1070
MAMMAL	272	245	317	153	167	190	43	1387
VEGETABLE	181	105	290	161	225	96	29	1087
VEHICLE	216	289	251	238	207	115	50	1366
total	1847	2469	2730	1769	1786	1019	407	12027

Table A.5.2: Collection: frequency of type classes in the selected FDs

	TAXONOMIC	PART OF	PERCEPTUAL	USAGE	LOCATIONAL	EVENTS & ATTR.	ASSOCIATION	total
BIRD	36	49	61	12	46	46	14	264
BODY PART	23	66	60	71	4	34	22	280
BUILDING	44	70	68	50	60	22	15	329
CLOTHING	57	46	78	65	58	52	22	378
FRUIT	43	61	83	42	89	31	17	366
FURNITURE	46	61	79	61	81	34	20	382
IMPLEMENT	57	34	62	69	38	37	23	320
MAMMAL	66	66	67	51	59	70	17	396
VEGETABLE	47	30	68	59	77	33	12	326
VEHICLE	59	68	70	73	63	44	25	402
total	478	551	696	553	575	403	187	3443

Table A.5.3: Collection: frequency of type classes in the normalized FDs

A.6 Supplementary material for chapter 7

A.6.1 List of relations in the Kelly et al. (2012)

List of relations generated by Kelly et al. (2012) from the British English version of the McRae norms by Taylor et al. (2011). From a personal communication with Colin Kelly.

becomes, bought in, branches, builds, carries, causes, chases, chews on, climbs, comes, comes from, comes in, comes on, cuts, different, digs, does, doesn't, dries, drinks, dropped, eaten, eaten as, eaten at, eaten by, eaten for, eaten in, eaten on, eaten with, eats, eg, fans, fires, flies, flows, found, found at, found below, found beside, found by, found in, found near, found on, found over, gets, gives, goes, grows, grows in, grows on, has, hasn't, hates, herded, holds, hunted by, is, isn't, juts, killed, king, launched, lays, likes, lives by, lives in, lives near, lives on, loses, made by, made from, made in, made of, made with, makes, owned, part, pollinates, pops, pricks, produces, puts, requires, runs, sees, shoots, sits, sleeps, smells, sold in, sounds, spins, spouts, sprays, stands, sticks, stores, strangles, suckles, sucks, surrounds, swims, swings, symbol, tastes, tells, travels in, used, used a, used as, used at, used by, used by blowing, used by connecting, used by firing, used by hanging, used by moving, used by pulling, used by riding, used by throwing, used for, used for aerating, used for attaching, used for avoiding, used for blowing, used for boiling, used for breaking, used for browning, used for building, used for burning, used for buying, used for calling, used for carrying, used for catching, used for chewing, used for chopping, used for cleaning, used for closing, used for colouring, used for connecting, used for controlling, used for cooking, used for cooling, used for covering, used for cutting, used for delivering, used for digging, used for dispensing, used for dividing, used for diving, used for docking, used for draining, used for drawing, used for eating, used for enlarging, used for ensuring, used for expelling, used for feeding, used for freezing, used for getting, used for grating, used for heating, used for hitting, used for holding, used for housing, used for inflicting, used for keeping, used for killing, used for listening, used for loosening, used for lying, used for making, used for measuring, used for mixing, used for moving, used for opening, used for ordering, used for performing, used for picking, used for playing, used for pounding, used for preserving, used for preventing, used for producing, used for protecting, used for providing, used for prying, used for pulling, used for putting, used for regulating, used for removing, used for repelling, used for resting, used for riding, used for sealing, used for seating, used for seeing, used for selecting, used for sending, used for serving, used for showing, used for shredding, used for sliding, used for smoking, used for smoothing, used for starting, used for staying, used for storing, used for telling, used for throwing, used for tightening, used for tilling, used for transporting, used for traveling, used for turning, used for unlocking, used for waking, used for walking, used for washing, used for watching, used for watering, used for wiping, used for writing, used in, used long, used on, used when, used with, uses, walks, worn, worn around, worn as, worn at, worn by, worn for, worn for blocking, worn for covering, worn for exposing, worn for holding, worn for keeping, worn for protecting, worn for riding, worn for supporting, worn for walking, worn in, worn on, worn over, worn through, worn to, worn under, worn with.

A.6.2 List of linking verbs retrieved from our FDs collection

List of verbs linking the source to the target concepts in our norms. Absolute frequencies are reported between brackets.

essere (1048), usare (524), avere (285), trovare (180), fare (109), produrre (51), vivere (40), potere (39), comporre (37), contenere (28), crescere (27), nascere (25), coinvolgere (19), costruire (18), venire (18), mangiare (18), trasportare (11), leggere (8), coltivare (8), associare (7), vedere (7), prendere (7), nutrire (7), suscitare (6), tessere (5), attraversare (5), raccogliere (5), cucinare (5), dividere (5), generare (5), tagliare (5), proteggere (4), sviluppare (4), alimentare (4), afferrare (4), parcheggiare (5), tenere (4), dormire (4), lavorare (4), scaldare (4), coprire (3), profumare (3), correre (3), affettare (3), stringere (3), cavalcare (3), pettinare (3), curare (3), provocare (3), riposare (3), dipingere (3), cuocere (3), sostenere (3), pulire (3), vestire (3), riparare (3), toccare (2), giocare (2), incontrare (2), difendere (2), viaggiare (2), imbottire (2), far (2), servire (2), spostare (2), scrivere (2), cibare (2), volare (2), indicare (2), lavare (2), guidare (2), avvistare (2), muovere (2), battere (2), appoggiare (2), stare (2), bere (2), odorare (2), indossare (2), camminare (2), sorreggere (2), allevare (2), andare (2), protegge (1), portare (1), condire (1), rilassare (1), piegare (1), combattere (1), flettere (1), spazzare (1), abitare (1), sbucciare (1), friggere (1), accendere (1), piangere (1), verniciare (1), cantare (1), scopare (1), piantare (1), arrampicare (1), macchiare (1), suonare (1), affilare (1), impastare (1), gustare (1), soffriggere (1), raccoglie (1), pensare (1), collegare (1), manipolare (1), sapere (1), chiudere (1), puzzare (1), picchiettare (1), tostare (1), premere (1), catturare (1), gareggiare (1), ferire (1), sbottonare (1), permettere (1), truccare (1), accarezzare (1), pizzicare (1), abbellire (1), controllare (1), cescere (1), aprire (1), sentire (1), avvitare (1), appuntire (1), riporre (1), accogliere (1), stirare (1), pescare (1), navigare (1), sorvegliare (1), tendere (1), lanciare (1), accomodare (1), salutare (1), diffondere (1), studiare (1), slacciare (1), emettere (1), togliere (1), aiutare (1), dare (1), posteggiare (1), ornare (1), bruciare (1), uccidere (1), pregare (1), creare (1), rumore (1), appendere (1), sedere (1), bagnare (1), uscire (1), applaudire (1), pittare (1), spremere (1), sistemare (1), stendere (1), riscaldare (1), mettere (1), riprodurre (1), guardare (1), resistere (1).

synset id	synset lemmas	synset id	synset lemmas
a#01328712	{large}	a#01336443	{small, little}
a#01153628	{high}	a#01155404	{low}
a#00761969	{dull}	a#00762511	{sharp}
a#01967903	{square}	a#01966694	{round, circular}
a#00864017	{even}	a#00865276	{uneven}
a#02257422	{sour}	a#02257076	{sweet}
a#02281599	{tasteful}	a#02284574	{tasteless}
a#01830655	{pungent}	a#01831399	{bland}
a#01001840	{fragrant}	a#01002636	{malodorous}
a#02425348	{wet}	a#02429323	{dry}
a#01102688	{hard}	a#01104721	{soft}
a#01199751	{cold}	a#01195771	{hot}
a#00974331	{rigid}	a#00974677	{nonrigid}
a#01315243	{juicy}	a#01315616	{juiceless}
a#02130839	{smooth}	a#02132734	{rough, unsmooth}
a#00211427	{hairy, hirsute}	a#00210154	{hairless}
a#00355823	{chromatic}	a#01738511	{pointed}
a#01968566	{rounded}	a#01464354	{metallic}
a#00364634	{achromatic, colorless}	a#01740196	{pointless, unpointed}
a#01971048	{angular, angulate}	a#01465951	{nonmetallic, nonmetal}
a#00945545	{feathered}	a#00947052	{unfeathered, featherless}

Table A.6.1: List of polar adjectives of all the clusters for which the summed frequency of their member in the STaRS norms is $f \geq 2$.

synset id	synset lemmas
n#06684175	{part, piece}
n#00011937	{artifact}
n#00017954	{group, grouping}
n#06684175	{part, piece}
n#00010572	{substance, matter}
n#00014887	{location}
n#09771631	{linear measure, long measure}
n#00015185	{shape, form}
n#05488770	{sound}
n#00003731	{causal agent, cause, causal agency}
n#00009457	{object, physical object}
n#00017297	{event}
n#10843624	{time period, period, period of time, amount of time}
n#00012865	{psychological feature}

Table A.6.2: Relations Least Common Subsumers extracted from the STaRS norms. Synset lemmas from PWM 1.6.

Aeroplano

Da Wikipedia, l'enciclopedia libera.

L'**aeroplano**, o **aereo** (in forma abbreviata), è un mezzo di trasporto. Esso è più pesante dell'aria, nonostante ciò può volare grazie a principi fisico-meccanici. In base alla classificazione scientifica, gli aeroplani sono dei velivoli, insieme agli idrovolanti e agli anfibi.[1] In quanto tali, sono in grado di volare utilizzando una forza aerodinamica (detta portanza), generata grazie al moto relativo dell'aria lungo una superficie fissa (chiamata ala). Differiscono dagli alianti, in quanto dotati di uno o più motori e per questo motivo rientrano nella più grande categoria delle aerodine a motore, a cui appartengono anche gli elicotteri e altri aeromobili, che però non hanno ali fisse.

Gli aerei sono divisi principalmente in due categorie: militari e civili. Quelli militari a loro volta si dividono in aereo da caccia, bombardieri, aerei da attacco al suolo, aerei da addestramento, aerei da ricognizione, aerei da trasporto. Quelli civili si dividono in aerei di linea, aerei per trasporto merci (detti anche *cargo*), jet executive e aerei da turismo. In generale poi si hanno aerei acrobatici che di solito sono aerei da caccia, da addestramento o da turismo, a volte modificati per adattarli alle particolari sollecitazioni del volo acrobatico.

Altra suddivisione è tra aerei "treassi", dove un qualunque aeroplano in cui il pilota abbia il controllo dell'asse di imbardata, asse di beccheggio e asse di rollio, e aerei "pendolari", dove l'imbardata è assente.

L'Airbus A380 è l'aereo di linea più grande al mondo

Fixed-wing aircraft

From Wikipedia, the free encyclopedia

A **fixed wing aircraft** is an aircraft capable of flight using wings that generate lift due to the vehicle's forward airspeed and the shape of the wings. Fixed-wing aircraft are distinct from rotary-wing aircraft in which the wings form a rotor mounted on a spinning shaft and ornithopters in which the wings flap in similar manner to a bird.

The wings of a fixed-wing aircraft are not necessarily rigid; kites, hang-gliders and aeroplanes using wing-warping or variable geometry are all regarded as fixed-wing aircraft.

A powered fixed-wing aircraft that gains forward thrust from an engine is typically called an **aeroplane**, **airplane**, or simply a **plane**. Aeroplanes include powered paragliders, powered hang gliders and some ground effect vehicles.

Unpowered fixed-wing aircraft, including free-flying gliders of various kinds and tethered kites, can use moving air to gain height.

Most fixed-wing aircraft are flown by a pilot on board the aircraft, but some are designed to be remotely or computer-controlled.

A Boeing 737 airliner - an example of a fixed-wing aircraft

Figure A.6.1: Lead sections of the "*Aeroplano*" article in the Italian Wikipedia (above) and of the "*Airplane*" article in the English Wikipedia (below).
http://it.wikipedia.org/wiki/Aeroplano
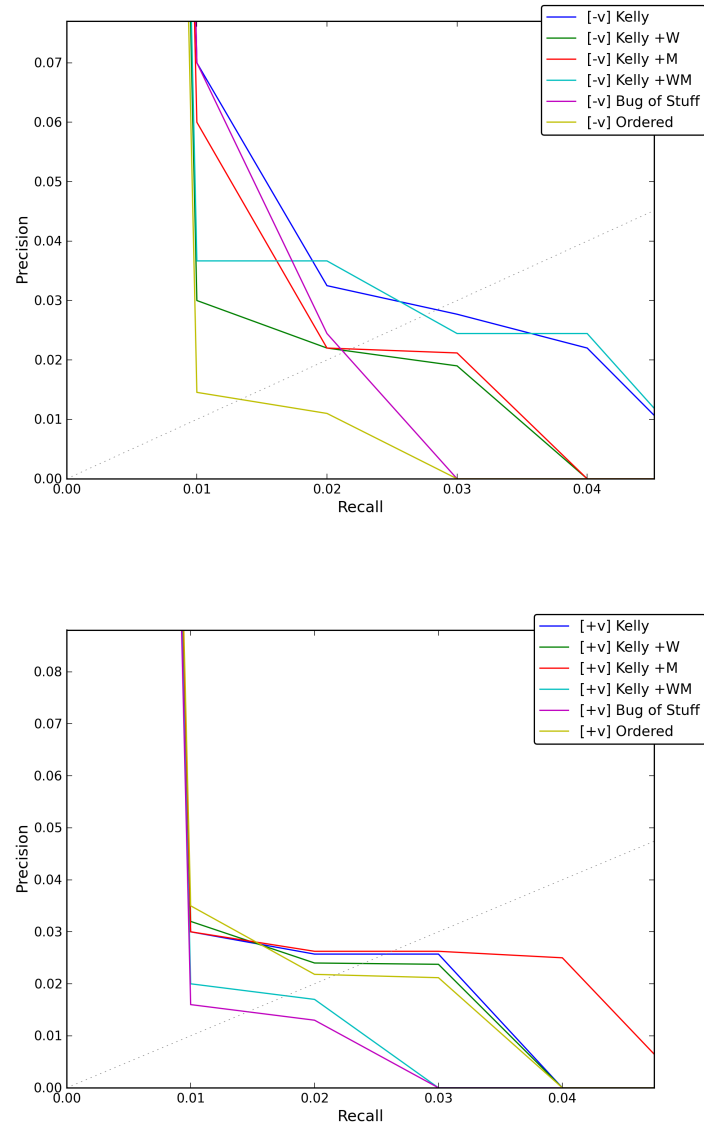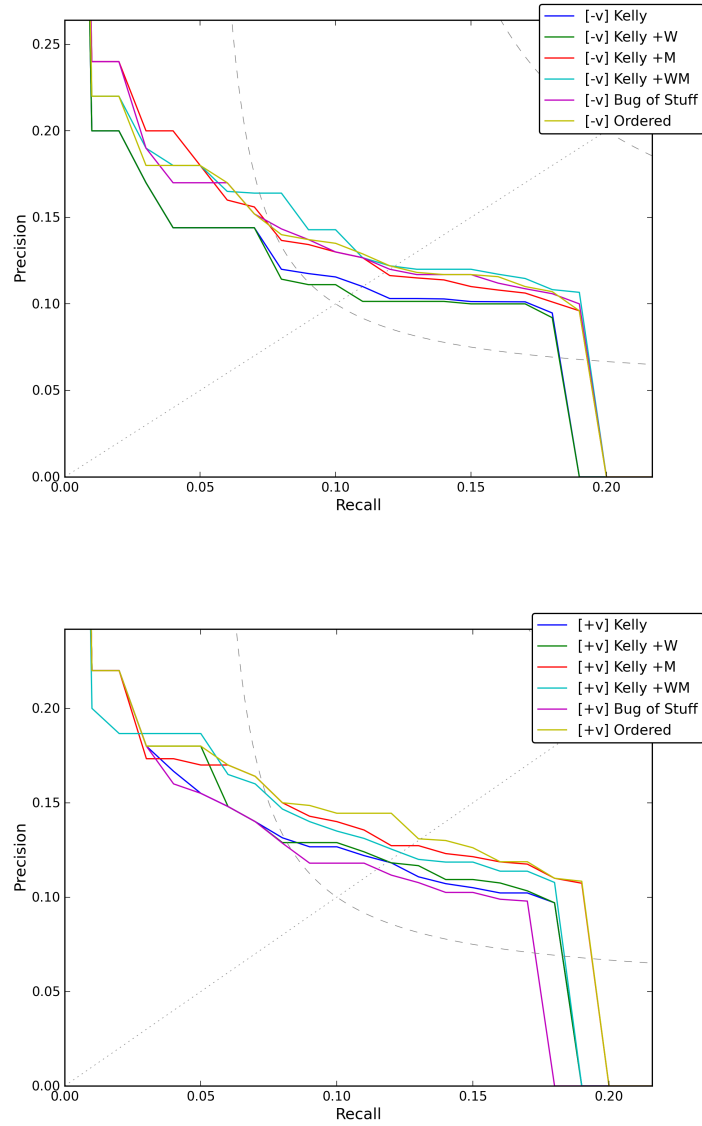http://en.wikipedia.org/wiki/Fixed-wing_aircraft

Figure A.6.2: Best models performance on triples selection excluding verbal target concepts (above) or including them (below). Curves calculated by selecting the top *n* concepts, where n ∈ {5 ≤ n ≤ 100 | n/5 ∈ ℕ}.

Figure A.6.3: Best models performance on pairs selection excluding verbal target concepts (above) or including them (below). Curves calculated by selecting the top *n* concepts, where n ∈ {5 ≤ n ≤ 100 | n/5 ∈ ℕ}.
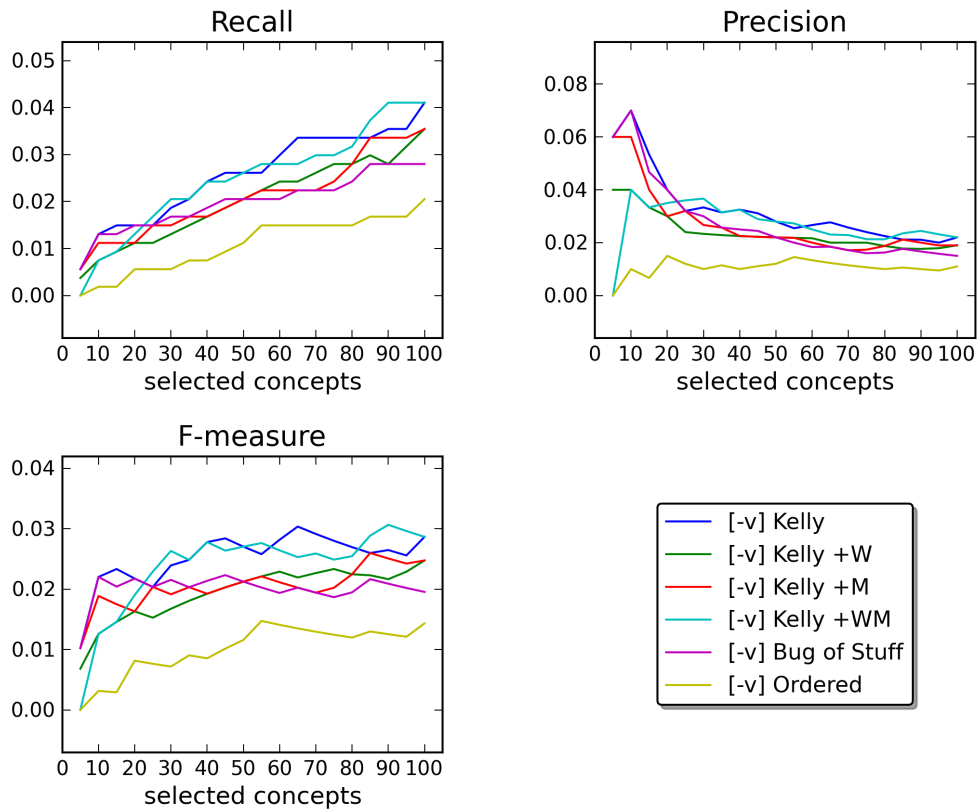
Figure A.6.4: Best models performance on triples selection excluding verbal target concepts: single measures over number of top concepts selected.
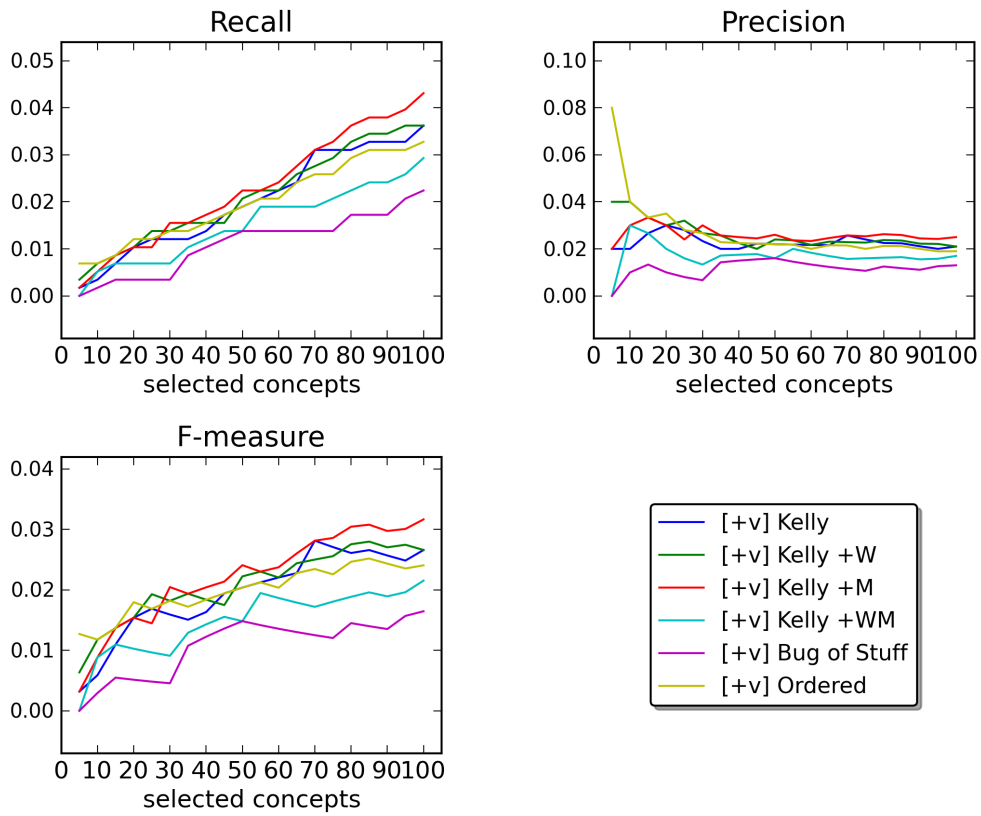
Figure A.6.5: Best models performance on triples selection including verbal target concepts: single measures over number of top concepts selected.
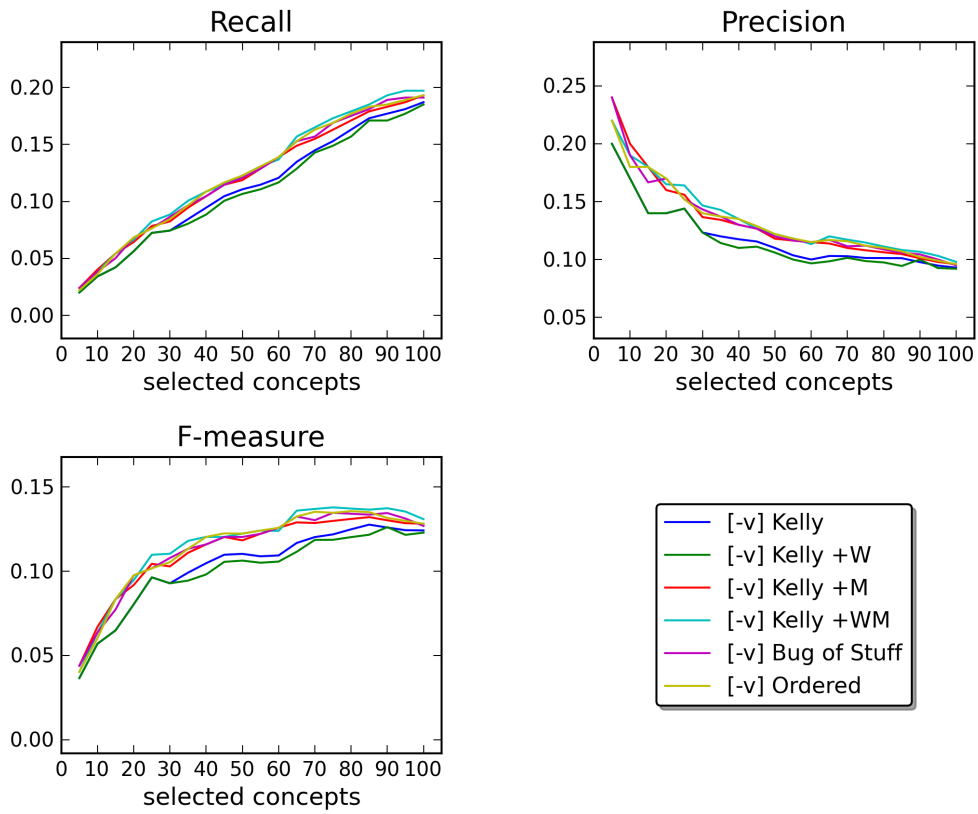
Figure A.6.6: Best models performance on pairs selection excluding verbal target concepts: single measures over number of top concepts selected.
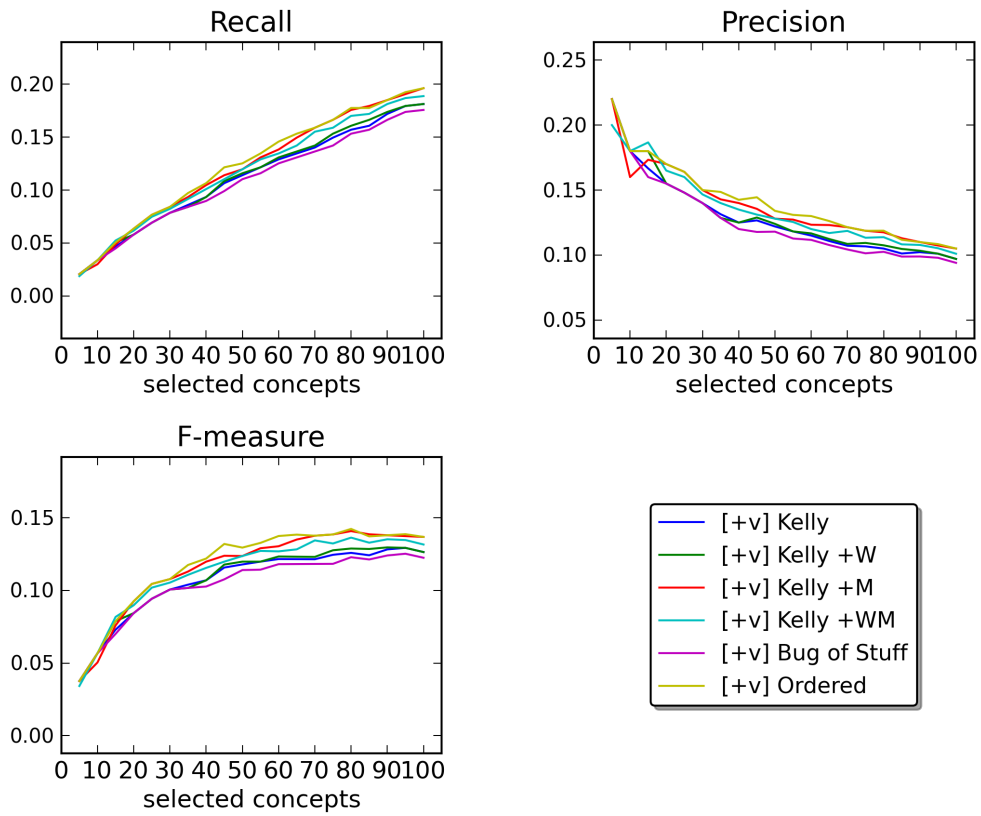
Figure A.6.7: Best models performance on pairs selection including verbal target concepts: single measures over number of top concepts selected.

List of the top 20 *{source_concept, target_concept}* pairs per source concept selected by the "*Kelly*" model trained and tested on nominal and adjectival target concepts. Pairs that are present in the STaRS.sys norms are marked with an asterisk "*".

aeroplano: *{aeroplano, motofalciatrice}, {aeroplano, idrovolante}, {aeroplano, aviatore}, {aeroplano, caffettiera}, {aeroplano, giradischi}, {aeroplano, kalashnikov}, {aeroplano, elica}, {aeroplano, detector}, {aeroplano, fusoliera}, {aeroplano, hangar}, {aeroplano, aviazione}, {aeroplano, piroscafo}, {aeroplano, automobile}, {aeroplano, modellino}, {aeroplano, pilota}, {aeroplano, volo}, {aeroplano, elicottero}, {aeroplano, mitragliatrice}, {aeroplano, paracadute}, {aeroplano, locomotiva}*

calzino: *{calzino, sandalo}, {calzino, calza}*, {calzino, scarpa}*, {calzino, canottiera}, {calzino, camicia}, {calzino, maglietta}, {calzino, ciabatta}, {calzino, maglione}, {calzino, conio}, {calzino, biancheria}, {calzino, calzone}, {calzino, cotone}*, {calzino, polpaccio}, {calzino, asciugamano}, {calzino, cravatta}, {calzino, slip}, {calzino, pigiama}, {calzino, giacca}, {calzino, lana}*, {calzino, paio}*

coniglio: *{coniglio, cacciatora}, {coniglio, tacchino}, {coniglio, colombaccio}, {coniglio, coturnice}, {coniglio, pollo}, {coniglio, ruggito}, {coniglio, lepre}*, {coniglio, pollame}, {coniglio, tortora}, {coniglio, coniglia}, {coniglio, anatra}, {coniglio, coniglietto}, {coniglio, gallina}, {coniglio, volpe}, {coniglio, fagiano}, {coniglio, donnola}, {coniglio, conigliera}, {coniglio, porchetta}, {coniglio, maiale}, {coniglio, volatile}*

garage: *{garage, ripostiglio}, {garage, cantina}*, {garage, lavanderia}, {garage, seminterrato}, {garage, mansarda}, {garage, parcheggio}*, {garage, soffitta}, {garage, autorimessa}, {garage, solarium}, {garage, appartamento}, {garage, terrazzo}, {garage, terrazza}, {garage, sauna}, {garage, villetta}, {garage, scantinato}, {garage, ascensore}, {garage, auto}*, {garage, lavatrice}, {garage, saracinesca}, {garage, miniappartamento}*

patata: *{patata, fecola}*, {patata, carota}*, {patata, cipolla}, {patata, pomodoro}, {patata, fagiolino}, {patata, barbabietola}, {patata, tubero}*, {patata, insalata}, {patata, fagiolo}, {patata, crocchetta}, {patata, bietola}, {patata, melanzana}, {patata, raffaello}, {patata, verdura}*, {patata, rapa}, {patata, riso}, {patata, prezzemolo}, {patata, topinambur}, {patata, cavolfiore}, {patata, carciofo}*

pera: *{pera, mela}*, {pera, senato}, {pera, susina}, {pera, prugna}, {pera, cotogna}, {pera, kiwi}, {pera, albicocca}, {pera, nettarina}, {pera, ciliegia}, {pera, cacio}, {pera, ananas}, {pera, mostarda}, {pera, sidro}, {pera, arancia}, {pera, mandarino}, {pera, agrume}, {pera, frutta}*, {pera, banana}*, {pera, fragola}, {pera, zabaione}*

picchio: *{picchio, picco}, {picchio, marte}, {picchio, ghiandaia}, {picchio, astore}, {picchio, gheppio}, {picchio, torcicollo}, {picchio, nibbio}, {picchio, allocco}, {picchio, pettirosso}, {picchio, rapace}, {picchio, cuculo}, {picchio, rondone}, {picchio, gufo}, {picchio, muratore}, {picchio, fringuello}, {picchio, civetta}, {picchio, usignolo}, {picchio, avifauna}, {picchio, scoiattolo}, {picchio, uccello}**

scopa: *{scopa, manico}*, {scopa, spazzolone}*, {scopa, strofinaccio}, {scopa, befana}, {scopa, ramazza}, {scopa, briscola}, {scopa, sgabuzzino}, {scopa, spazzolino}, {scopa, strega}*, {scopa, spazzino}*, {scopa, spazzola}*, {scopa, stanzino}, {scopa, paletta}*, {scopa, lavastoviglie}, {scopa, secchio}, {scopa, pomo}, {scopa, aspirapolvere}*, {scopa, dentifricio}, {scopa, lavandino}, {scopa, piumino}*

sedia: *{sedia, rotella}, {sedia, schienale}*, {sedia, tavolo}*, {sedia, impagliatore}, {sedia, tavolino}, {sedia, sgabello}*, {sedia, ombrellone}, {sedia, spalliera}, {sedia, armadio}, {sedia, bracciolo}*, {sedia, scrittoio}, {sedia, comodino}, {sedia, divano}, {sedia, ruota}, {sedia, sofà}, {sedia, scrivania}*, {sedia, poltroncina}, {sedia, efflorescenza}, {sedia, panca}, {sedia, lavabo}*

testa: *{testa, essere}, {testa, giramento}, {testa, colpo}, {testa, cross}, {testa, corner}, {testa, traversa}, {testa, spacco}, {testa, collo}, {testa, acetabolo}, {testa, classifica}, {testa, male}, {testa, piede}, {testa, scappellotto}, {testa, palla}, {testa, mano}, {testa, emicrania}, {testa, zucchetto}, {testa, lisca}, {testa, cranio}, {testa, coda}*

List of the top 20 *{source_concept, relation, target_concept}* triples per source concept selected by the "*Kelly*" model trained and tested on nominal and adjectival target concepts. Triples that are present in the STaRS.sys norms are marked with an asterisk "*".

aeroplano: *{aeroplano, has Component, motofalciatrice}, {aeroplano, has Component, idrovolante}, {aeroplano, is Associated with, idrovolante}, {aeroplano, Made of, idrovolante}, {aeroplano, Coordination, idrovolante}, {aeroplano, has Component, aviatore}, {aeroplano, Made of, aviatore}, {aeroplano, has Component, caffettiera}, {aeroplano, has Component, giradischi}, {aeroplano, has Component, kalashnikov}, {aeroplano, has Component, elica}, {aeroplano, Made of, elica}, {aeroplano, has Component, detector}, {aeroplano, has Component, fusoliera}, {aeroplano, Made of, fusoliera}, {aeroplano, has Component, hangar}, {aeroplano, has Size, hangar}, {aeroplano, Made of, hangar}, {aeroplano, has Component, aviazione}, {aeroplano, Made of, aviazione}*

calzino: *{calzino, has Component, sandalo}, {calzino, Coordination, sandalo}, {calzino, is Associated with, sandalo}, {calzino, is Used with, sandalo}, {calzino, has Component, calza}, {calzino, Coordination, calza}*, {calzino, is Associated with, calza}, {calzino, has Shape, calza}, {calzino, has Size, calza}, {calzino, Made of, calza}, {calzino, is Used by, calza}, {calzino, isA, calza}*, {calzino, has Component, scarpa}, {calzino, Coordination, scarpa}, {calzino, is Associated with, scarpa}, {calzino, Made of, scarpa}, {calzino, has Size, scarpa}, {calzino, isA, scarpa}, {calzino, is Used by, scarpa}, {calzino, has Component, canottiera}*

coniglio: *{coniglio, has Component, cacciatora}, {coniglio, has Size, cacciatora}, {coniglio, has Colour, cacciatora}, {coniglio, is Used with, cacciatora}, {coniglio, has Component, tacchino}, {coniglio, Coordination, tacchino}, {coniglio, has Size, tacchino}, {coniglio, Made of, tacchino}, {coniglio, has Colour, tacchino}, {coniglio, is Used by, tacchino}, {coniglio, has Component, colombaccio}, {coniglio, Coordination, colombaccio}, {coniglio, has Size, colombaccio}, {coniglio, is Associated with, colombaccio}, {coniglio, is the Category of, colombaccio}, {coniglio, has Component, coturnice}, {coniglio, Made of, coturnice}, {coniglio, is Used with, coturnice}, {coniglio, is Associated with, coturnice}, {coniglio, Coordination, coturnice}*

garage: *{garage, has Component, ripostiglio}, {garage, has Size, ripostiglio}, {garage, Coordination, ripostiglio}, {garage, Made of, ripostiglio}, {garage, has Colour, ripostiglio}, {garage, is Used with, ripostiglio}, {garage, is Associated with, ripostiglio}, {garage, is Used by, ripostiglio}, {garage, has Component, cantina}, {garage, Coordination, cantina}*, {garage, has Size, cantina}, {garage, is Used with, cantina}, {garage, Made of, cantina}, {garage, has Colour, cantina}, {garage, is Associated with, cantina}, {garage, isA, cantina}, {garage, is Used by, cantina}, {garage, has Component, lavanderia}, {garage, Coordination, lavanderia}, {garage, Made of, lavanderia}*

patata: *{patata, has Component, fecola}, {patata, Made of, fecola}, {patata, has Size, fecola}, {patata, is Used with, fecola}, {patata, has Component, carota}, {patata, Coordination, carota}*, {patata, Made of, carota}, {patata, is Associated with, carota}, {patata,*

is Used with, carota}, {patata, has Size, carota}, {patata, has Colour, carota}, {patata, is Used by, carota}, {patata, has Component, cipolla}, {patata, Coordination, cipolla}, {patata, has Size, cipolla}, {patata, Made of, cipolla}, {patata, is Associated with, cipolla}, {patata, has Colour, cipolla}, {patata, is Used with, cipolla}, {patata, isA, cipolla}

pera: *{pera, has Component, mela}, {pera, Coordination, mela}*, {pera, Made of, mela}, {pera, has Colour, mela}, {pera, has Size, mela}, {pera, is Associated with, mela}, {pera, is Used with, mela}, {pera, is Used by, mela}, {pera, isA, mela}, {pera, has Shape, mela}, {pera, has Component, senato}, {pera, has Component, susina}, {pera, Coordination, susina}, {pera, is Associated with, susina}, {pera, Made of, susina}, {pera, has Component, prugna}, {pera, Coordination, prugna}, {pera, has Size, prugna}, {pera, Made of, prugna}, {pera, is Used by, prugna}*

picchio: *{picchio, has Component, picco}, {picchio, is Used by, picco}, {picchio, Coordination, picco}, {picchio, has Component, marte}, {picchio, has Size, marte}, {picchio, has Component, ghiandaia}, {picchio, has Component, astore}, {picchio, has Component, gheppio}, {picchio, has Component, torcicollo}, {picchio, is Used with, torcicollo}, {picchio, Coordination, torcicollo}, {picchio, Made of, torcicollo}, {picchio, has Component, nibbio}, {picchio, Coordination, nibbio}, {picchio, has Component, allocco}, {picchio, has Size, allocco}, {picchio, has Component, pettirosso}, {picchio, has Size, pettirosso}, {picchio, Made of, pettirosso}, {picchio, has Component, rapace}*

scopa: *{scopa, has Component, manico}*, {scopa, Made of, manico}, {scopa, has Size, manico}, {scopa, has Component, spazzolone}, {scopa, Coordination, spazzolone}*, {scopa, Made of, spazzolone}, {scopa, is Associated with, spazzolone}, {scopa, has Component, strofinaccio}, {scopa, Made of, strofinaccio}, {scopa, Coordination, strofinaccio}, {scopa, is Associated with, strofinaccio}, {scopa, has Component, befana}, {scopa, has Size, befana}, {scopa, Made of, befana}, {scopa, is Associated with, befana}, {scopa, Coordination, befana}, {scopa, is Used with, befana}, {scopa, has Component, ramazza}, {scopa, Coordination, ramazza}, {scopa, Made of, ramazza}*

sedia: *{sedia, has Component, rotella}, {sedia, Made of, rotella}, {sedia, has Colour, rotella}, {sedia, has Size, rotella}, {sedia, is Used by, rotella}, {sedia, is Associated with, rotella}, {sedia, has Component, schienale}*, {sedia, Made of, schienale}, {sedia, has Size, schienale}, {sedia, has Colour, schienale}, {sedia, is Associated with, schienale}, {sedia, is Used with, schienale}, {sedia, is Used by, schienale}, {sedia, has Component, tavolo}, {sedia, has Component, impagliatore}, {sedia, Made of, impagliatore}, {sedia, has Component, tavolino}, {sedia, Coordination, tavolino}, {sedia, Coordination, tavolo}, {sedia, is Associated with, tavolino}*

testa: *{testa, has Size, essere}, {testa, has Component, essere}, {testa, has Component, giramento}, {testa, Made of, giramento}, {testa, has Size, giramento}, {testa, isA, giramento}, {testa, is Associated with, giramento}, {testa, is Used by, giramento}, {testa, Coordination, giramento}, {testa, has Component, colpo}, {testa, has Component, cross}, {testa, has Size, cross}, {testa, Made of, cross}, {testa, has Colour, cross}, {testa, Coordination, cross}, {testa, is Associated with, cross}, {testa, is Used with, cross}, {testa, isA, cross}, {testa, has Component, corner}, {testa, has Component, traversa}*

List of the top 20 *{source_concept, target_concept}* pairs per source concept selected by the "*Kelly +WM*" model trained and tested on nominal and adjectival target concepts. Pairs that are present in the STaRS.sys norms are marked with an asterisk"*".

aeroplano: *{aeroplano, motofalciatrice}, {aeroplano, idrovolante}, {aeroplano, aviatore}, {aeroplano, caffettiera}, {aeroplano, giradischi}, {aeroplano, kalashnikov}, {aeroplano, pilota}, {aeroplano, elica}, {aeroplano, detector}, {aeroplano, volo}, {aeroplano, fusoliera}, {aeroplano, hangar}, {aeroplano, automobile}, {aeroplano, aviazione}, {aeroplano, piroscafo}, {aeroplano, modellino}, {aeroplano, elicottero}, {aeroplano, mitragliatrice}, {aeroplano, paracadute}, {aeroplano, registratore}*

calzino: *{calzino, sandalo}, {calzino, calza}*, {calzino, scarpa}*, {calzino, canottiera}, {calzino, camicia}, {calzino, maglietta}, {calzino, ciabatta}, {calzino, maglione}, {calzino, conio}, {calzino, biancheria}, {calzino, cotone}*, {calzino, calzone}, {calzino, polpaccio}, {calzino, asciugamano}, {calzino, cravatta}, {calzino, slip}, {calzino, pigiama}, {calzino, lana}*, {calzino, paio}, {calzino, giacca}*

coniglio: *{coniglio, cacciatora}, {coniglio, pollo}, {coniglio, tacchino}, {coniglio, lepre}*, {coniglio, ruggito}, {coniglio, colombaccio}, {coniglio, coturnice}, {coniglio, pollame}, {coniglio, volpe}, {coniglio, tortora}, {coniglio, anatra}, {coniglio, gallina}, {coniglio, coniglia}, {coniglio, coniglietto}, {coniglio, fagiano}, {coniglio, (volatile)}, {coniglio, maiale}, {coniglio, donnola}, {coniglio, selvaggina}, {coniglio, conigliera}*

garage: *{garage, appartamento}, {garage, cantina}*, {garage, camera}, {garage, parcheggio}*, {garage, ripostiglio}, {garage, cucina}, {garage, soggiorno}, {garage, auto}*, {garage, lavanderia}, {garage, mansarda}, {garage, seminterrato}, {garage, costruzione}*, {garage, soffitta}, {garage, piano}, {garage, villetta}, {garage, sauna}, {garage, autorimessa}, {garage, giardino}, {garage, solarium}, {garage, terrazza}*

patata: *{patata, riso}, {patata, insalata}, {patata, pomodoro}, {patata, cipolla}, {patata, carota}*, {patata, fecola}*, {patata, g}, {patata, tubero}*, {patata, fagiolino}, {patata, olio}, {patata, barbabietola}, {patata, fagiolo}, {patata, minestra}, {patata, verdura}*, {patata, cucchiaio}, {patata, oliva}, {patata, buccia}*, {patata, prezzemolo}, {patata, melanzana}, {patata, crocchetta}*

pera: *{pera, presidente}, {pera, senato}, {pera, università }, {pera, mela}*, {pera, riso}, {pera, susina}, {pera, prugna}, {pera, kiwi}, {pera, cotogna}, {pera, albicocca}, {pera, nettarina}, {pera, frutta}*, {pera, insalata}, {pera, ciliegia}, {pera, cacio}, {pera, arancia}, {pera, ananas}, {pera, sidro}, {pera, mostarda}, {pera, mandarino}*

picchio: *{picchio, picco}, {picchio, ghiandaia}, {picchio, marte}, {picchio, astore}, {picchio, gheppio}, {picchio, torcicollo}, {picchio, nibbio}, {picchio, allocco}, {picchio, pettirosso}, {picchio, rapace}, {picchio, cuculo}, {picchio, rondone}, {picchio, gufo}, {picchio, muratore}, {picchio, fringuello}, {picchio, civetta}, {picchio, usignolo}, {picchio, avifauna}, {picchio, scoiattolo}, {picchio, uccello}**

scopa: *{scopa, manico}*, {scopa, spazzolone}*, {scopa, strofinaccio}, {scopa, befana}, {scopa, ramazza}, {scopa, briscola}, {scopa, strega}*, {scopa, sgabuzzino}, {scopa, spazzolino}, {scopa, spazzino}*, {scopa, spazzola}*, {scopa, paletta}*, {scopa, stanzino}, {scopa, lavastoviglie}, {scopa, secchio}, {scopa, pomo}, {scopa, aspirapolvere}*, {scopa, dentifricio}, {scopa, lavandino}, {scopa, piumino}*

sedia: *{sedia, rotella}, {sedia, tavolo}*, {sedia, schienale}*, {sedia, tavolino}, {sedia, impagliatore}, {sedia, sgabello}*, {sedia, armadio}, {sedia, sedere}, {sedia, ruota}, {sedia, ombrellone}, {sedia, spalliera}, {sedia, letto}, {sedia, bracciolo}*, {sedia, scrittoio}, {sedia, divano}, {sedia, comodino}, {sedia, arredamento}*, {sedia, scrivania}*, {sedia, banco}, {sedia, sofà }*

testa: *{testa, essere}, {testa, colon}, {testa, mano}, {testa, colpo}, {testa, cosa}, {testa, uomo}*, {testa, volta}, {testa, occhio}*, {testa, palla}, {testa, regina}, {testa, piede}, {testa, corpo}*, {testa, parte}, {testa, grande}, {testa, cross}, {testa, classifica}, {testa, collo}, {testa, area}, {testa, spalla}, {testa, giorno}*

List of the top 20 *{source_concept, relation, target_concept}* triples per source concept selected by the "*Kelly +WM*" model trained and tested on nominal and adjectival target concepts. Triples that are present in the STaRS.sys norms are marked with an asterisk "*".

aeroplano: *{aeroplano, is the Category of, motofalciatrice}, {aeroplano, Space Located, motofalciatrice}, {aeroplano, is the Category of, idrovolante}, {aeroplano, is Used by, aviatore}, {aeroplano, is the Category of, aviatore}, {aeroplano, is the Category of, caffettiera}, {aeroplano, is the Category of, giradischi}, {aeroplano, is the Category of, kalashnikov}, {aeroplano, is Used by, pilota}, {aeroplano, is the Category of, elica}, {aeroplano, Space Located, elica}, {aeroplano, is the Category of, detector}, {aeroplano, Space Located, detector}, {aeroplano, is the Category of, volo}, {aeroplano, Space Located, fusoliera}, {aeroplano, is the Category of, fusoliera}, {aeroplano, Space Located, hangar}, {aeroplano, has Colour, hangar}, {aeroplano, is the Category of, hangar}, {aeroplano, is the Category of, automobile}*

calzino: *{calzino, is the Category of, sandalo}, {calzino, Space Located, sandalo}, {calzino, Coordination, sandalo}, {calzino, is Used with, sandalo}, {calzino, has Colour, sandalo}, {calzino, is Associated with, sandalo}, {calzino, is the Category of, calza}, {calzino, is the Category of, scarpa}, {calzino, is Used with, calza}, {calzino, is the Origin of, calza}, {calzino, Space Located, calza}, {calzino, Space Located, scarpa}, {calzino, Coordination, scarpa}, {calzino, is Used with, scarpa}*, {calzino, is the Category of, canottiera}, {calzino, Space Located, canottiera}, {calzino, is Associated with, canottiera}, {calzino, is the Category of, camicia}, {calzino, has Colour, camicia}, {calzino, Coordination, camicia}*

coniglio: *{coniglio, is the Category of, cacciatora}, {coniglio, Space Located, cacciatora}, {coniglio, has Size, cacciatora}, {coniglio, is Used with, cacciatora}, {coniglio, is Used by, pollo}, {coniglio, is the Category of, tacchino}, {coniglio, is the Category of, pollo}, {coniglio, Coordination, tacchino}, {coniglio, is the Category of, lepre}, {coniglio, is the Category of, ruggito}, {coniglio, has Colour, tacchino}, {coniglio, is Used with, tacchino}, {coniglio, has Size, tacchino}, {coniglio, is Associated with, tacchino}, {coniglio, is Used by, tacchino}, {coniglio, is the Category of, colombaccio}, {coniglio, is the Category of, coturnice}, {coniglio, is Used with, colombaccio}, {coniglio, Coordination, colombaccio}, {coniglio, has Colour, colombaccio}*

garage: *{garage, is the Category of, appartamento}, {garage, is the Category of, cantina}, {garage, Space Located, camera}, {garage, is the Category of, parcheggio}, {garage, is the Category of, ripostiglio}, {garage, Coordination, ripostiglio}, {garage, Space Located, ripostiglio}, {garage, has Size, ripostiglio}, {garage, is the Category of, cucina}, {garage, Coordination, cantina}*, {garage, is the Category of, soggiorno}, {garage, is the Category of, auto}, {garage, is the Category of, lavanderia}, {garage, is the Category of, mansarda}, {garage, is the Category of, seminterrato}, {garage, Space Located, cantina}, {garage, has Size, cantina}, {garage, is Associated with, cantina}, {garage, has Colour, cantina}, {garage, Coordination, lavanderia}*

155

patata: *{patata, is the Category of, riso}, {patata, is the Category of, insalata}, {patata, has Colour, riso}, {patata, is the Category of, pomodoro}, {patata, is the Category of, cipolla}, {patata, is the Category of, carota}, {patata, is Used with, carota}, {patata, is Used with, fecola}, {patata, is the Category of, fecola}, {patata, has Colour, fecola}, {patata, Space Located, fecola}, {patata, is the Category of, g}, {patata, is Used with, cipolla}, {patata, has Colour, carota}, {patata, Space Located, carota}, {patata, is Used by, carota}, {patata, Coordination, carota}*, {patata, has Size, carota}, {patata, is Used by, insalata}, {patata, has Colour, cipolla}*

pera: *{pera, is Used by, presidente}, {pera, is the Category of, senato}, {pera, is the Category of, universitÃ }, {pera, is the Category of, mela}, {pera, is Used with, mela}, {pera, has Colour, mela}, {pera, is Used by, senato}, {pera, Coordination, mela}*, {pera, Space Located, mela}, {pera, is Used by, mela}, {pera, is the Category of, riso}, {pera, is Used with, susina}, {pera, is the Category of, susina}, {pera, Coordination, susina}, {pera, is the Category of, prugna}, {pera, is Used with, prugna}, {pera, has Colour, prugna}, {pera, is Used with, kiwi}, {pera, is Used by, prugna}, {pera, is the Category of, kiwi}*

picchio: *{picchio, is the Category of, picco}, {picchio, is the Origin of, picco}, {picchio, Space Located, picco}, {picchio, is Used with, picco}, {picchio, Coordination, picco}, {picchio, is the Category of, ghiandaia}, {picchio, is the Category of, marte}, {picchio, is Used by, marte}, {picchio, is Used by, ghiandaia}, {picchio, is the Category of, astore}, {picchio, is the Category of, gheppio}, {picchio, is the Category of, torcicollo}, {picchio, Coordination, torcicollo}, {picchio, is Used with, torcicollo}, {picchio, is the Origin of, torcicollo}, {picchio, is Associated with, torcicollo}, {picchio, is the Category of, nibbio}, {picchio, Coordination, nibbio}, {picchio, is the Category of, allocco}, {picchio, is Used by, allocco}*

scopa: *{scopa, is the Category of, manico}, {scopa, Space Located, manico}, {scopa, has Colour, manico}, {scopa, has Size, manico}, {scopa, is the Category of, spazzolone}, {scopa, is Used with, spazzolone}, {scopa, Coordination, spazzolone}, {scopa, is the Category of, strofinaccio}, {scopa, is Associated with, strofinaccio}, {scopa, Coordination, strofinaccio}, {scopa, has Size, strofinaccio}, {scopa, is the Category of, befana}, {scopa, has Colour, befana}, {scopa, is Associated with, befana}, {scopa, is the Category of, ramazza}, {scopa, Space Located, ramazza}, {scopa, is the Category of, briscola}, {scopa, is Used with, briscola}, {scopa, has Size, briscola}, {scopa, has Colour, briscola}*

sedia: *{sedia, is the Category of, rotella}, {sedia, Space Located, rotella}, {sedia, has Colour, rotella}, {sedia, is the Category of, tavolo}, {sedia, Space Located, tavolo}, {sedia, Space Located, schienale}, {sedia, is the Category of, schienale}, {sedia, is the Category of, tavolino}, {sedia, Space Located, tavolino}, {sedia, has Colour, schienale}, {sedia, has Size, schienale}, {sedia, is Used with, schienale}, {sedia, is Used by, impagliatore}, {sedia, is the Category of, impagliatore}, {sedia, is the Category of, sgabello}, {sedia, is Used with, tavolo}*, {sedia, Coordination, tavolino}, {sedia, is Used with, tavolino}, {sedia, is Associated with, tavolino}, {sedia, is the Category of, armadio}*

testa: *{testa, is the Category of, essere}, {testa, has Colour, essere}, {testa, is the Cate-*

gory of, colon}, {testa, has Colour, colon}, {testa, is the Category of, mano}, {testa, is the Category of, colpo}, {testa, is the Category of, cosa}, {testa, is Used by, uomo}*, {testa, is the Category of, volta}, {testa, is the Category of, occhio}, {testa, is the Category of, palla}, {testa, is the Category of, regina}, {testa, is the Category of, piede}, {testa, is the Category of, corpo}, {testa, is the Category of, parte}, {testa, has Size, grande}, {testa, is the Category of, cross}, {testa, is Used by, classifica}, {testa, is the Category of, collo}, {testa, is the Category of, area}

References

Aftonomos, L., Appelbaum, J., and Steele, R. (1999). Improving outcomes for persons with aphasia in advanced community-based treatment programs. *Stroke*, 30:1370–1379.

Almuhareb, A. and Poesio, M. (2004). Attribute-based and value-based clustering: An evaluation. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pages 158–165.

Almuhareb, A. and Poesio, M. (2005). Concept learning and categorization from the web. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, pages 103–108.

Alonge, A., Calzolari, N., Vossen, P., Bloksma, L., Castellon, I., Marti, M., and Peters, W. (1998). The linguistic design of the eurowordnet database. *Computers and the Humanities, Special Issue on EuroWordNet*, 32:91–115.

Amaro, R., Chaves, R., Marrafa, P., and Mendes, S. (2006). Enriching wordnets with new relations and with event and argument structures. In *Proceeding of the 7th International Conference on Intelligent Text Processing and Computational Linguistics*, pages 28–40.

Andrews, M., Vigliocco, G., and Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological review*, 116:463–498.

Arcuri, L. and Girotto, V. (1985). Giudizi di tipicità per esemplari di 6 categorie naturali. uno studio evolutivo. Report del Dipartimento di Psicologia dello SViluppo e della Socializzazione.

Artstein, R. and Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34:555–596.

Ashcraft, M. (1978). Property norms for typical and atypical items from 17 categories: A description and discussion. *Memory & Cognition*, 6:227–232.

Attardi, G. (2006). Experiments with a multilanguage non-projective dependency parser. In *Proceedings of the Tenth Conference on Computational Natural Language Learning*, pages 166–170.

Baker, C. F., Fillmore, C. J., and Lowe, J. B. (1998). The berkeley framenet project. In *Proceedings of the 17th International Conference on Computational Linguistics*, pages 86–90.

Barbu, E. (2008). Combining methods to learn feature-norm-like concept descriptions. In *Proceedings of the ESSLLI Workshop on Distributional Lexical Semantics*, pages 9–16.

Barbu, E. and Poesio, M. (2008). A comparison of wordnet and feature norms. In *Proceedings of the fourth Global Wordnet Conference*.

Baroni, M., Bernardini, S., Ferraresi, A., and Zanchetta, E. (2009). The wacky wide web: a collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation*, 43:209–226.

Baroni, M., Evert, S., and Lenci, A. (2008). *Bridging the gap between semantic theory and computational simulations: proceedings of the ESSLLI 2008. Workshop on distributional semantics.*

Baroni, M., Murphy, B., Barbu, E., and Poesio, M. (2010). Strudel: a corpus-based semantic model based on properties and types. *Cognitive Science*, 34:222–254.

Barsalou, L. (2008). Grounded cognition. *Annual Review of Psychology*, 59:617–645.

Bentivogli, L., Bocco, A., Pianta, E., et al. (2004a). Archiwordnet: integrating wordnet with domain-specific knowledge. In *Proceedings of the 2nd International Conference of the Global WordNet Association*, pages 39–46.

Bentivogli, L., Forner, P., Magnini, B., and Pianta, E. (2004b). Revising the wordnet domains hierarchy: semantics, coverage and balancing. In *Proceedings of the Workshop on Multilingual Linguistic Ressources*, pages 101–108.

Bentivogli, L. and Pianta, E. (2003). Beyond lexical units: Enriching wordnets with phrasets. In *Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics*, pages 67–70.

Bentivogli, L. and Pianta, E. (2004). Extending wordnet with syntagmatic information. In *Proceedings of the 2nd International WordNet Conference*, pages 47–53.

Bertinetto, P., Burani, C., Laudanna, A., Marconi, L., Ratti, D., Rolando, C., and Thorton, A. (2005). Corpus e lessico di frequenza dell'italiano scritto (CoLFIS).

Bird, S. and Loper, E. (2006). Nltk: The natural language toolkit. In *Proceedings of the COLING/ACL on Interactive presentation sessions*, pages 69–72.

Blundo, C., Ricci, M., and Miller, L. (2006). Category-specific knowledge deficit for animals in a patient with herpes simplex encephalitis. *Cognitive neuropsychology*, 23:1248–1268.

Bond, F. and Paik, K. (2012). A survey of wordnets and their licenses. In *Proceedings of the 6th Global WordNet Conference*.

Borgo, F. and Shallice, T. (2001). When living things and other "sensory quality" categories behave in the same fashion: a novel category specificity effect. *Neurocase*, 7:201–220.

Bosco, C. and Mazzei, A. (2011). The evalita 2011 parsing task: the dependency track. *Working Notes of EVALITA 2011*.

Bouma, G. (2009). Normalized (pointwise) mutual information in collocation extraction. *Proceedings of the Biennial GSCL Conference 2009*, pages 31–40.

Boyd-Graber, J., Fellbaum, C., Osherson, D., and Schapire, R. (2006). Adding dense, weighted connections to wordnet. In *Proceedings of the Third International WordNet Conference*, pages 29–36.

Boyle, M. and Coelho, C. (1995). Application of semantic feature analysis as a treatment for aphasic dysnomia. *American Journal of Speech-Language Pathology*, 4:94–98.

Brainerd, C., Yang, Y., Reyna, V., Howe, M., and Mills, B. (2008). Semantic processing in "associative" false memory. *Psychonomic bulletin & review*, 15:1035–1053.

Brust, J. C., Shafer, S. Q., Richter, R. W., and Bruun, B. (1976). Aphasia in acute stroke. *Stroke*, 7:167–174.

Budanitsky, A. and Hirst, G. (2001). Semantic distance in wordnet: An experimental, application-oriented evaluation of five measures. In *Proceedings of the Workshop on WordNet and Other Lexical Resources*, pages 29–34.

Budanitsky, A. and Hirst, G. (2006). Evaluating wordnet-based measures of lexical semantic relatedness. *Computational Linguistics*, 32:13–47.

Bunescu, R. and Mooney, R. (2005). Subsequence kernels for relation extraction. In *Proceedings of the 19th Conference on Neural Information Processing Systems*.

Buxbaum, L., Veramontil, T., and Schwartz, M. (2000). Function and manipulation tool knowledge in apraxia: knowing "what for" but not "how". *Neurocase*, 6:83–97.

Capitani, E., Laiacona, M., Mahon, B., and Caramazza, A. (2003). What are the facts of semantic category-specific deficits? a critical review of the clinical evidence. *Cogn Neuropsychol*, 20:213–261.

Caramazza, A. and Hillis, A. E. (1990). Where do semantic errors come from? *Cortex*, 26:95–122.

Caramazza, A. and Hillis, A. E. (1991). Lexical organization of nouns and verbs in the brain. *Nature*, 349:788–790.

Caramazza, A., Hillis, A. E., Rapp, B. C., and Romani, C. (1990). The multiple semantics hypothesis: multiple confusions? *Cognitive Neuropsychology*, 7:161–189.

Caramazza, A., Papagno, C., and Ruml, W. (2000). The selective impairment of phonological processing in speech production. *Brain Lang*, 75:428–450.

Caramazza, A. and Shelton, J. (1998). Domain-specific knowledge systems in the brain: The animate-inanimate distinction. *Journal of cognitive neuroscience*, 10:1–34.

Cassidy, P. (2000). An investigation of the semantic relations in the roget⊠s thesaurus: Preliminary results. In *CICLing-2000: Conference on Intelligent Text Processing and Computational Linguistics, February*, volume 13, pages 181–204.

Church, K. and Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational linguistics*, 16:22–29.

Clark, S. and Curran, J. (2007). Wide-coverage efficient statistical parsing with ccg and log-linear models. *Computational Linguistics*, 33:493–552.

Colby, K., Christinaz, D., Parkison, R., Graham, S., and Karpf, C. (1981). A word-finding computer program with a dynamic lexical-semantic memory for patients with anomia using an intelligent speech prosthesis. *Brain and Language*, 14:272–281.

Collins, A. and Loftus, E. (1975). A spreading-activation theory of semantic processing. *Psychological review*, 82:407–428.

Collins, A. and Quillian, M. (1969). Retrieval time from semantic memory. *Journal of verbal learning and verbal behavior*, 8:240–247.

Conger, A. (1980). Integration and generalization of kappas for multiple raters. *Psychological Bulletin*, 88:322.

Connolly, A., Gleitman, L., and Thompson-Schill, S. (2007). Effect of congenital blindness on the semantic representation of some everyday concepts. *Proceedings of the National Academy of Sciences*, 104:8241–8246.

Cree, G. and McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of *chipmunk, cherry, chisel, cheese,* and *cello* (and many other such concrete nouns). *Journal of Experimental Psychology: General*, 132:163–201.

Cresswell, M. and Hughes, G. (2012). *A New Introduction to Modal Logic.* Routledge.

Crosson, B., Moberg, P., Boone, J., Gonzalez Rothi, L., and Raymer, A. (1997). Category-specific naming deficit for medical terms after dominant thalamic/capsular hemorrhage. *Brain and Language*, 60:407–442.

Cruse, D. (1986). *Lexical semantics.* Cambridge University Press.

Cuetos, F., Aguado, G., and Caramazza, A. (2000). Dissociation of semantic and phonological errors in naming. *Brain Lang*, 75:451–460.

Davis, L. and Stanton, S. (2005). Semantic feature analysis as a functional therapy tool. *Contemporary Issues in Communication Science and Disorders*, 32:85–92.

De Deyne, S., Verheyen, S., Ameel, E., Vanpaemel, W., Dry, M., Voorspoels, W., and Storms, G. (2008). Exemplar by feature applicability matrices and other dutch normative data for semantic concepts. *Behavior Research Methods*, 40:1030–1048.

De Mauro, T., Mancini, F., Vedovelli, M., and Voghera, M. (1993). *Lessico di frequenza dell'italiano parlato (LIP)*. Etaslibri.

Devereux, B., Pilkington, N., Poibeau, T., and Korhonen, A. (2009). Towards unrestricted, large-scale acquisition of feature-based conceptual representations from corpus data. *Research on Language & Computation*, 7:137–170.

Devlin, J., Gonnerman, L., Andersen, E., and Seidenberg, M. (1998). Category-specific semantic deficits in focal and widespread brain damage: A computational account. *Journal of Cognitive Neuroscience*, 10:77–94.

Doesborgh, S., van de Sandt-Koenderman, M., Dippel, D., van Harskamp, F., Koudstaal, P., and Visch-Brink, E. (2004). Effects of semantic treatment on verbal communication and linguistic processing in aphasia after stroke a randomized controlled trial. *Stroke*, 35:141–146.

Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational linguistics*, 19:61–74.

Egan, J., Worrall, L., and Oxenham, D. (2004). Accessible internet training package helps people with aphasia cross the digital divide. *Aphasiology*, 18:265–280.

Ellis, A., Young, A., and Critchley, E. (1989). Loss of memory for people following temporal lobe damage. *Brain*, 112:1469–1483.

Evert, S. (2008). Corpora and collocations. In Lüdeling, A. and Kytö, M., editors, *Corpus Linguistics. An International Handbook*, volume 2, pages 1212–1248. Mouton de Gruyter.

Fellbaum, C. (1998a). A semantic network of english verbs. In Fellbaum, C., editor, *WordNet: an electronic lexical database*, pages 69–104. MIT Press.

Fellbaum, C. (1998b). *WordNet: An electronic database*. MIT Press.

Fellbaum, C. (2002). On the semantics of troponymy. In Green, B. and (2002), M., editors, *The Semantics of Relationships*, pages 23–34. Kluwer.

Fellbaum, C. (2006). Wordnet and wordnets. In Brown, K., editor, *Encyclopedia of Language and Linguistics*, pages 665–670. Elsevier.

Fellbaum, C., Grabowski, J., Landes, S., et al. (1997). Analysis of a hand-tagging task. In *Proceedings of ANLP-97 Workshop on Tagging Text with Lexical Semantics: Why, What, and How*, pages 34–40.

Firth, J. (1957). A synopsis of linguistic theory, 1930-1955. *Studies in linguistic analysis. Special volume of the Philological Society*, pages 1–32.

Fleiss, J. (1971). Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76:378.

Fodor, J. (1998). *Concepts: Where cognitive science went wrong.* Oxford University Press.

Frassinelli, D. and Lenci, A. (2012). Concepts in context: Evidence from a feature-norming study. In *Proceedings of the Annual Meeting of the Cognitive Science Society.*

Frith, U. (1980). *Cognitive processes in spelling.* Academic Press.

Gangemi, A., Guarino, N., Masolo, C., and Oltramari, A. (2003). Sweetening wordnet with dolce. *AI magazine*, 24:13.

Garrard, P., Ralph, M., Hodges, J., and Patterson, K. (2001). Prototypicality, distinctiveness, and intercorrelation: Analyses of the semantic attributes of living and nonliving concepts. *Cognitive Neuropsychology*, 18:125–174.

Gerstl, P. and Pribbenow, S. (1995). Midwinters, eng games, and body parts: a classification of part-whole relations. *International journal of human-computer studies*, pages 865–889.

Girju, R., Badulescu, A., and Moldovan, D. (2006). Automatic discovery of part-whole relations. *Computational Linguistics*, 32:83–135.

Girju, R., Nakov, P., Nastase, V., Szpakowicz, S., Turney, P., Yuret, D., et al. (2007). Semeval-2007 task 04: Classification of semantic relations between nominals. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 13–18.

Gonnerman, L., Andersen, E., Devlin, J., Kempler, D., and Seidenberg, M. (1997). Double dissociation of semantic categories in alzheimer's disease. *Brain and language*, 57:254–279.

Goodglass, H. and Wingfield, A. (1997). *Anomia: Neuroanatomical and Cognitive Correlates.* Foundations of Neuropsychology. Academic Press.

Guarino, N. and Welty, C. (2000). A formal ontology of properties. *Knowledge Engineering and Knowledge Management Methods, Models, and Tools*, pages 191–230.

Hart, J., Berndt, R., and Caramazza, A. (1985). Category-specific naming deficit following cerebral infarction. *Nature*, 316:439–440.

Hearst, M. (1992). Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 14th conference on Computational linguistics*, pages 539–545.

Hearst, M. (1998). Automated discovery of wordnet relations. In Fellbaum, C., editor, *WordNet: an electronic lexical database*, pages 131–151. MIT Press.

Hendrickx, I., Kim, S., Kozareva, Z., Nakov, P., Séaghdha, D., Padó, S., Pennacchiotti, M., Romano, L., and Szpakowicz, S. (2009). Semeval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals. In *Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions*, pages 94–99.

Hillis, A. and Caramazza, A. (1991). Category-specific naming and comprehension impairment: A double dissociation. *Brain*, 114:2081–2094.

Hinton, G. and Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological review*, 98:74–95.

Hirst, G. and St-Onge, D. (1998). Lexical chains as representations of context for the detection and correction of malapropisms. In Fellbaum, C., editor, *WordNet: an electronic lexical database*, pages 305–332. MIT Press.

Horton, S. and Byng, S. (2002). Semantic therapy in day-to-day clinical practice: perspectives on diagnosis and therapy related to semantic impairments in aphasia. In *The Handbook of Adult Language Disorders Integrating Cognitive Neuropsychology, Neurology, and Rehabilitation*, pages 229–252. Psychology Press.

Izquierdo, R., Suárez, A., and Rigau, G. (2007). Exploring the automatic selection of basic level concepts. In *Proceedings of Recent Advances in Natural Language Processing 2009*, volume 7.

Jiang, J. and Conrath, D. (1997). Semantic similarity based on corpus statistics and lexical taxonomy. In *Proc. of International Conference Research on Computational Linguistics*.

Johannsen-Horbach, H., Cegla, B., Mager, U., Schempp, B., and Wallesch, C. (1985). Treatment of chronic global aphasia with a nonverbal communication system. *Brain and language*, 24:74–82.

Jurafsky, D. and Martin, J. (2009). *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics [2nd edition]*. Prentice-Hall.

Katz, R. (2009). Application of computers to the treatment of us veterans with aphasia. *Aphasiology*, 23:1116–1126.

Katz, R. and Nagy, V. (1982). A computerized treatment system for chronic aphasic patients. In *Proceedings of the Clinical Aphasiology Conference*.

Kay, J. (1992). *Palpa: Psycholinguistic Asessments Of Language Processing In Aphasia*. Taylor & Francis.

Kay, J. and Ellis, A. (1987). A cognitive neuropsychological case study of anomia. implications for psychological models of word retrieval. *Brain*, 110 (Pt 3):613–629.

Kelly, C., Devereux, B., and Korhonen, A. (2010). Acquiring human-like feature-based conceptual representations from corpora. In *Proceedings of the First Workshop on Computational Neurolinguistics*, pages 61–69.

Kelly, C., Devereux, B., and Korhonen, A. (2012). Semi-supervised learning for automatic conceptual property extraction. In *Proceedings of the NAACL HLT 2012*, pages 11–21.

Kitzing, P., Ahlsén, E., and Jönsson, B. (2005). Communication aids for people with aphasia. *Logopedics Phonatrics Vocology*, 30:41–46.

Kohl, K., Jones, D., Berwick, R., and Nomura, N. (1998). Representing verb alternations in wordnet. In Fellbaum, C., editor, *WordNet: an electronic lexical database*, pages 153–178. MIT Press.

Kremer, G. and Baroni, M. (2010). Predicting cognitively salient modifiers of the constitutive parts of concepts. In *Proceedings of the 2010 Workshop on Cognitive Modeling and Computational Linguistics*, pages 54–62.

Kremer, G. and Baroni, M. (2011). A set of semantic norms for german and italian. *Behavior Research Methods*, 43:97–109.

Krippendorff, K. (2004). Reliability in content analysis: some common misconceptions and recommendations. *Human Communication Research*, 30:411–433.

Krippendorff, K. (2008). Testing the reliability of content analysis data. In Bock, M., editor, *The content analysis reader*, pages 350–357. Sage Publications.

Laham, D. (1997). Latent semantic analysis approaches to categorization. In *Proceedings of the 19th annual conference of the Cognitive Science Society*, page 979.

Laiacona, M., Barbarotto, R., and Capitani, E. (1993a). Perceptual and associative knowledge in category specific impairment of semantic memory: A study of two cases. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*.

Laiacona, M., Barbarotto, R., Trivelli, C., and Capitani, E. (1993b). Dissociazioni semantiche intercategoriali: descrizione di una batteria standardizzata e dati normativi. *Archivio di Psicologia, Neurologia e Psichiatria*.

Laiacona, M., Capitani, E., and Barbarotto, R. (1997). Semantic category dissociations: A longitudinal study of two cases. *Cortex*, 33:441–461.

Landauer, T. and Dumais, S. (1997). A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review; Psychological Review*, 104:211.

Laparra, E., Rigau, G., and Cuadros, M. (2010). Exploring the integration of wordnet and framenet. In *Proceedings of the 5th Global WordNet Conference*.

Lebani, G. E. and Pianta, E. (2010a). Exploiting lexical resources for therapeutic purposes: the case of WordNet and STaRS.syslebani2010:cogalex. In *Proceedings of the 2nd Workshop on Cognitive Aspects of the Lexicon*, pages 12–17.

Lebani, G. E. and Pianta, E. (2010b). A feature type classification for therapeutic purposes: a preliminary evaluation with non-expert speakers. In *Proceedings of the Fourth Linguistic Annotation Workshop*, pages 157–161.

Lebani, G. E. and Pianta, E. (2010c). Human language technologies supporting therapeutic practices for language disorders: the project STaRS.sys. In *Proceedings of the 7th annual meeting of the Italian Society of Cognitive Science*, pages 52–56.

Lebani, G. E. and Pianta, E. (2012). Encoding commonsense lexical knowledge into Word-Net. In *Proceedings of the 6th Global WordNet Conference*, pages 159–166.

Lenat, D. (1995). Cyc: A large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38:33–38.

Lenci, A., Busa, F., Ruimy, N., Gola, E., Monachini, M., Calzolari, N., Zampolli, A., Pustejovsky, J., Ogonowski, A., McCawley, C., et al. (2000). Simple linguistic specifications. *Deliverable D2*.

Lincoln, N., Mulley, G., Jones, A., McGuirk, E., Lendrem, W., and Mitchell, J. (1984). Effectiveness of speech therapy for aphasic stroke patients: a randomised controlled trial. *The Lancet*, 323:1197–1200.

Liu, H. and Singh, P. (2004). Commonsense reasoning in and over natural language. In *Proceedings of the 8th International Conference on Knowledge-Based Intelligent Information and Engineering Systems*, pages 293–306.

LoBue, P. and Yates, A. (2011). Types of common-sense knowledge needed for recognizing textual entailment. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*.

Lund, K. and Burgess, C. (1996). Hyperspace analog to language (hal): A general model of semantic representation. *Language and Cognitive Processes*.

Lyons, J. (1977). *Semantics*. Cambridge University Press.

Magnini, B. and Speranza, M. (2001). Integrating generic and specialized wordnets. *Proceedings of the Euroconference Recent Advances in NLP 2001*.

Mahon, B. and Caramazza, A. (2009). Concepts and categories: a cognitive neuropsychological perspective. *Annual review of psychology*, 60:27.

Malt, B., Sloman, S., Gennari, S., Shi, M., and Wang, Y. (1999). Knowing versus naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory and Language*, 40:230–262.

Mawson, C. O. S. (1911). *Roget's Thesaurus of English Words and Phrases Classified and Arranged so as to Facilitate the Expression of Ideas and Assist in Literary Composition*. McDevitt-Wilson's.

McRae, K. and Cree, G. (2002). Factors underlying category-specific semantic deficits. In Forde, E. and Humphreys, G., editors, *Category specificity in brain and mind*, page 211. Psychology Press.

McRae, K., Cree, G., Seidenberg, M., and McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, 37:547–559.

McRae, K., Cree, G., Westmacott, R., and De Sa, V. (1999). Further evidence for feature correlations in semantic memory. *Canadian Journal of Experimental Psychology*, 53:360.

McRae, K., De Sa, V., and Seidenberg, M. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, 126:99–130.

Meyer, D., Zeileis, A., and Hornik, K. (2006). The strucplot framework: Visualizing multi-way contingency tables with vcd. *Journal of Statistical Software*, 17.

Miceli, G., Capasso, R., Daniele, A., Esposito, T., Magarelli, M., and Tomaiuolo, F. (2000). Selective deficit for people's names following left temporal damage: An impairment of domain-specific conceptual knowledge. *Cognitive Neuropsychology*, 17:489–516.

Miller, G. (1998a). Nouns in wordnet. In Fellbaum, C., editor, *WordNet: an electronic lexical database*, pages 23–46. MIT Press.

Miller, G., Beckwith, R., Fellbaum, C., Gross, D., and Miller, K. (1990). Introduction to wordnet: An on-line lexical database. *International journal of lexicography*, 3:235–244.

Miller, G. and Hristea, F. (2006). Wordnet nouns: Classes and instances. *Computational Linguistics*, 32:1–3.

Miller, G. A. (1995). Wordnet: a lexical database for english. *Communications of the ACM*, 38:39–41.

Miller, K. (1998b). Modifiers in wordnet. In Fellbaum, C., editor, *WordNet: an electronic lexical database*, pages 47–67. MIT Press, Cambridge, MA.

Milner, A. and Goodale, M. (1993). Visual pathways to perception and action. In Hicks, T., Molotchnikoff, S., and Ono, T., editors, *Progress in Brain Research*, volume 95, pages 317–337. Elsevier Science.

Mohammad, S. and Hirst, G. (2012). Distributional measures as proxies for semantic relatedness. *arXiv preprint arXiv:1203.1889*.

Morato, J., Marzal, M., Lloréns, J., and Moreiro, J. (2004). Wordnet applications. In *Proceedings of the second Global Wordnet Conference*, volume 2004.

Morris, J. and Hirst, G. (2004). Non-classical lexical semantic relations. In *Proceedings of the HLT-NAACL Workshop on Computational Lexical Semantics*, pages 46–51.

Moschitti, A. (2004). A study on convolution kernels for shallow semantic parsing. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, page 335.

Moss, H., Tyler, L., and Devlin, J. (2002). The emergence of category-specific deficits in a distributed semantic system. In Humphreys, E. M. E. F. . G. W., editor, *Category Specificity in Brain and Mind*, pages 115–147. Psychology Press.

Murphy, G. L. (2002). *The big book of concepts*. MIT Press.

Nickels, L. (2002). Therapy for naming disorders: Revisiting, revising, and reviewing. *Aphasiology*, 16:935–979.

Niemann, E. and Gurevych, I. (2011). The people's web meets linguistic knowledge: Automatic sense alignment of wikipedia and wordnet. In *Proceedings of the 9th International Conference on Computational Semantics*, pages 205–214.

Nikolova, S., Boyd-Graber, J., and Cook, P. (2009a). The design of viva: a mixed-initiative visual vocabulary for aphasia. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, pages 4015–4020.

Nikolova, S., Boyd-Graber, J., and Fellbaum, C. (2012). Collecting semantic similarity ratings to connect concepts in assistive communication tools. *Modeling, Learning, and Processing of Text Technological Data Structures*, pages 81–93.

Nikolova, S., Boyd-Graber, J., Fellbaum, C., and Cook, P. (2009b). Better vocabularies for assistive communication aids: connecting terms using semantic networks and untrained annotators. In *Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility*, pages 171–178.

Oltramari, A., Gangemi, A., Guarino, N., and Masolo, C. (2002). Restructuring wordnet's top-level: The ontoclean approach. In *Proceedings of the Third International Language Resources and Evaluation*.

Palmer, M., Dang, H., and Fellbaum, C. (2007). Making fine-grained and coarse-grained sense distinctions, both manually and automatically. *Natural Language Engineering*, 13:137.

Patwardhan, S., Banerjee, S., and Pedersen, T. (2003). Using measures of semantic relatedness for word sense disambiguation. *Computational Linguistics and Intelligent Text Processing*, pages 241–257.

Pazienza, M., Stellatoa, A., and Tudoracheab, A. (2008). A bottom-up comparative study of eurowordnet and wordnet 3.0 lexical and semantic relations. In *Proceedings of the Sixth International Language Resources and Evaluation*, pages 2293–2299.

Pedersen, P. M., Jorgensen, H. S., Nakayama, H., Raaschou, H. O., and Olsen, T. S. (1995). Aphasia in acute stroke: incidence, determinants, and recovery. *Annals of Neurology*, 38:659–666.

Petheram, B. (2004). Special issue on computers and aphasia. *Aphasiology*, 18:187–282.

Petheram, B. and Enderby, P. (2008). The role of electronic devices in the rehabilitation of language disorders. In Stemmer, B. and Whitaker, H., editors, *Handbook of the Neuroscience of Language*, pages 427–432. Elsevier Science.

Pianta, E., Bentivogli, L., and Girardi, C. (2002). Developing an aligned multilingual database. In *Proceedings of Global WordNet Conference 2002*.

Pianta, E., Girardi, C., and Zanoli, R. (2008). The textpro tool suite. In *Proceedings of the Sixth International Language Resources and Evaluation*.

Plaut, D. (2002). Graded modality-specific specialisation in semantics: A computational account of optic aphasia. *Cognitive Neuropsychology*, 19:603–639.

Poesio, M. and Almuhareb, A. (2005). Identifying concept attributes using a classifier. In *Proceedings of the ACL-SIGLEX Workshop on Deep Lexical Acquisition*, pages 18–27.

Poesio, M., Barbu, E., Giuliano, C., and Romano, L. (2008). Supervised relation extraction for ontology learning from text based on a cognitively plausible model of relations. *Proceedings of Worksop Ontology Learning and Population*, pages 1–5.

Poprat, M., Beisswanger, E., and Hahn, U. (2008). Building a BIOWORDNET by using wordnet's data formats and wordnet's software infrastructure–a failure story. *Software Engineering, Testing, and Quality Assurance for Natural Language*, pages 31–39.

Pustejovsky, J. (1995). *The generative lexicon*. MIT press.

Raymer, A. and Rothi, L. G. (2002). Clinical diagnosis and treatment of naming disorders. In Hillis, A., editor, *The Handbook of Adult Language Disorders*, pages 163–182. Taylor & Francis.

Raymer, A. and Rothi, L. G. (2000). Aphasia and language: Theory to practice. In Nadeau, S. E., Rothi, L. J. G., and Crosson, B., editors, *The Handbook of Adult Language Disorders*, pages 108–132. Guilford Press.

Riddoch, M. and Humphreys, G. (1987). A case of integrative visual agnosia. *Brain*, 110:1431–1462.

Rosch, E. and Mervis, C. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive psychology*, 7:573–605.

Roventini, A., Alonge, A., Bertagna, F., Magnini, B., and Calzolari, N. (2000). Italwordnet: a large semantic database for italian. In *Proceedings of LREC*, pages 783–790.

Roventini, A. and Marinelli, R. (2004). Extending the italian wordnet with the specialized language of the maritime domain. In *Proceedings of the 2nd International Global Wordnet Conference*, pages 193–198.

Sagri, M., Tiscornia, D., and Bertagna, F. (2004). Jur-wordnet. In *Proceedings of the 2nd International Global Wordnet Conference*, pages 305–310.

Sahlgren, M. (2006). *The Word-Space Model: Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces*. PhD thesis.

Samson, D. and Pillon, A. (2003). A case of impaired knowledge for fruit and vegetables. *Cognitive Neuropsychology*, 20:373–400.

Sartori, G. and Lombardi, L. (2004). Semantic relevance and semantic disorders. *Journal of Cognitive Neuroscience*, 16:439–452.

Semenza, C. (1999). Lexical-semantic disorders in aphasia. In Denes, G. and Pizzamiglio, L., editors, *Handbook of clinical and experimental neuropsychology*, pages 215–244. Psychology Press.

Shapiro, K. and Caramazza, A. (2003). The representation of grammatical categories in the brain. *Trends in Cognitive Sciences*, 7:201–206.

Snow, R., O'Connor, B., Jurafsky, D., and Ng, A. (2008). Cheap and fast—but is it good?: evaluating non-expert annotations for natural language tasks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 254–263. Association for Computational Linguistics.

Springer, L. (2008). Therapeutic approaches in aphasia rehabilitation. In Stemmer, B. and Whitaker, H., editors, *Handbook of the Neuroscience of Language*, pages 397–406. Elsevier Science.

Steele, R., Weinrich, M., Wertz, R., Kleczewska, M., and Carlson, G. (1989). Computer-based visual communication in aphasia. *Neuropsychologia*, 27:409–426.

Steyvers, M. (2010). Combining feature norms and text data with topic models. *Acta Psychologica*, 133:234–243.

Steyvers, M., Smyth, P., and Chemuduganta, C. (2010). Combining background knowledge and learned topics. *Topics in Cognitive Science*, 3:18–47.

Storms, G., Navarro, D., and Lee, M. (2010). *Acta psychologica special issue on formal modeling of semantic concepts*, volume 133. Elsevier.

Strapparava, C. and Valitutti, A. (2004). Wordnet-affect: an affective extension of wordnet. In *Proceedings of the Fourth International Language Resources and Evaluation*, pages 1083–1086.

Taylor, K., Devereux, B., Acres, K., Randall, B., and Tyler, L. (2011). Contrasting effects of feature-based statistics on the categorisation and basic-level identification of visual objects. *Cognition*.

Tsochantaridis, I., Joachims, T., Hofmann, T., and Altun, Y. (2006). Large margin methods for structured and interdependent output variables. *Journal of Machine Learning Research*, 6:1453–1484.

Turgeon, Y. and Macoir, J. (2008). Classical and contemporary assessment of aphasia and acquired disorders of language. In Stemmer, B. and Whitaker, H., editors, *Handbook of the Neuroscience of Language*, pages 3–11. Elsevier Science.

Turney, P., Pantel, P., et al. (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, 37:141–188.

Ulivieri, M., Bianco, I., Guazzini, E., and Molino, S. (2006). Some aspects of the parole-simple-clips semantic layer: Uses and advantages. *Literary and linguistic computing*, 21:143–156.

Uyeda, K. and Mandler, G. (1980). Prototypicality norms for 28 semantic categories. *Behavior Research Methods*, 12:587–595.

van de Sandt-Koenderman, M. (2004). High-tech aac and aphasia: Widening horizons? *Aphasiology*, 18:245–263.

van de Sandt-Koenderman, W. (2011). Aphasia rehabilitation and the role of computer technology: Can we keep up with modern times? *International journal of speech-language pathology*, 13:21–27.

Vapnik, V. (1979). *Estimation of dependences based on empirical data (in Russian)*. Nauka.

Vieu, L. and Aurnague, M. (2005). Part-of relations, functionality and dependence. In M. Aurnague, M. H. and Vieu, L., editors, *The Categorization of Spatial Entities in Language and Cognition*, pages 307–336. John Benjamins.

Vigliocco, G., Vinson, D., Lewis, W., Garrett, M., et al. (2004). Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology*, 48:422–488.

Vigliocco, G., Warren, J., Siri, S., Arciuli, J., Scott, S., and Wise, R. (2006). The role of semantics and grammatical class in the neural representation of words. *Cerebral Cortex*, 16:1790–1796.

Vinson, D. and Vigliocco, G. (2008). Semantic feature production norms for a large set of objects and events. *Behavior Research Methods*, 40:183–190.

Vinson, D., Vigliocco, G., Cappa, S., and Siri, S. (2003). The breakdown of semantic knowledge: Insights from a statistical model of meaning representation. *Brain and Language*, 86:347–365.

Visch-Brink, E., Bajema, I., and Van De Sandt-Koenderman, M. (1997). Lexical semantic therapy: Box. *Aphasiology*, 11:1057–1078.

Vossen, P. (1998). Introduction to eurowordnet. *Computers and the Humanities, Special Issue on EuroWordNet*, 32:73–89.

Wade, D. T., Hewer, R. L., David, R. M., and Enderby, P. M. (1986). Aphasia after stroke: natural history and associated deficits. *Journal of Neurology, Neurosurgery & Psychiatry*, 49:11–16.

Wahn, C. (2002). Pcad: Portable communication assistant for people with acquired dysphasia–eine elektronische kommunikationshilfe für sprach-und sprechgestörte menschen. *Computer helfen heilen und leben. Bad Honnef: Hippocampus*, pages 295–304.

Wallesch, C. and Johannsen-Horbach, H. (2004). Computers in aphasia therapy: Effects and side-effects. *Aphasiology*, 18:223–228.

Warrington, E. and McCarthy, R. (1987). Categories of knowledge further fractionations and an attempted integration. *Brain*, 110:1273–1296.

Warrington, E. and McCarthy, R. (1994). Multiple meaning systems in the brain: A case for visual semantics. *Neuropsychologia*, 32:1465–1473.

Warrington, E. and Shallice, T. (1984). Category specific semantic impairments. *Brain*, 107:829–853.

Winston, M., Chaffin, R., and Herrmann, D. (1987). A taxonomy of part-whole relations. *Cognitive science*, 11:417–444.

Wittgenstein, L. (1953). *Philosophical investigations*. Macmillan,.

Wolf, E. and Gurevych, I. (2010). Aligning sense inventories in wikipedia and wordnet. In *Proceedings of the 1st Workshop on Automated Knowledge Base Construction*, pages 24–28.

Wu, L. and Barsalou, L. (2009). Perceptual simulation in conceptual combination: Evidence from property generation. *Acta Psychologica*, 132:173–189.

Zelenko, D., Aone, C., and Richardella, A. (2003). Kernel methods for relation extraction. *The Journal of Machine Learning Research*, 3:1083–1106.

Acknowledgments

First and foremost, I'd like to thank all the people that contributed to the work presented here. My advisor Emanuele Pianta, the participants of the evaluations reported in chapter 5 and those of the norming study in chapter 6; Marco Baroni and Gerhard Kremer for their priceless suggestions and for giving us a non-normalized version of their norms; prof. Gabriele Miceli, Rita Capasso and Alessia Monti for their suggestions on the application scenario in chapter 3; Colin Kelly, Massimo Poesio for crucial explanations.

My gratitude goes also to all those people that supported me in a more indirect way: Massimo Turatto (5) and Thomas Mastropasqua (8) from CIMeC FC; Sara Tonelli and Bernardo Magnini from FBK; Andrew Anderson, Brian Murphy, Federico Boschetti and Eva Maria Vecchi from the CLIC lab at CIMeC; Katya Tentori, Veronica Mazza and Claudia Bonfiglioli from CIMeC; Stella Tsigas, Sharon Shan-Shan Chan, Sara Fabbri, Raffaele Tucciarelli, Elisa Infanti, Guido and MC Rummolino, Benedetta Heimler and some of the other CoBraS; Luca Polonio and Francesco Vespignani from DiSCoF.

Finally, I'd like to thank the most important people around me, those that patiently understood my absence and my cyclic bad mood: Mum and Dad, Marianna and Andrea, Roberta, Cataldo, Lucilla. Really love each one of you. Now what?

Oltrona al Lago,
November 5, 2012