



## **Danish Academic Vocabulary**

### **Four studies on the words of academic written Danish**

Jakobsen, Anne Sofie

*Publication date:*  
2018

*Document version*  
Publisher's PDF, also known as Version of record

*Document license:*  
[CC BY-NC-ND](#)

*Citation for published version (APA):*  
Jakobsen, A. S. (2018). *Danish Academic Vocabulary: Four studies on the words of academic written Danish*. Det Humanistiske Fakultet, Københavns Universitet.

UNIVERSITY OF COPENHAGEN  
DEPARTMENT OF ENGLISH, GERMANIC AND ROMANCE STUDIES



---

## **PhD Thesis**

Anne Sofie Jakobsen

# Danish Academic Vocabulary

Four studies on the words of academic written Danish

Primary supervisor: Professor MSO Birgit Henriksen

Submitted on: 31 August 2018

Name of department: Department of English, Germanic and Romance Studies; Centre for Internationalisation and Parallel Language Use

Author: Anne Sofie Jakobsen

Title: Danish Academic Vocabulary – Four studies on the words of academic written Danish

Danish Title: Dansk akademisk ordforråd – Fire undersøgelser af ordene i dansk akademisk skriftsprog

Primary supervisor: Birgit Henriksen

Secondary supervisors: Averil Coxhead and Sussi Olsen

Submitted: 31 August 2018

## Acknowledgements

First of all, I must thank my primary supervisor, Birgit Henriksen for her continuing encouragement to embark on a PhD, for her enormously supportive, insightful, and creative mentoring, and for keeping my spirits up and providing me with invaluable input in my writings. As my secondary supervisor, Averil Coxhead has been an inspiring and encouraging supervisor and I cannot thank her enough for her input and guidance throughout the project. Moreover, I owe Sussi Olsen as my other secondary supervisor so much for her knowledgeable input which in particular has been integral to my work on compiling a corpus of academic Danish.

The following persons have helped me in various ways and I sincerely thank them: Bart Jongejan, the Centre for Language Technology, Jørg Asmussen, Det Danske Sprog- og Litteraturselskab, Sabine Kirchmeier and Philip Didrichsen, Dansk Sprognævn, and Suzanne Løje, the Faculty of Humanities, UCPH.

I would also like to thank all the members of the LUNAS<sup>1</sup> research network for their inspiring work on Norwegian and Swedish academic word lists.

Thank you to the Department of English, Germanic and Romance Studies and to the Centre for Internationalisation and Parallel Language Use for housing me during these three years. Thanks to all my colleagues for being the best colleagues in the world. I could not have done this without you. Especially thanks to Karen-Margrete Frederiksen for all your encouragement and good advice. Thank you to Charlotte Øhrstrøm, Martin C. Unger, Anne Marie Dyrberg, and Joyce Kling Sørensen for all their help and support. Thank you to Sanne Larsen for agreeing to teach a course on academic language and vocabulary with me, and for all your input and for all the discussions we have had. I am especially grateful to Sophie Swerts Knudsen for her proofreading of this thesis.

A special thanks to my PhD friend Camilla Falk Rønne Nissen whose company during the last three years has been invaluable.

Last, but not least, I am eternally grateful for the support from family and friends during these three PhD years.

---

<sup>1</sup> The LUNAS research network (2013-2016, funded by Nordplus) was established with the aim of exploring the academic vocabularies of Swedish, Norwegian, and Danish ([www.cip.ku.dk/english/research/network/lunas/](http://www.cip.ku.dk/english/research/network/lunas/)).

# Table of contents

Acknowledgements .....	1
Table of contents .....	2
List of tables .....	6
List of figures .....	9
List of appendices .....	10
Chapter 1. Introduction .....	11
1.1. Framing the four studies: Challenges of academic language and academic vocabulary .....	12
1.2. Research purpose .....	15
1.3. Thesis outline .....	15
Chapter 2. Vocabulary .....	17
2.1. Defining words .....	17
2.1.2. Unit of counting .....	18
2.2. Word knowledge .....	23
2.2.2. How many words are needed to comprehend language? .....	25
2.3. Vocabulary categories .....	26
2.3.1. Specialised vocabulary .....	27
2.3.2. Frequency of occurrence and general high frequency vocabulary .....	29
2.3.3. Divisions of vocabulary in the Nordic languages .....	31
2.3.4. Academic vocabulary .....	37
2.4. Functions of academic vocabulary .....	41
2.5. Summary .....	46
Chapter 3. Word lists .....	47
3.1. Introduction .....	47
3.2. Word lists in general .....	47
3.3. Danish word lists of general high frequency vocabulary .....	48
3.3.1. Most frequently used lemmas in Danish .....	48
3.3.2. Danske kerneord (Danish core words) .....	50
3.3.3. Dansk frekvensordbog (Danish frequency dictionary) .....	52
3.4. Pedagogical word lists .....	53
3.4.1. Developing pedagogical word lists .....	55
3.4.2. Academic word lists .....	60
3.5. Rationale for the studies of this thesis .....	74
Chapter 4. Corpora .....	76
4.1. Introduction .....	76

4.2. Overview of corpora used in this thesis.....	76
4.3. Corpus design .....	78
4.4. The AcaDan corpus – design and compilation.....	78
4.4.1. Language represented: Professional academic writing.....	79
4.4.2. Academic disciplines .....	81
4.4.3. Corpus platform.....	84
4.4.4. Text collection .....	84
4.4.5. Size of the AcaDan Corpus .....	89
4.4.6. Authors, publications, and time of publication in the AcaDan Corpus .....	90
4.4.7. Metadata in the AcaDan Corpus.....	92
4.4.8. Copyright of the AcaDan Corpus texts.....	92
4.4.9. Text preparation, corpus upload, and sub-corpora .....	93
4.4.10. Summary: Description of the AcaDan Corpus .....	94
4.5. The second academic language corpus.....	95
4.6. General language corpora.....	96
4.7. Summary.....	97
Chapter 5. Study 1 – General and academic high frequency vocabulary in Danish.....	98
5.1. Introduction and research questions .....	98
5.2. Methodology.....	99
5.2.1. Data.....	99
5.2.2. Data analysis.....	100
5.3. Results .....	105
5.4. Discussion.....	112
5.5. Limitations.....	116
5.6. Rationale for Study 2.....	117
Chapter 6. Study 2 – A Danish Academic Word List.....	118
6.1. Introduction and research questions .....	118
6.2. Identifying academic words - vocabulary selection for the DAWL.....	120
6.2.1. Corpora data .....	120
6.2.2. Unit of counting.....	122
6.2.3. Data analysis.....	123
6.3. Summary.....	148
6.4. Results: A Danish Academic Word List.....	148
6.4.1. Research question 1: The Danish Academic Word List.....	148
6.4.2. Research question 2: Parts of speech of the DAWL lemmas .....	152
6.4.3. Research question 3: Evaluation of the DAWL .....	157

6.4.4. Research question 4: General high frequency words among the DAWL lemmas .....	159
6.4.5. Research question 5: Lexical coverage of the general high frequency words in the DAWL .....	162
6.4.6. Specifications of the Danish Academic Word List.....	165
6.5. Discussion.....	166
6.5.1. Limitations.....	170
6.6. Summary and rationale for Study 3 .....	171
Chapter 7. Study 3 – Exploring the functions of the DAWL words .....	173
7.1. Introduction and research question.....	173
7.2. A functional categorisation of the DAWL words.....	174
7.2.1. The first analysis: Macro categories .....	174
7.2.2. The second analysis: A functional framework of Danish academic vocabulary .....	175
7.4. Discussion and limitations.....	203
7.5. Rationale for Study 4.....	205
Chapter 8. Study 4 – Supplementing the DAWL.....	207
8.1. Introduction and research questions .....	207
8.2. Supplementing the DAWL .....	208
8.2.1. Morphological relatedness: derivations and compounds.....	209
8.2.2. Procedure for supplementing the DAWL .....	211
8.2.3. Measuring the lexical coverage of the S-DAWL.....	213
8.3. The S-DAWL .....	213
8.3.1. Description of the added lemmas: abbreviations, compounds, and derivations .....	216
8.3.2. Description of the S-DAWL.....	226
8.3.3. Lexical coverage of the added lemmas and the S-DAWL.....	229
8.4. Summary.....	231
8.5. Danish academic vocabulary represented by the S-DAWL .....	232
8.6. Word groups .....	234
8.7. Limitations and further research.....	236
8.8. Rationale for Chapter 9.....	237
Chapter 9. Key findings and conclusion .....	238
9.1. Introduction .....	238
9.2. Key findings and contributions.....	239
9.2.1. The nature of Danish general vocabulary .....	239
9.2.2. The nature of Danish academic vocabulary.....	240
9.2.3. The overlaps between vocabulary categories .....	241
9.2.4. Implications for pedagogy .....	243
9.3. Suggestions for further research.....	245

9.4. Concluding remarks.....	246
References.....	248
Abstract.....	271
Abstract in Danish.....	272
Appendices.....	273



## List of tables

Table 2.1. Overview of units of counting in word lists.....	19
Table 2.2. The seven levels of the word family framework (Bauer and Nation, 1993).....	20
Table 2.3. What is involved in knowing a word (Nation, 2013).....	24
Table 2.4. High, mid-, and low frequency vocabulary according to frequency levels .....	30
Table 2.5. Categories of textbook vocabulary (Järborg, 2007).....	35
Table 2.6. Categories of textbook vocabulary (Kanebrant, 2015) .....	36
Table 2.7. Martin's (1976) categorisation of academic vocabulary with examples .....	43
Table 2.8. A pragmatic-functional taxonomy of academic formulas (Simpson-Vlach & Ellis, 2010) .....	44
Table 3.1. The first ten lemmas in the DSL list .....	49
Table 3.2. Excerpt from Danske Kerneord (Ruus, 1995) .....	51
Table 3.3. Extract from Dansk frekvensordbog (Bergenholtz, 1992).....	53
Table 3.4. Four academic word lists in English.....	68
Table 3.5. Overview of the Swedish Academic Word List .....	70
Table 3.6. Overview of the Norwegian Academic Word List .....	72
Table 4.1. Overview of corpora used in this thesis .....	77
Table 4.2. The six research domains as outlined in the OECD classification .....	82
Table 4.3. Classifications of academic disciplines .....	83
Table 4.4. The four academic disciplines represented in the AcaDan corpus .....	84
Table 4.5. Texts and running words in the AcaDan corpus.....	87
Table 4.6. Three sub-corpora of the AcaDan corpus used in Studies 1 and 2 .....	87
Table 4.7. Six sub-corpora of the AcaDan corpus used in Study 2 .....	87
Table 4.7. Sub-disciplines and OECD disciplines in the AcaDan corpus .....	88
Table 4.8. Authors in the AcaDan corpus .....	90
Table 4.9. Overview of the number of sources for the texts in the AcaDan corpus .....	91
Table 4.10. Time of publication for the texts in the AcaDan corpus .....	92
Table 4.11. Development and content of the AcaDan corpus .....	95
Table 4.12. The Second Academic Language Corpus .....	96
Table 4.13. The Journalisten.dk corpora.....	96
Table 5.1. Specifications for the lists of the 2,000 most frequently used words of Danish.....	104
Table 5.2. Coverage of the first and second 1,000-word lists of Danish in general language.....	106

Table 5.3. Coverage of the first and second 1,000-word lists of Danish over academic language .	107
Table 5.4. Coverage and means of the first and second 1,000 word lists of Danish over Medical research articles (M).....	108
Table 5.5. Coverage and means of first and second 1,000 word lists of Danish over Information science research articles (IS).....	108
Table 5.6. Twenty high frequency academic words in Danish.....	111
Table 6.1. Overview of corpora used in Study 2 .....	121
Table 6.2. Sub-corpora of the AcaDan Corpus .....	121
Table 6.3. An overview of the six sub-corpora in the AcaDan Corpus .....	122
Table 6.4. Results of the overlap analysis.....	128
Table 6.5. Results from the second overlap analysis .....	129
Table 6.6. Number of lemmas between different D values .....	132
Table 6.7. Number of lemmas at different cut-off D values .....	132
Table 6.9. Examples of lemmas with a dispersion value between 0.60 and 0.69.....	134
Table 6.10. Examples of technical lemmas with a D value between 0.70 and 0.79 and above 0.80 .....	137
Table 6.11. Overview of error items .....	140
Table 6.12. Overview of the base word lists developed on the basis of the DAWL lemmas.....	145
Table 6.13. Changes made to the 18 lemmas in the lemma list.....	147
Table 6.14. Distribution of the 758 lemmas in frequency belts of 1,000 across the AcaDan Corpus .....	149
Table 6.15. Frequency distribution of the 630 lemmas with a frequency below 1,000 across the AcaDan Corpus.....	150
Table 6.16. Examples of DAWL lemmas <100.....	150
Table 6.17. Dispersion of the DAWL lemmas in the AcaDan Corpus.....	151
Table 6.18. Distribution of part of speech in the DAWL .....	152
Table 6.19. The 21 DAWL lemmas with multiple parts of speech .....	154
Table 6.20. The 12 lemmas with are listed with separate parts of speech.....	156
Table 6.21. Nation and Parent's (2016) definition of homoforms.....	157
Table 6.22. Overview of the base word lists developed on the basis of the DAWL lemmas.....	158
Table 6.23. The coverage of the DAWL over academic and general language .....	158
Table 6.24. The coverage of the DAWL over the four disciplines in the AcaDan corpus .....	159

Table 6.25. Examples of the 212 DAWL lemmas that overlap with the 402 lemmas from Study 1 .....	159
Table 6.26. Examples of general high frequency lemmas among the DAWL .....	160
Table 6.27. The DAWL divided into frequency bands of the DAWL.....	161
Table 6.28. Comparison of frequencies for DAWL lemmas with more than one part of speech....	162
Table 6.29. The coverage of the DAWL with and without general high frequency items .....	163
Table 6.30. Coverage of the DAWL with and without general high frequency lemmas in the four academic disciplines of the AcaDan Corpus.....	164
Table 6.31. Specifications of the DAWL.....	165
Table 7.1. Framework for the functional categorisation of academic vocabulary.....	176
Table 7.2. The functional framework for the final categorisation of the DAWL words .....	179
Table 7.3. Types of linking words in the DAWL .....	185
Table 7.4. Examples of DAWL words in the scholarly processes sub-category .....	189
Table 7.5. DAWL words with stance functions.....	195
Table 7.6. DAWL words only occurring as phrasal elements .....	200
Table 7.7. DAWL words occurring as phrasal elements and as single-word units .....	201
Table 7.8. Modifiers in the phrase ‘i x grad’ in the AcaDan corpus.....	203
Table 8.1. The basis for the supplementation analysis .....	208
Table 8.2. Derivations and inflections of <b>anvende</b> .....	210
Table 8.3. Examples of DAWL lemmas and the added lemmas .....	215
Table 8.4. Distribution of the added lemmas according to dispersion and morphological relatedness .....	216
Table 8.5. S-DAWL compounds and their ACL equivalents .....	219
Table 8.6. Derivations with prepositional prefixes .....	223
Table 8.7. General high frequency words among the added lemmas .....	226
Table 8.8. Dispersion of the S-DAWL lemmas in the AcaDan Corpus .....	227
Table 8.9. Frequency distribution of the added lemmas in the AcaDan Corpus .....	228
Table 8.10. Part of speech distribution of the added lemmas and of the DAWL .....	229
Table 8.11. The coverage of the S-DAWL and the added lemmas over academic and general language .....	230
Table 8.12. The coverage of the S-DAWL and the added lemmas over the four disciplines in the AcaDan corpus.....	230
Table 8.13. Grouping of lemmas related to <b>forstå</b> .....	236

## List of figures

Figure 2.1. The vocabulary circle – the vocabulary categories of academic language.....	28
Figure 2.2. Hirsh’s (2004) framework for functional analysis of academic words .....	45
Figure 4.1. Overview of the development of the AcaDan corpus.....	79
Figure 5.1. The lemmas være (to be) and have (to have).....	103
Figure 5.2. Text excerpts from Medicine and Information Science .....	109
Figure 5.3. Three examples of the 402 academic high frequency lemmas in Danish research articles .....	112
Figure 6.1. The vocabulary selection process of the DAWL.....	124
Figure 6.2. The process of the range analysis .....	126
Figure 6.3. Development of the DAWL .....	143
Figure 6.3. The lemma <b>accept</b> in the base word list.....	145
Figure 8.1. The S-DAWL represented as a circle based on dispersion values .....	233
Figure 8.2. The S-DAWL represented as a circle based on morphological relatedness.....	234
Figure 9.1. The vocabulary circle – the vocabulary categories of academic language.....	242

## List of appendices

Appendix A. Texts in the AcaDan Corpus arranged according to academic discipline .....	274
Appendix B. Texts in the Second Academic Corpus arranged according to academic discipline ..	298
Appendix C. Texts in the General Language Corpus arranged according to type.....	300
Appendix D. Co-author Declaration and Confirmation.....	316
Appendix E. 402 academic lemmas .....	319
Appendix F. The DAWL .....	329
Appendix G. The S-DAWL .....	347

## Chapter 1. Introduction<sup>2</sup>

Academic language, an essential aspect of academic competence, is an important tool for gaining, sharing, and developing knowledge within any field of study, enabling students and researchers to develop and convey abstract and technical ideas, and facts about complex phenomena. Learning to use a technical language, as well as a general academic language, i.e. developing academic literacy skills, allows academics to analyse, synthesize and express relations between concepts and phenomena. Academics therefore need to master both technical vocabulary within their own field of study as well as a more general academic vocabulary used frequently in a range of study areas across faculty boundaries. The hallmark of academic discourse is precision and nuance in expression, often at word choice level, e.g. created through the appropriate use of general, technical and academic vocabulary. Academic vocabulary can be considered the glue of academic language, serving a range of functions in relation to presenting information, building argumentation, scaffolding, signposting, quantifying, and stance-setting the information to be conveyed. Thus, in order to convey information and create argumentation in a clear and convincing way, academic vocabulary must be easily accessible to academic readers and writers, enabling them to focus their attentional resources on the central content and the academic argument itself. Research on academic language must therefore not only focus on the macro-level of discourse structure, but also on the micro-level of academic vocabulary. A large body of English academic vocabulary research exists, both in the form of academic language corpora and studies of academic language use in different disciplines and contexts (e.g. Hyland, 2004; Biber, 2006; Snow & Uccelli, 2009; Ädel & Erman, 2012), as well as studies on academic words and phrases (e.g. Coxhead, 2000; Simpson-Vlach & Ellis, 2010; Gardner & Davies, 2014; Malmström, Pecorari, & Gustafsson, 2016). In contrast, very little research on Danish academic vocabulary has been carried out, and no corpora of Danish academic language or no academic word list for Danish exist.

The purpose of the research presented in this thesis is therefore to address this gap in Danish lexical research by investigating the academic words of Danish professional academic writing. This is done through four separate but related studies:

Study 1: An investigation of Danish general high frequency vocabulary and the lexical coverage of these high frequency words in general and academic Danish.

---

<sup>2</sup> Parts of Chapter 1 are adapted from Jakobsen (2017).

Study 2: The identification of Danish academic words and the vocabulary selection for a Danish academic word list (DAWL) together with an investigation of the lexical coverage of these items in academic Danish.

Study 3: A functional analysis of the DAWL words.

Study 4: Identification of words in the AcaDan Corpus that are morphologically and semantically related to words in the DAWL. This study thus provides an expanded, supplementary list: the S-DAWL.

An explicit focus on Danish academic vocabulary is motivated by the vast array of research, especially in English, which shows that this type of vocabulary is an essential component of academic language skills (e.g. Coxhead, 2000; Lindberg & Johansson Kokkinakis, 2007; Snow & Uccelli, 2009; Paquot, 2010; Nagy & Townsend, 2012; Ranney, 2012; Nation, 2013; Gardner & Davies, 2014; Golden, 2016). The outcome of the four studies of this thesis is a description of Danish academic vocabulary in relation to distribution, function, and use, which will enhance our understanding of Danish academic language. The four studies also shed light on the construct of academic vocabulary in relation to discipline-specific and general language use, and thus contribute to a general understanding of the delineations between different vocabulary categories. The theoretical and methodological insights gained for Danish academic language will enable researchers of Danish to embark on further research on the micro-level of academic discourse and within other academic genres, written as well as spoken. The compilation of the corpora of Danish academic language (the AcaDan Corpus), the extraction of the DAWL, the functional analysis, and the supplementary list will also enable researchers to compare Danish academic language use with academic language use in other languages, e.g. the other Nordic languages, English or other languages studied in the Danish educational context. Furthermore, the knowledge of Danish academic vocabulary provided by the studies of this thesis is essential for developing pedagogical tools and materials for language skills development in Danish speaking students and researchers, and for assessing their academic language proficiency.

### 1.1. Framing the four studies: Challenges of academic language and academic vocabulary

Nagy and Townsend (2012, p. 92) define academic language as “the specialized language, both written and oral, of academic settings that facilitates communication and thinking about disciplinary content.” The acquisition of academic language skills, or what Cummins (1980, 2008) terms

Cognitive Academic Language Proficiency (CALP), is therefore a matter of not only learning a new form of language, but also learning to do new things with this new language (Nagy & Townsend, 2012, p. 93). This cognitive definition of academic language emphasises that academic language is not only a conduit for academic thinking but also a facilitator of it. As such, academic language can be an obstacle in relation to understanding and producing knowledge in specialized areas of study, e.g. science subjects, and therefore demands special attention in all curricula to ensure optimal subject area learning (Snow, 2010; Snow & Uccelli, 2009; Ulriksen, Murning, & Ebbensgaard, 2009). Content learning is central in all disciplines, so the focus in teaching will often be on the more technical vocabulary of a certain discipline in order to ensure understanding of the disciplinary content. Thus, students at all levels of education need to learn how to think and function academically within specialized domains while at the same time acquiring the code to operate with and convey the acquired knowledge, i.e. not only developing general and technical vocabulary knowledge, but also developing their general academic literacy skills. In a primary or secondary school context, learning most often takes place in a country's majority language, e.g. Danish in the Danish school context, often with a mixed population of students with Danish as their L1 or L2. In tertiary education, teaching is conducted in both Danish and English, which creates a growing demand to develop academic literacy skills in both languages.

As mentioned above, little research focuses directly on the vocabulary of academic language. Danish research literature on academic language use has concentrated on the challenges of and the attitudes to academic language in the Danish educational context. This research literature indicates that Danish L1 tertiary education students struggle with understanding and especially producing academic language because of its implicitness, i.e. the fact that little direct language support is directed at mastering academic language skills (Blom et al., 2017; Kristiansen, 2010; Skov, 2006, 2013). This is further emphasised by Knudsen (Knudsen, 2009, p. 50) who, in her discussion of the invisibility of academic language, points out that mastering the academic senses of general words is a challenge for most students, and that there is a need for making the different components of academic language, including vocabulary, more visible to students. In line with this, Ulriksen, Murning, and Ebbensgaard (2009) describe how upper secondary school students experience the language, and in particular the words used by their teachers, as a foreign language distanced from everyday Danish language use. The authors compare this failure to understand the academic language used by teachers of upper secondary school with bilingual students' challenges of comprehending the grey-zone language of education (Lund & Bertelsen, 2008a).



Recent years have experienced a growing focus on students with Danish as their second language and their linguistic challenges in academia. In this context, Lund and Bertelsen (2008a, 2008b, pp. 49–50) define academic language as a distinct and cognitively demanding form of language use, and they argue that it can be difficult for L2 students to understand since their comprehension of it depends to a high degree on their understanding of the L2 in general. Lund and Bertelsen showed that L2 students who passed the Higher Education Examination, an entry exam for tertiary education L2 students with a non-qualifying entrance exam, were not adequately prepared for studying in Danish. Based on these findings, Lund and Bertelsen argue that L2 students’ struggle with understanding academic language is connected to the fact that academic language is often used in context-reduced settings that offer only few or no non-linguistic remedies for the L2 students’ comprehension of the academic issue at hand. In a large-scale investigation of Icelandic students studying in Denmark and their experiences with and needs for Danish language skills (Hauksdóttir, 2012), it was reported that the students experience shortcomings in their productive skills in relation to academic language use. Specifically, it is the knowledge of and ability to navigate the various genre demands of academic language use that pose a challenge. Moreover, vocabulary is reported as a recurring problem for the Icelandic students in their mastering of Danish academic language skills (Hauksdóttir, 2012, pp. 235–237). In her study on the disciplinary and linguistic difficulties encountered by university students with Danish as a second language, Laursen (2013) found that these difficulties exist in a “grey zone between language and discipline” (K. Å. Laursen, 2013, p. 78, my translation). Moreover, it was found that academic language functions such as analyse, define, compare, and discuss (Bailey, Butler, & Sato, 2007) used by teachers in e.g. exam questions were rarely explained to the students. Laursen argues that by rendering visible the linguistic meaning of the academic language functions, the discipline teachers can help the students in learning not only the academic language but also the disciplinary content (K. Å. Laursen, 2013, pp. 78–80). Odgaard’s (2014) study is similar to Laursen (2013) in its focus on university students with Danish as a second language and their needs for and attitudes to academic language. Through a survey and interviews with staff and students at a university department, the study confirms the implicitness of academic language norms from both staff and students’ points of view, and Odgaard (2014, pp. 93–95) argues for ways of making these norms explicit. Another investigation of the linguistic and disciplinary challenges of university students with Danish as their second language showed that 67 percent of L2 students contacting the university’s student support unit reported difficulties with their written proficiency, while 13 percent reported that vocabulary was a specific challenge for them (Møller, 2015, p. 11). Likewise, in an article on Nordic L2 students studying in Denmark, Holmen (2016) argues for a university pedagogical approach which

includes more focus on the linguistic challenges experienced by these students. Although the research reviewed above focus on different groups of students, they all point to various linguistic aspects of academic language, including vocabulary, as significant challenges for students irrespective of language background. This, in turns, highlights the fact that more empirically based research on the vocabulary categories of Danish academic language, including a specific focus on academic vocabulary, is needed.

## 1.2. Research purpose

To address the need for more research-based knowledge of Danish academic vocabulary, the investigations of this thesis have been carried out with two overall aims:

- 1) To identify a Danish academic vocabulary and provide a description of this lexical inventory.
- 2) To investigate the nature of general high frequency vocabulary in academic language.

Each of the four studies are guided by a number of research questions that are listed in the relevant chapters. The rationale for the sequence of the presentation of the studies is given in the last section of Chapter 3. Methodologically, the studies of this thesis are primarily quantitative using corpora and lexical coverage for exploring Danish academic vocabulary. Consequently, the research in this thesis required the compilations of both academic and general language corpora, and the development of word lists to be used for lexical coverage analyses. Integral to the four studies is the development and use of a corpus of professional written academic language, the AcaDan Corpus.

## 1.3. Thesis outline

This thesis consists of nine chapters.

**Part I** provides the framework for the thesis and comprises Chapters 1 to 3.

In this first chapter, I have framed and given the purpose of the four studies of this thesis.

**Chapter 2** introduces basic issues concerning vocabulary and vocabulary knowledge. The majority of Chapter 2 centres on different classifications of vocabulary with particular attention given to academic vocabulary and the functions of this lexical inventory in academic writing.

**Chapter 3** consists of an account of word lists used for research and pedagogical purposes. Three Danish studies of general vocabulary are introduced together with different academic word lists in English, Swedish, and Norwegian.

**Part II** consists of Chapters 4-8 in which the four studies of this thesis are presented together with the corpora developed and used in these four studies.

**Chapter 4** introduces the corpora used in the four studies. The majority of the chapter addresses the design and compilation of a corpus of Danish academic language, the AcaDan Corpus, developed in particular for the studies of this thesis.

**Chapter 5** presents the first study of this thesis, Study 1, which focuses on the nature of general high frequency vocabulary in both academic and general language use. A central issue is the lexical coverage analysis and the comparison with other international studies. This study is published in a revised form in the Nordand Journal, May 2018 (Jakobsen, Coxhead, & Henriksen, 2018).

**Chapter 6** presents Study 2, which involves the development of an academic word list, the DAWL, and a description of this list, including a comparison with other international studies on word lists and lexical coverage.

**Chapter 7** presents Study 3. This study comprises a functional analysis of the words identified for the Danish academic word list in Study 2.

**Chapter 8** presents the fourth and last study of this thesis. The focus of Study 4 is to identify words in the AcaDan Corpus that are morphologically and semantically related to words in the DAWL and discuss their relations to the DAWL words. In this way, Study 4 provides a supplementary, expanded list, the S-DAWL, which adds to our understanding of which lexical items may potentially be included in an extended pedagogical academic word list for Danish.

**Part III**, the final part of the thesis, draws the findings of the four studies together and consists of **Chapter 9**, which concludes the thesis by highlighting the contributions of this thesis nationally and internationally, the main discussions raised in relation to the four studies, as well as the main limitations and pedagogical implications of the research project.

## Chapter 2. Vocabulary

In this chapter, I give an account of some central issues related to the four studies presented in this thesis. The first part of the chapter centres on what a word is and what is involved in knowing a word. In the second part of the chapter, I describe and discuss different conceptualisations of vocabulary categories with a particular focus on vocabulary categorisations in the Nordic context, and on academic vocabulary and the functions it performs in academic discourse.

### 2.1. Defining words

Central to studies of vocabulary is the conceptualisation of a word. In this section, I briefly describe two types of words, single-word units, and multi-word units. Then I introduce different terms in relation to counting words. Words comprising a single string of characters surrounded by space or punctuation are in the vocabulary literature commonly referred to as single-word units to distinguish them from words consisting of more than one string of characters. This is also how most people think of a word, but investigating vocabulary also includes focusing on lexical items that are comprised by more than one string of characters<sup>3</sup>. For example, in English, compounds comprise at least two separate strings of characters which make them multi-word items as well. In contrast, in Danish, compounds are commonly written as one string of characters and can thus be perceived as single-word units. Consider the Danish noun *forskningsprojekt* and the English equivalent *research project*. In form, a multi-word unit consists of individual items, but the meaning is derived from the constellation of these items. Multi-word units also include phrases such as *på grund af* (due to), *i forbindelse med* (in connection with), and *stemme overens med* (correspond to). There is growing evidence that multi-word units are learned and stored as whole parts and not as individual items, and are therefore seen as essential for the development of language proficiency (Ellis, Simpson-Vlach, & Maynard, 2008; Wray, 2004, 2008). While much word list research in academic vocabulary has tended to focus on single words (e.g. Coxhead, 2000; Gardner & Davies, 2014), attention to multi-word units in academic discourse has increased (e.g. Biber, Conrad, & Cortes, 2004; Biber, 2006; Simpson-Vlach & Ellis, 2010; Ädel & Erman, 2012; Simonsen, 2015; Henriksen & Westbrook, 2017). Despite the importance of multi-word units, the studies of this thesis focus primarily on single-word items. This focus is motivated by the fact that there is only limited research available on Danish

---

<sup>3</sup> It should be noted that the word definition given here relies heavily on an orthographic perception of words as strings of characters with or without spaces between them.

academic vocabulary and, therefore, it seems sensible to start with single-word units as the basic word construct.

An important consideration in relation to defining a word is how to count them. A central question is whether to count every instance of a word even if it occurs multiple times within the same text or to count each item once. The terms **tokens** and **types** are used to refer to how words are counted. Tokens or running words are commonly used when giving the size of a corpus. For example, the AcaDan Corpus, used in the studies of this thesis, comprises around 3 million tokens. In contrast to tokens, types refer to unrepeatd instances of a word in a text. In the sentence “Mary gave the flowers to the maid”, there are seven tokens, also called running words, but six types as ‘the’ occurs twice. Words can also be counted in alternative ways as lemmas or word families. These two terms are described in detail in the next section under the heading “Unit of counting”. This term especially refers to how lexical items are conceptualised and organised in a word list.

### 2.1.2. Unit of counting

In deciding on an appropriate unit of counting when for example developing word lists, the morphological relationship between lexical items needs to be taken into consideration. In other words, one needs to consider and decide whether and to what degree word forms of different parts of speech such as ‘agreed’ (verb) and ‘agreement’ (noun) should be listed together or separately. Moreover, the choice of unit of counting for a particular word list should be closely connected to purpose of the list (Nation, 2016, p. 21) which will be detailed below. The unit of counting for word lists is typically either **word families** or **lemmas**, but words can also be counted as **types**, **token** or **flemmas** (Pinchbeck, 2014 in Nation, 2016, p. 26). Table 2.1 provides an overview of different ways of counting words in relation to word lists based on Nation (2013, 2016). The two primary units of counting, word families and lemmas, are discussed in detail below. This section concludes with a justification of the choice of lemmas as the unit of counting for Study 1 and Study 2 of this thesis.

*Table 2.1. Overview of units of counting in word lists*

<b>Unit of counting</b>	<b>Definition</b>
Tokens/running words	The number of word forms in a text. The sentence "Mary gave the flowers to the maid." contains seven tokens.
Types	Unrepeated word forms in a text. The sentence "Mary gave the flowers to the maid." contains six types since "the" occurs twice and is counted as one type.
Lemmas	Word forms sharing the same stem and part of speech. A lemma consists of a base form and the inflected form of the base form. The lemma "maid" is a noun and contains the base form, "maid", and the inflected plural and genitive forms: "maids", "maid's" and "maids". "The head maid makes the work plan for the rest of the maids" contains nine lemmas: the, head, maid, make, work, plan, for, rest, of.
Flemmas	Word forms sharing the same stem but not necessarily the same part of speech. The flemma consists of a base form and the inflected forms pertaining to the different parts of speech of the flemma. In the sentence "The smile he gave her made her smile", there are six flemmas (but seven lemmas). The flemma "smile" contains the following forms: smile (noun + verb), smiles (noun + verb), smiled (verb), smiling (verb).
Word families	Word forms sharing the same stem and part of speech plus closely related derivations. A word family consists of a head word plus inflections and derivations of this head word. Derivations are included in varying degree according to Bauer and Nation's (1993) word family scale.

### *Word families*

The last type of unit of counting in Table 2.1 is the word family which is a framework (Bauer & Nation, 1993) developed to make the creation of reliable frequency-based word lists possible for use in vocabulary tests and in vocabulary load analysis programmes such as Range (Nation, Heatly, & Coxhead, 2002) or AntWordProfiler (Anthony, 2014) (Nation, 2016, p. 26). As an attempt to establish a morphological taxonomy in relation to what is useful for learners to know, the framework consists of a seven level scale as outlined in Table 2.2 which is based on Bauer and Nation (1993, pp. 258–262) and Nation (2016, p. 27). The scale moves from the most basic and transparent members of a word family (Level 1) to the least transparent family members in Level 7. Inclusion of word forms in each level was based on criteria of frequency, productivity, predictability, and regularity (Bauer & Nation, 1993). As described in Table 2.1, only word forms sharing the same stem are allowed to be grouped together in a word family. Additionally, only free morphemes can be a headword in a word family. As such, the adjective 'present' and the noun 'presence' are not part of the same word family but act as headwords of their own word families because the stem 'pres' cannot stand alone. Both '-'

ent’ and ‘-ence’ are suffixes (Nation, 2016, p. 29). It is important to note that the definition of a word family given in Table 2.1, consisting of a headword plus inflections and derivations of this head word, covers Levels 3 to 7 of the scale. The first two levels correspond to word types and lemmas, respectively. The scale is accumulative in that knowledge of a word family in e.g. Level 6 assumes knowledge of all the inflections and derivations of the preceding levels.

*Table 2.2. The seven levels of the word family framework (Bauer and Nation, 1993)*

<b>Level</b>	<b>Definition</b>	<b>Affixes</b>
Level 1	Each form is a different word (word forms)	‘agrees’ and ‘agreed’ are counted as different words
Level 2	Inflectional suffixes – regular inflections of the head word (lemma)	8 affixes: Plural, third person singular present tense, past tense, past participle, -ing, comparative, superlative, and possessive
Level 3	The most frequent and regular derivational affixes	10 affixes: -able, -er, -ish, -less, -ly, -ness, -th, -y, non-, and un-
Level 4	Frequent, orthographically regular affixes	11 affixes: -al, -action, -ess, -ful, -ism, -ist, -ity, -ize, -ment, -ous, and in-
Level 5	Regular, but infrequent affixes	50 affixes: e.g. -age, -al, -an, -hood, -let, anti-, inter-, and pro-
Level 6	Frequent but irregular affixes	12 affixes: -able, -ee, -ic, -ify, -ion, -ist, -ition, -ive(ative), -th, -y, pre-, and re-
Level 7	Classical roots and affixes	e.g. ab-, ad-, com-, de-, dis-, ex-, and sub-

The word family framework has been criticised for assuming a linear word knowledge development in learners, i.e. that the learner moves from Level 1 to 7 in a linear process of vocabulary acquisition (Gardner, 2007). In particular, learners are more likely to be exposed to the inflected and derived forms of a given head word of a word family, and thus having difficulties “recognizing and utilizing the common morphemic stems of a “*Word Family*” (Gardner, 2007, p. 248). This can be seen as an unfair criticism as the framework was not developed to represent learner knowledge development but, as Dang (2017) argues, outlines what is useful for learners to know. However, in relation to the use of word families in word lists, and in particular in relation to Coxhead’s Academic Word List (2000), Gardner and Davies (2014) argue against the use of word families because 1) the members of

a word family may not share the same meaning, which can have implications for lexical coverage analysis, and 2) derivational word knowledge is developed later than inflectional knowledge, which makes the word family too complex, especially for non-advanced learners (Gardner & Davies, 2014, pp. 307–308). Similarly, Dang and Webb (2016) argue against using the Level 6 word list family, which was used in word lists such as the General Service List (West, 1953) and the BNC/COCA2000 lists (Nation, 2006, 2012), in beginners' word lists since the Level 6 word family contains both highly frequent family members and very low-frequent members which makes it unsuitable for beginner learners of English. Dang and Webb (2016, p. 154) give the example of the head word 'study' which as a word family contains high frequency members such as 'studies' and 'studied', but also low frequency members such as 'studious' and 'studiously'. In contrast, the Level 2 word family, which corresponds to the lemma, only contains the forms 'studies', 'studied', and 'studying' besides the headword 'study'.

Another more practical toned argument against the word family is that using the word family as the unit of counting for word lists requires a great deal of manual checking and editing as Nation (2016, p. 30) reports because derivational affixes are more irregular than inflectional affixes which can easily be handled by computer-driven lemmatisation. Nonetheless, the word family, or more specifically word families of Levels 3 to 7, has been used as the unit of counting for word lists such as the General Service List (West, 1953), the Academic Word List (Coxhead, 2000), the Academic Spoken Word List (Dang, Coxhead, & Webb, 2017), and the BNC/COCA word lists (Nation, 2006, 2012). Especially word lists created for receptive purposes such as the Academic Word List make use of word families as it is assumed that learners are able to relate derivations to already known words when reading and listening even if they are not able to do so productively. It should be mentioned that for Danish, the concept of word families has not been developed and validated as it has in English, and it is not within the scope of this project to develop a framework for Danish such as Bauer and Nation's (1993) word family scale for English. In Study 4, however, an attempt is made to group Danish academic words together via morphological and semantical relatedness in connection with a process of supplementing a core Danish academic word list with related items.

### *Lemmas*

As outlined in Table 2.1, a lemma consists of a base form and the inflected forms of this base form. It corresponds to Level 2 in Bauer and Nation's word family framework (1993). In most corpus-linguistic vocabulary research, the lemma is the most used conceptualisation of the morphological relationship between a set of lexical items. Francis and Kučera (1982, p. 1) defined the lemma as "a



set of lexical forms having the same stem and belonging to the same major word class, differing only in inflection and/or spelling”. As such, ‘smile’ as a verb and as a noun constitutes two different lemmas. Also irregular forms are included in the lemma even though they do not have the same stem as the base form of the lemma (e.g. ‘go’ and ‘went’), and this non-transparent relationship between base forms and irregular forms may be problematic in a learning perspective as argued by Gardner (2007, p. 244). In general, however, it is assumed that learning a lemma (base form plus inflected forms) is easier than learning a word family (base form plus inflected forms plus derivations). When a learner has gained knowledge of the inflectional system of the language in question, it is believed that they can quickly learn the inflected forms of a base form (Nation, 2013, p. 10).

Another issue related to the lemma in relation to vocabulary learning is the semantic relationship between a lemma’s base form and its inflections as Francis and Kučera’s (1982) definition only takes the grammatical class into account. Similarly, Ruus in her investigation of core words of Danish merges homographs such as *koste*, which can mean either to hunt or to cost, into one lemma instead of listing them as two lemmas (Ruus, 1995, p. 21). Pedagogically, this may be problematic because of the difference in meaning (Gardner, 2007, p. 244). In their discussion of polysemy in Coxhead’s Academic Word List (2000), Gardner and Davies (2014, p. 308) draw attention to how a lemma-based approach can via “grammatical identification” solve the problem of discerning between ‘proceeds’ as a verb and as a noun. Homoforms as well as polysemous word forms certainly constitute a challenge in using the lemmas as a counting unit, but until we have fully semantically tagged corpora, the lemmas definition offered by Francis and Kučera (1982) is still the most wide-spread within corpus-linguistic vocabulary studies. Even though the word family is widely used in word list development in English applied vocabulary studies, there are also important examples of lemma-based lists in English as well as arguments for and against it. Gardner and Davies’s (2014) Academic Vocabulary List is lemma-based as is Brezina and Gablasova’s New General Service List (2015). The academic word lists developed for Swedish (Ribeck, Jansson, & Sköldberg, 2014) and Norwegian (Hagen, Johannessen, & Saidi, 2016) are also lemma-based. As will be outlined in the sections of general high frequency and academic word lists, Danish word lists such as Ruus’ (1995) *Danske Kerneord* (Danish core words) and the Danish Language and Literature Society’s list of the 10,000 most used words in Danish (Det Danske Sprog- og Litteraturselskab, n.d.-d) are also lemma-based. To my knowledge, no Danish word lists have been based on word families, which is contributable to the fact that this concept has not been applied in Danish as mentioned above. This,

in turn, may have to do with the fact that developments of word lists in Danish have been carried out primarily for non-pedagogical purposes.

### *The unit of counting in the four studies*

While the word family framework for English has its weaknesses as outlined above, it does offer a conceptualisation of a word that takes into account that words not only have inflectional affixes but also derivational affixes and knowledge of both are central to vocabulary use and learning. Given that this framework has not yet been explored in Danish, the unit of counting for the word lists developed in studies 1 and 2 is the lemma. The use of the lemma as the unit of counting in the Danish academic word list as well as in the general high frequency word list developed in Study 1 also has the advantage of being more suitable for non-advanced learners of Danish and for learners of Danish using the list for productive purposes. When Dang and Webb (2016) developed The Essential Word List, they argued against the Level 6 word family as described above. Instead, they chose to use the Level 2 word family corresponding to the lemma because beginners' morphological knowledge may not yet include knowledge of derivations. However, they decided to include different parts of speech into the same lemma or Level 2 word family thus making the unit of counting correspond to the unit of counting termed *flemma* in Table 2.1. Likewise, in both Study 1 and Study 2 of this thesis, the lemma is expanded to include different parts of speech. This has mainly to do with extraction measures and will be elaborated in Chapters 5 and 6. In the final list of Danish academic vocabulary, the DAWL, a small number of items with more than one frequently occurring part of speech are listed as separate lemmas in the list. For example, the item *styrke* occurs both as a verb (strengthen) and as a noun (force, strength) and is thus listed twice.

In the preceding sections, I have given an account of what a word is and how to count them. I have also justified the unit of counting for the studies of this thesis. In the next section, I outline what it means to know a word and present an overview of what the research literature tells us about vocabulary size and comprehension.

## 2.2. Word knowledge

Word knowledge is commonly envisioned as involving at least two dimensions: productive word knowledge and receptive word knowledge. **Productive word knowledge** is related to the skills of writing and speaking and involves knowing not only how a word is pronounced and spelled, but also where and when to use a given word, e.g. in a certain register or genre. In addition, knowing words productively includes knowing which other words a word can be used together with. **Receptive word**

**knowledge**, on the other hand, encompasses being able to connect spoken and written forms of a word to the word itself, e.g. understanding the meaning in a specific linguistic context. These different aspects involved in both productive and receptive word knowledge are categorised by Nation (2013) into three major aspects: **form**, **meaning**, and **use**. Using these three aspects, Nation develops a comprehensive framework for word knowledge in which he elaborates on the three aspects in relation to productive and receptive knowledge. This widely used framework is given in Table 2.3 adapted from Nation (2013, p. 49) .

*Table 2.3. What is involved in knowing a word (Nation, 2013)*

Form	spoken	R	What does the word sound like?
		P	How is the word pronounced?
	written	R	What does the word look like?
		P	How is the word written and spelled?
	word parts	R	What parts are recognisable in this word?
		P	What word parts are needed to express the meaning?
Meaning	form and meaning	R	What meaning does this word form signal?
		P	What word form can be used to express this meaning?
	concept and references	R	What is included in the concept?
		P	What items can the concept refer to?
	associations	R	What other words does this make us think of?
		P	What other words could we use instead of this one?
Use	grammatical functions	R	In what patterns does the word occur?
		P	In what patterns must we use this word?
	collocations	R	What words or types of words occur with this one?
		P	What words or types of words must we use with this one?
	constraints of use (register, frequency...)	R	Where, when, and how often would we expect to meet this
		P	word? Where, when, and how often can we use this word?

As can be seen from Table 2.3, the **form** aspect includes not only knowledge about orthography and pronunciation, receptively and productively, but also morphological knowledge, i.e. being able to recognise the parts of a word and use this knowledge for understanding the word. The **meaning** aspect, on the other hand, includes knowing how to relate the form of the word to the meaning of the word. Additionally, the meaning aspect involves knowing what the word refers to and what other words it associates with. The third and last aspect, the **use** aspect, relates to three issues: grammatical functions, collocations, and pragmatics, and this knowledge of context of use may be the most complicated aspect of word knowledge which often develops late for L2 learners.

### 2.2.2. How many words are needed to comprehend language?

Word knowledge and comprehension are closely related (Laufer, 1989; N. Schmitt, Jiang, & Grabe, 2011; van Zeeland & Schmitt, 2013). In the cited studies, comprehension is conceptualised as the number of words needed to be known to comprehend spoken and written language measured via text coverage. Laufer (1989) suggests that reasonable comprehension of a text can be reached at 95 percent coverage of a text. This means that 95 percent of the words of a text need to be known by the reader for them to understand it. To reach 95 percent, second language learners of English need to know 2,000–3,000 word families to comprehend spoken texts in English (Nation, 2006; van Zeeland & Schmitt, 2013). For written texts, learners need to know more vocabulary, i.e. 4,000–5,000 word families (Laufer & Ravenhorst-Kalovski, 2010; Nation, 2006) to reach the 95 percent coverage. In line with this, Dang and Webb (2014, p. 67) propose 95 percent for reasonable coverage and 98 percent for ideal comprehension of academic spoken discourse. The issue of vocabulary size and comprehension has also been explored in other languages than English, however to a limited degree. Hazenberg and Hulstijn (1996) investigated how many words were needed to understand first-year university reading materials in Dutch, and found that it requires up to 10,000 lemmas to read and understand such texts. Conversely, Cobb and Horst (2004) found that for French, knowledge of the 2,000 most frequently occurring lemmas in French were adequate for academic text comprehension. The question is whether this surprising difference in how many words are needed to comprehend Dutch versus French academic texts can be explained by language-related differences or by the applied methods of the two studies. Regardless of the cause, the difference emphasises the need for further studies on the vocabulary load of academic texts in other languages than English and its relation to reading comprehension. There is research carried out on the link between lexis and comprehension in Danish language, but these are primarily with a focus on young readers and less on readers at higher academic levels in the educational system, e.g. upper secondary school or university (e.g. Gellert, 2003; Nielsen, Daugaard, & Juul, 2017). Figures for English are therefore commonly referred to (e.g. Lund & Bertelsen, 2008b; Andersen & Henriksen, 2014). Study 1 of this thesis investigates the lexical coverage of the 2,000 most frequently used lemmas of Danish in academic and general texts, and is thus an important additional step in exploring the relationship between vocabulary size and reading comprehension by highlighting the nature and importance of high frequency vocabulary, e.g. in relation to lexical coverage.

## 2.3. Vocabulary categories

Having outlined the most central issues related to word knowledge in the preceding section, I now turn to how vocabulary can be classified into different categories.

First of all, words can be categorised according to their part of speech, a categorisation that forms the basis for the lemma-based way of counting words as described in Section 2.1.2. Even though some words have multiple parts of speech, this type of categorisation is rather clear and unproblematic. However, part of speech categorisation of vocabulary tells us only little about how the words are used in actual language besides their syntactical and morphological nature. Another way of classifying words that tell us more about how the words are used is the distinction between **content words** and **function words**. Content words, or lexical words as they are also referred to, are those words that carry some kind of content that is integral to the proposition set forth in a given text. Their presence in a text enables us to understand what the text is about. They comprise the four word classes, nouns, verbs, adjectives, and adverbs. Function words are equally important, but do not carry lexical meaning with them in the same way that content words do. Instead, they are used for structuring the content, i.e. for grammatical, syntactical and discourse purposes, and comprise the word classes of pronouns, articles, prepositions, conjunctions and auxiliary verbs. While words can be added to the group of content words as new concepts arise, function words are a closed set of words. However, a group of content words, especially nouns and verbs, is similar to function words in that the meanings of these words are dependent on the context in which they occur and these words have as such undergone a process of delexicalisation (cf. Meyer, 1990 in Nation, 2013, p. 300). For example, the meanings of words like *aspekt* (aspect), *del* (part), *problem* (problem), and *forhold* (forhold) are primarily available from the context in which they occur. This is an issue I return to in Study 3. The distinction between content and function words is interesting in relation to developing word lists and describing the inventory in such word lists, e.g. the development of a Danish academic word list.

In the next section, I focus on categories of specialised vocabulary as a prelude to the following two sections of which the first one introduces Nordic vocabulary categorisations related to education and research. The last section of this chapter provides an account of the concept of academic vocabulary and its functions. It is important to stress that the distinction between different types of vocabulary is related to different modes, registers and genres in a language, and thus for a systemic characterization of language use. The distinction is also crucial in relation to language development, both for L1 and L2 learners, e.g. in describing learning problems and developing language teaching tools in the form

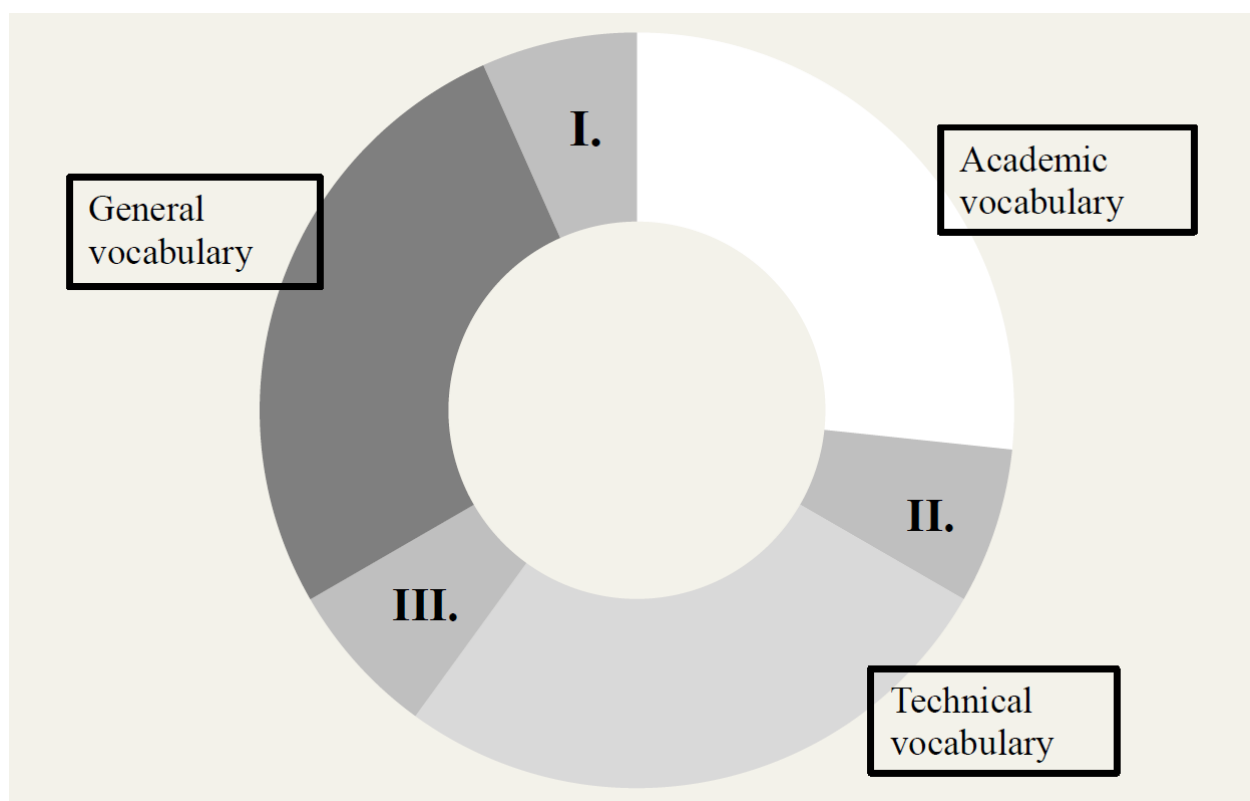
of word lists. Much of the research reported below has been motivated by concerns for language learning and teaching, and thus outline specifications of language use with this focus in mind.

### 2.3.1. Specialised vocabulary

Vocabulary can be divided into two broad functional categories: general and specialised vocabulary (Nation, 2013). Specialised vocabulary comprises both academic and technical vocabulary. The model presented in Figure 2.1 illustrates how the words of an academic text, whether written or spoken, can be divided into three macro vocabulary categories: general, academic and technical<sup>4</sup>. The category of general vocabulary will be addressed in Section 2.3.2 together with a discussion of high frequency vocabulary. The two other categories will be the focus of this section. As the name of the technical category suggests, words belonging to this category refer to technical entities and concepts, and the words are primarily known by those working and/or studying in the specific discipline or domain. These words are often referred to as domain specific vocabulary or specified even more precisely in relation to the domain they are used in, e.g. medical words, technical vocabulary of plumbing or gardening. Within the field of linguistics, terms like morphology, pronouns, and grammaticalisation are technical words. In contrast, academic vocabulary, as described in Chapter 1, are those words that are used across different disciplines and sub-disciplines to refer to general academic activities and academic functions. I will elaborate further on the definitions of academic vocabulary in Section 2.3.4.

---

<sup>4</sup> This model was first presented by Henriksen (2014) who calls it the vocabulary circle.



*Figure 2.1. The vocabulary circle – the vocabulary categories of academic language*

The areas with the numbers I., II., and III. in Figure 2.1. illustrate the overlaps between the three macro types of vocabulary. Area III. represents the fact that many general words can take on technical meanings which I will refer to as **pre-technical** words<sup>5</sup>. This term combines Fraser's (2003, 2006, 2008, 2009) two categories of crypto-technical and lay-technical words. Fraser uses the term 'crypto-technical' for those words which have additional technical meanings in specific disciplines besides their general meanings. Moreover, Fraser uses the term 'lay-technical' (2006) for general words with additional technical meanings which are supposedly known not only by specialist but also, as the name suggests, by lay people. Both sub-types are here termed pre-technical and can be described as belonging to overlap zone III in the model in Figure 2.1.

In relation to pre-technical words, Chung and Nation (2004) found that both West's (1953) General Service List and the Academic Word List (Coxhead, 2000) comprise a large number of words with specific technical senses. The issue of pre-technical words has received a lot of attention in the Nordic

<sup>5</sup> Fraser (Fraser, 2003, 2006, 2008, 2009)(2003, 2006, 2008, 2009) uses the term 'crypto-technical' (coined by Howard (1991)) for those words which have additional technical meanings in specific disciplines besides their general meanings. Moreover, Fraser uses the term 'lay-technical' (2006) for general words with additional technical meanings which are supposedly known not only by specialist but also, as the name suggests, by lay people. Both sub-types are here termed pre-technical and can be described as belonging to overlap zone III in the model.

context in relation to second language education, and will be expanded on in Section 2.3.3. Area II represents the overlap between technical and academic vocabulary. Academic words can have additional technical senses in some disciplines and some technical words may also function as academic words. This is an issue that I return to in Study 4 of this thesis. Area I. represents the fact that general words can also be academic, a notion that is explored in Study 1 of this thesis. It is also an issue that is of central concern in the creation and identification of academic word lists, and will consequently be an issue discussed in Chapters 3 and 5 as well. The vocabulary circle presented in Figure 2.1 is a model that illustrates the three macro types of vocabulary as well as the overlap between them. It does not, however, indicate the extent of those overlaps nor does it capture what can tentatively be termed **discipline-dependent polysemy**. This term is used to signal that general, academic, and technical words may take on additional senses dependent on the disciplinary contexts in which they occur.

The categorisation of words according to types is, as mentioned above, not just a characteristic of languages in general, but is also closely linked to vocabulary teaching and learning. While traditional terminology research has primarily been concerned with identifying the terms used in specific domains, dividing words according to the domains, registers, and genres for learning purposes have been the focus of much vocabulary research in both second and foreign language research. The distinction between general and specialised uses of vocabulary is commonly drawn in relation to learning purposes. In Nation's view, specialised vocabulary should be taught "when learners have mastered the 2,000-3,000 high frequency words of general usefulness in English (...)". At this point, it is advisable "to direct vocabulary learning to more specialised areas, depending on the aims of the learners." (Nation, 2013, p. 289).

### 2.3.2. Frequency of occurrence and general high frequency vocabulary

Viewing vocabulary in relation to its frequency of occurrence in a given language or a given text stands at the centre of most lexical research (N. Schmitt, 2010, p. 63). Frequency is a key aspect of vocabulary acquisition, processing, and use. Consequently, attention has been given to how vocabulary can also be classified according to their frequency of occurrence in the language or in specific texts. Frequency-based divisions of vocabulary divide words in a given language or a given text according to frequency of occurrence, and typically operate with three categories: high-frequency, mid-frequency, and low-frequency vocabulary. Words can be grouped into these three categories according to their frequency of occurrence in the language, using 1,000 word levels of frequency as shown in Table 2.4 based on Nation (2013, pp. 21–23).



Table 2.4. High, mid-, and low frequency vocabulary according to frequency levels

Categories	Frequency levels
High frequency vocabulary	1-2,000
Mid-frequency vocabulary	2,001-9,000
Low frequency vocabulary	>9,000

Table 2.4 provides a frequency division based on general language texts, and it is somewhat dependent on text type which words are the most frequent. For example, in a medical article, words such as *venøs* (venous), *symptomatisk* (symptomatic), and *profylaktisk* (prophylactic) may be high frequency words within that particular text while in general Danish they may belong to low frequency vocabulary. There is, however, a rather large set of words that occurs very frequently across all text types, and these are commonly referred to as **general high frequency vocabulary**. For example, the ten most frequently occurring words of Danish are *i*, *være*, *og*, *en*, *den*, *på*, *til*, *det*, *at*, and *af* (in, be, and, one/a, it, on, to, it, that, and of) (Det Danske Sprog- og Litteraturselskab, n.d.-d). These are words that we can expect to meet in all types of texts, irrespective of genre or domain. What also characterise the ten words exemplified here is that most of them belong to closed word classes such as prepositions and conjunctions, and are function words. Besides function words, general high frequency vocabulary also includes lexical words such as *sige* (say), *god* (god), and *stor* (big). These three words occur at the top of a list of the 10,000 most frequently used lemmas in Danish, and as with the first 10 words, these are words used in all text types. General high frequency vocabulary typically includes at least the 2,000 most frequent words in a language (Nation, 2013; N. Schmitt & Schmitt, 2014).

General high frequency vocabulary can make up a large proportion of any text, from 75% in academic written texts in English (Coxhead, 2000) to 90% of fiction texts (Hirsh & Nation, 1992). Knowledge of general high frequency vocabulary is therefore an essential goal for any language user and language learner (Nation, 2013, p. 24). Given that general high frequency words occur in all kinds of texts, language users, both L1 and L2 learners, will encounter these words many times in their listening and reading, and they will also be expected to be able to use these words when speaking and writing. Nation (2016, p. 5) uses the ‘cost/benefit principle’ in relation to high frequency vocabulary in ‘that learners should get the best return for their learning effort’. Learning high frequency words first, according to Nation (2016, p. 5), provides learners with “[...] the greatest opportunities to enrich their knowledge through later meetings with the words, and [...] the greatest opportunity to produce what they know”. The cut-off of 2,000 for high frequency vocabulary has been disputed as **mid-frequency vocabulary** to a high degree also include general purpose words (Nation, 2013, p. 18). Moreover,

knowledge of high and mid frequency vocabulary up to the fourth level (see Table 2.4) provides a coverage of about 95%, which is the coverage required for reasonable text comprehension as described above. In line with this, Schmitt and Schmitt (2014) argue for including the third 1,000 words into high frequency vocabulary as these are in particular useful for general purposes. The last frequency-based category of vocabulary, **low frequency vocabulary**, is in fact the largest group of words as it includes around 50,000 words in English (Nation, 2013, p. 23), and there is no reason to believe it is different for Danish. While a number of frequency-based studies on words exist in Danish (e.g. Bergenholtz, 1992; Ruus, 1995, see Chapter 3), no studies have been carried out to establish lexical coverage of different frequency bands in Danish. In Study 1 of this thesis, I explore the nature of Danish general high frequency vocabulary in relation to its coverage in both general and academic language use, and thus provide the needed research-based knowledge into this issue for Danish. In the next section, I will focus on Nordic conceptualisations of specialised vocabulary related to education and research.

### 2.3.3. Divisions of vocabulary in the Nordic languages<sup>6</sup>

The three vocabulary macro-types of general, academic, and technical words in the vocabulary circle in Figure 2.1 are also relevant when we turn to the research carried out in Sweden, Norway, and Denmark in relation to the vocabulary of education and research. These studies have primarily focused on primary and secondary school students' vocabulary knowledge with a focus on academic and technical vocabulary. This research has led to a threefold division of the vocabulary in academic and school-related texts into: 1) **topic-related words**, 2) **school-related but topic-neutral words**, and 3) **general words used in technical senses** (Enström, 2004; Golden, 2016; Lindberg & Johansson Kokkinakis, 2007). The first type is equal to technical or discipline-specific words as described above. The second type is similar to definitions of academic vocabulary in that they are used across a broad range of disciplines in ways that are supportive of the content proper but not central. The third type, general words used in technical senses, which corresponds to the third overlap zone in Figure 2.1, has, as mentioned, received a great deal of attention in Nordic second language research. Gimbel (1995, 1998) termed this type of vocabulary **pre-subject**<sup>7</sup> words (*førfaglige ord*). He was inspired by Golden and Hvenekilde (1983), Golden (1984), and Jørgensen (1984) to investigate both L1 and L2 primary school children's comprehension of the words in textbooks from different subjects. Jørgensen (1984) found that a number of words in primary school textbooks were

<sup>6</sup> Parts of Section 2.3.3 are adapted from Jakobsen (2017).

<sup>7</sup> This is Gimbel's own translation of *førfaglig* given in an English version of Gimbel (1995).

difficult to understand for children with Danish as a second language primarily because these words were assumed to be known to the children beforehand, and they had therefore not been explained. These words correspond to the words described by Fraser (2008) as lay-technical. Golden and Hvenekilde (1983) investigated the vocabulary of 40 primary school textbooks used in the teaching of history, geography, and physics with the purpose of developing materials to aid especially learners of Norwegian as a second language in their reading of textbooks. They based their study on the assumption that there were two types of vocabulary that would be problematic for L2 students. The first type comprised the technical words such as ‘negative electron’ and ‘republic’, and are words that are recognized as domain specific, and therefore warrant more attention in the teaching of a specific topic. The second type included general words with a higher frequency in disciplinary texts than in non-disciplinary texts such as ‘decrease’(Golden, 1984, p. 170). The latter type of words would be known by most L1 students and would therefore not be explained by the teacher. In contrast, most L2 students may not have had the opportunities to encounter these words outside school, and therefore they would struggle comprehending the text book. Based on frequency counts and teacher insight, Golden and Hvenekilde divided the words of the textbooks into three categories (Golden, 1984):

- 1) Known words
- 2) Technical words
- 3) Non-technical words

The first category consisted of words that the L2 students would surely know. The second category comprised words that occurred frequently in the textbook. Teachers of the three disciplines were asked to indicate which of these words they would explain to the students. Those words marked as such were categorised as technical by Golden and Hvenekilde. The third and last category, non-technical words, comprised the words that were not included in any of the other categories. Frequency counts showed that the non-technical words comprised more of the words in the subjects of geography and history than in the physics subject. Moreover, for the non-technical words, a surprising pattern emerged: Since these words were not technical words, they were expected to occur with a more even distribution across the three subjects. However, more than half of these words, 1,220 words out of 2,196 or 55 percent, were highly frequent in only one subject. This finding has important implications for L2 comprehension. As Lindberg (2007, p. 25) puts it, these words were not so general that it could be expected that L2 students automatically knew them, but they were central to the comprehension of the textbooks. In contrast to the technical words which could be expected to be explained by the teachers, these non-technical words were not explained to the students due to the assumption that they

would know them since the words are part of a general, everyday vocabulary. Thus, L2 students encounter technical words and unknown general words. They also, as Golden (1984, p. 175) concludes, encounter already known words in a different sense. The investigation carried out by Golden and Hvenekilde (1983) emphasised the aforementioned discipline-dependent polysemy illustrated by the overlap zones in Figure 2.1.

As mentioned above, general words that were used in context in a discipline-specific sense such as those identified by Golden and Hvenekilde (1983) were later termed pre-subject (*førfaglige*) words by Gimbel (1995, 1998) in his investigation of primary school children with Danish and Turkish backgrounds and their comprehension of the vocabulary of textbooks. To underline that these words belong in overlap zone III. of the vocabulary circle (Figure 2.1), I will henceforth use the term **pre-technical** (see Section 2.3.1). One motivation for Gimbel's study was that Jørgensen's (1984) study only included L2 children, and therefore Gimbel (1995, 1998) included 16 children with Danish and 16 children with Turkish background in his study. Words from textbooks from the subjects of history, geography, and biology were assessed by teachers who were asked to mark those words that they would explain in class. The remaining words (n=50) were presented individually to the 32 children in both spoken and written form. The children were asked to explain and use the words in context. Gimbel (1995) gives the following examples of these words: *afgrøde* (crop), *ansvar* (responsibility), *appetit* (appetite), *bevidstløs* (unconscious), *bønder* (farmers), *dyrke* (grow), *døgn* (day-24 hours), *energi* (energy), *fattig* (poor), and *flod* (river). The Danish children knew in average 42 out of 50 words while the Turkish children knew in average 15 out of 50 words. One of the interesting findings in the analysis of the children's explanations and use of the words was that the Turkish children relied more on the phonetic interpretation of the words than the Danish children which caused them to give wrong explanations and uses. Conversely, the Danish children employed their background knowledge and knowledge of other words in their answers, and were less dependent on the pronunciation of the words. Even though the investigation's population was small (n=32) as Gimbel acknowledged, Gimbel posited that the findings were in line with teacher experiences. Based on his finding that the children with Danish as their L2 had some difficulties explaining these words in comparison with L1 children, Gimbel called for more focus on vocabulary in the teaching of Danish as a second language (Gimbel, 1995, pp. 31–33). Moreover, his investigation supported the findings of Golden and Hvenekilde (1983) and, as mentioned, pre-technical words have since been the focus of much vocabulary teaching in Danish as a second language.

Lund and Bertelsen (2008a, 2008b), in their study on students with Danish as L2 in tertiary education (see also Chapter 1) employ the concept of grey-zone words (in Danish ‘gråzone’), based on both Golden and Hvenekilde’s (1983) and Golden’s (1984) findings, and on the work by Gimbel (1995, 1998). Grey-zone words are defined as used in a range of subjects, but with different senses depending on the subject they are used within (Lund, 2016, p. 85), i.e. combining pre-subject or pre-technical and academic words into the same category of grey-zone words. Content teachers will often use these grey-zone words, which they believe are well known to the students, to explain terminology. While language and words defined as grey-zone by Lund and Bertelsen (2008a, 2008b) and Lund (2016) arguably constitute a challenge for learners of Danish, the merging of what Nation (2013), among others, defines as two different types of vocabulary (technical vs academic, see also Figure 2.1) into a collective term, grey-zone words, risks making it difficult to operationalise such a term both linguistically and pedagogically. Moreover, pre-technical words have sometimes been likened to academic vocabulary (see H. P. Laursen, 2006, p. 39; Weber, 2009, pp. 52–53; K. Å. Laursen, 2013, p. 21). Such a merger of word types entirely misses the point that pre-technical and academic words are in fact two different types of vocabulary. As described above, pre-technical words are general words used as technical words to convey content matter. In contrast, academic words are primarily topic-neutral words used for organising the academic discourse, but also for the linguistic realisation of academic activities, processes, and tools (see the section on academic vocabulary below). Due to the fact that some words have multiple senses, i.e. are polysemous, some academic words may also occur as technical words as pointed out by Hyland and Tse (2007) and discussed in Section 2.3.4 below. These overlapping zones of word meaning and use, as illustrated in the vocabulary circle in Figure 2.1, between on the one hand topic-neutral academic words and topic-specific words (marked as II on the figure), and on the other hand between general and pre-technical words (marked as III on the figure) become even more obvious when turning to research on textbook vocabulary carried out in the Swedish context.

Also, in the Swedish context, researchers have focused on how to aid both L1 and L2 students in acquiring the necessary vocabulary knowledge in order to comprehend their written school materials. Below, I will describe three Swedish studies which more or less make use of the threefold vocabulary division between general, academic and technical vocabulary outlined in the beginning of this section: The OrdiL project (Lindberg & Johansson Kokkinakis, 2007), the T-Master project (Kanebrant et al., 2015), and a project on “Spåk- och kunskapsutveckling i NO-ämnena” (development of language and content skills in natural science subjects) (Johansson, 2017). Similarly, to the other studies discussed

in this section, the purpose of the OrdiL project was to aid students in mastering an adequate school-related vocabulary, but also to make teachers especially aware of the topic-neutral vocabulary, i.e. academic vocabulary. Similarly, the project “Spåk- och kunskapsutveckling i NO-ämnena” focused on language, reading and writing in Natural Science subjects in primary school and views vocabulary knowledge as a central component in the student literacy in Natural Science (Johansson, 2017). Kanebrant et al. (2015) aimed at developing a comprehensive assessment tool for reading skills in students aged 10 to 15. The two latter projects built largely on the vocabulary division set forth in the OrdiL project (Lindberg & Johansson Kokkinakis, 2007). In this project, the researchers found it useful to distinguish between topic-neutral and topic-related words. These two categories or types were further subdivided. The categories are summarised in Table 2.5. Järborg (2007, p. 87) states that category B to some degree is equivalent to academic vocabulary as defined in English vocabulary research. Category C, on the other hand, corresponds to what Gimbel (1995, 1998) termed pre-technical words and the 1,220 words identified by Golden and Hvenekilde (1983) as general but also discipline-specific.

*Table 2.5. Categories of textbook vocabulary (Järborg, 2007)*

Topic-neutral words		Topic-related words	
A	General words with frequent occurrences in both spoken and written language	C	General words with a technical meaning
B	General often abstract words used primarily in writing	D	Technical words

This categorisation of textbook and teaching material vocabulary was used and developed further in Kanebrant et al. (2015), and this categorisation was also used by Johansson (2017). The two overall categories of topic-neutral and topic-related words are maintained, but the sub-categories are more detailed as can be seen in Table 2.6 which is adapted and modified slightly from Kanebrant et al., (2015, p. 222). The modifications I have made relate in particular to the specifications added in capital letters in order to relate them to the vocabulary circle in Figure 2.1.

Table 2.6. *Categories of textbook vocabulary (Kanebrant, 2015)*

Topic-neutral words		Topic-related words	
1.	Most frequent words: The most common words which could appear in any text. (GENERAL VOCABULARY)	4.	Every day words (homonyms): Words which have a common every day meaning but also a subject or domain specific meaning. (PRE-TECHNICAL VOCABULARY)
2.	Middle-less frequent words: Less frequent words occurring in age-adapted texts. (GENERAL VOCABULARY)	5.	Subject typical: Words common to one type of texts, e.g. Natural Science texts. (TECHNICAL VOCABULARY)
3.	Genre typical words (academic, news, etc.): Academic words in school context, newspaper genre, descriptive texts. (ACADEMIC VOCABULARY)	6.	Subject specific: Words often only occurring in one type of text as unique words, e.g. a physics text on potential energy. (TECHNICAL VOCABULARY)

As can be seen, the researchers kept the C category from Järborg (2007) in Lindberg et al. (2007), but pointed out that these general or every day words were homonymous. Kanebrant et al. (2015) subdivided technical words into two sub-categories, subject typical and subject specific, based on the frequency of occurrence. The topic-neutral words also comprised three sub-types according to Kanebrant et al. (2015) based on frequency of occurrence with sub-category 1 corresponding to sub-category A in Table 2.5, and sub-category 3 corresponding to sub-category C. Thus, this latter sub-category corresponds to academic vocabulary as defined in the English vocabulary research literature.

In sum, topic-neutral vocabulary comprises general high frequency words as well as academic vocabulary. Topic-related vocabulary, on the other hand, comprises technical words as well as “vardagliga snedfördelade ord” which translates to general words that are unevenly distributed similar to the 1,220 words identified by Golden and Hvenekilde (1983) to be relatively discipline-specific (Golden, 2016, p. 7) and to the notion of pre-technical words as coined by Gimbel (1995, 1998). The Norwegian and Swedish research reviewed here proposes highly relevant vocabulary divisions based primarily on frequency of occurrence. To the best of my knowledge, no similar frequency-based research on the vocabulary of textbook has been carried out in the Danish context recently. The most recent frequency-based research on the vocabulary of textbook that I have found is Jansen (1973) which is a survey of the most common words in the subject of Danish in early primary school teaching. Based on especially Golden and Hvenekilde’s (1983), Golden (1984), and Gimbel’s (1995,

1998) studies, strong efforts have, however, been made in the Danish context to raise awareness among content teachers of the occurrence of pre-technical words in textbooks, and the comprehension difficulties these words entail for L2 students in particular (e.g. H. P. Laursen, 2006; Lund & Bertelsen, 2008a, 2008b; K. Å. Laursen, 2013; Henriksen, 2015; Lund, 2016).

As will be reported on in Chapter 3, both Swedish and Norwegian researchers have expanded the focus on vocabulary related to education by developing lists of academic vocabulary oriented towards higher education in particular. Such research is, however, also relevant for primary and secondary school education which is ascertained by the fact that academic vocabulary is included in the described vocabulary categorisations, and has been shown to create problems for both L1 and L2 learners. In the next section, I define academic vocabulary in more detail.

#### 2.3.4. Academic vocabulary

Academic vocabulary can occur across the different frequency levels of high, mid- and low frequency vocabulary (Nation, Coxhead, Chung, & Quero, 2016) outlined in Section 2.3.2. Academic vocabulary is commonly defined as words, both single-words and multi-word units, occurring with high frequencies across a broad range of academic disciplines. Due to this broad range, i.e. occurrence in multiple disciplines, academic words are often considered as context-independent or topic-neutral vocabulary “supportive of, but not central to the topics of the texts in which they occur.” (Coxhead, 2000, p. 214). This non-salient nature of academic vocabulary causes it to be a learning challenge for both L1 and L2 students across the educational system (Coxhead, 2000, p. 213; Nagy & Townsend, 2012; N. Schmitt et al., 2011). A strong motivation for exploring academic vocabulary<sup>8</sup>, and in particular for the development of academic word lists, has been to identify words relevant for L2 students irrespective of academic discipline. The concept has also received attention in relation to L1 students’ acquisition of academic language skills, especially in the North American context (see Nagy & Townsend, 2012; Ranney, 2012; Bailey et al., 2007). The high frequency nature of academic vocabulary is often considered relative to its occurrence in non-academic language. For example, Simpson-Vlach and Ellis (2010, p. 488) define academic formulas as being “significantly more

---

<sup>8</sup> It should be noted here that some researchers (e.g. Nagy & Townsend, 2012) use academic vocabulary as a cover term for two types of vocabulary occurring in academic language use: general academic vocabulary and discipline-specific academic vocabulary. The latter type encompasses lexical items occurring only within certain disciplines and is in this thesis termed discipline-specific or technical vocabulary. Conversely, the definition of general academic vocabulary covers what is here termed academic vocabulary. The discipline-specific academic vocabulary may be described as academic words that lie within overlap zone II in the vocabulary circle in Figure 2.1.



common in academic discourse than in non-academic discourse (...).” Defining academic vocabulary as more frequent in academic language than in non-academic language entails that also words belonging to general high frequency vocabulary can be considered academic. In fact, research has shown that academic vocabulary overlaps with general high frequency vocabulary. Comparisons of the Academic Word List (Coxhead, 2000) to Nation’s British National Corpus (BNC) lists have found that “a considerable number of the AWL word families” (Dang & Webb, 2014, p. 68) occur in the first 3,000 words of the BNC lists (e.g. Nation, 2004; N. Schmitt & Schmitt, 2014). Thus, even though the General Service List words were eliminated in the vocabulary selection for the Academic Word List, it still contains general high frequency words. The issue of the overlap between general high frequency vocabulary and academic vocabulary (as also illustrated in the vocabulary circle (Figure 2.1) is discussed in detail in Chapter 3 as it has implications for the development of academic word lists. It is also an issue with implications for language learning and teaching as it is central to the question of what is the nature of academic vocabulary, and I will return to the issue below.

The definition given above pertains primarily to occurrence and not so much to the form, meaning, and function of academic vocabulary. Nagy and Townsend (2012, pp. 93–95) offer some descriptive points in regard to the nature of academic vocabulary in English in their account of academic language, e.g. in relation to etymology and word class specifications. First of all, academic vocabulary comprises a high number of words with a Latin or Greek origin, and far from all of them have Germanic equivalents. This is supported by the fact that the Academic Word List contains 82 percent Graeco-Latin vocabulary (Coxhead, 2000, pp. 228–229). As Cobb and Horst (2004) argue, the prevalence of Latin and Greek words in English academic vocabulary is a distinctive feature of English academic vocabulary, and the question is if the same can be said of Danish academic vocabulary. There may be a tendency to use words of Graeco-Latin origin in Danish academic writing instead of the Germanic equivalents (e.g. *koncept* instead of *begreb*, *introduktion* instead of *indledning*), but this has not been investigated so far. If this is the case, it can also be contributed to the fact that much Danish academic writing, especially within the hard sciences, is influenced by English. A second feature of English academic vocabulary according to Nagy and Townsend (2012, p. 93) is that many academic words are morphologically complex due to affixation used for converting a word from one part of speech to another, e.g. verbs describing academic actions combined with nouns describing the process or result of this action (analyse and analysis). There is no reason to believe that this is not the case for Danish academic vocabulary since for example nominalisations are commonly found in Danish academic writing (cf. Rienecker & Jørgensen, 2012,

pp. 345–346). Extracting a Danish academic word list is the first necessary step in relation to describing Danish academic words as to origin and word class specifications.

Paquot (2010) offers some useful suggestions in regards to the nature of academic vocabulary by detailing the concept of sub-technical vocabulary. This term has often been used interchangeably with academic vocabulary and is described and discussed by Baker (1988). The term of sub-technical vocabulary was coined to meet the critique of the juxtaposition of general and specialised vocabulary in English for Specific Purposes. In particular, the critique was that this distinction did not take into account that for many language learners the problem is not so much understanding general or technical words, but more the words that “are neither highly technical and specific to a certain field of knowledge nor obviously general in the sense of being everyday words which are not used in a distinctive way in specialised texts.” (Baker, 1988, p. 91). Paquot (2010) seems to describe sub-technical vocabulary as a sub-type of academic vocabulary in that she defines academic vocabulary as comprising both sub-technical vocabulary as well as discourse-organising vocabulary. As such, sub-technical vocabulary, it seems, can be said to belong to overlap zone II., the overlap between academic and technical vocabulary in the vocabulary circle in Figure 2.1. Moreover, Paquot highlights the functions of academic vocabulary in arguing that it should be defined as “a set of options to refer to those activities that characterize academic work, organize scientific discourse and build the rhetoric of academic texts.” (Paquot, 2010, p. 28). Underlying Paquot’s functional definition is a focus on academic writing which differs from the more receptive focus found in e.g. Coxhead (2000), Dang, Coxhead, and Webb (2017), and Dang (2018a). Paquot claims that because the division between general and academic words is rooted in vocabulary research related to reading comprehension and text coverage, this division is more useful for receptive than for productive purposes, and she questions if for example all the words of the Academic Word List (Coxhead, 2000) should be taught productively (Paquot, 2010, pp. 15–16). Certainly, the issue of receptive-productive purposes is important in relation to academic word lists, however, the notion of a general academic vocabulary is important for both receptive and productive language use. In Section 2.4, I review additional research on the functions of academic vocabulary which is the focus of Study 3. In the following, I will briefly introduce some of the criticism that has been raised against the concept of academic vocabulary.

The most notable critique of the concept of academic vocabulary is expressed by Hyland and Tse (2007). Most importantly, they criticise the notion that students need to master a core academic vocabulary in order to comprehend academic texts as it strongly suggests that “there is a single

literacy which university students need to acquire to participate in academic environments.” (Hyland & Tse, 2007, p. 236). In sum, according to Hyland and Tse such an approach does not take into account that words behave collocationally and semantically differently dependent on contexts, and it does not represent how language is actually used in academic writing. Specifically, Hyland and Tse examined the words of Coxhead’s Academic Word List (2000) in relation to “frequency, range, preferred meanings and forms, and the collocational patterns of items in the AWL” (2007, p. 238) in order to explore its coverage in particular disciplines and how well the Academic Word List “represent the lexical composition of academic writing (...)” (2007, p. 238). Based on their findings, Hyland and Tse criticise the Academic Word List for not representing a vocabulary useful for students regardless of academic discipline in that many of the Academic Word List words are in fact discipline-specific even if they occur across disciplines. Put differently, they argue that the Academic Word List does not take discipline-dependent polysemy into account. As discussed by Durrant (2013, p. 3), this argument is closely connected to how vocabulary teaching and learning is envisioned. For example, Nation (2013, p. 77) argues that “[d]efining a word by looking for the general concept that runs through all its uses reduces the number of words to learn.” This suggests that the Academic Word List words should be taught according to their general meaning and not according to the possible multiple senses it may have. Hyland and Tse (2007) and Durrant (2013) seem to argue that semantic variation should be taken into account in relation to the teaching and learning of vocabulary related to education and research. This corresponds well with their position that only to a very limited degree is there such a thing as a general, core academic vocabulary. The idea of teaching a more discipline-specific academic vocabulary is also recognised by Gardner and Davies (2014, p. 311), but they maintain that to understand and determine what is discipline-specific vocabulary requires “a better understanding of what is common or core (...)”. Moreover, Gardner and Davies argue in line with Eldridge (2008) that there is in fact a core academic vocabulary relevant for English learners. Similarly, Malmström, Pecorari, and Shaw (2018, p. 37) assert that we should consider the context when defining words as academic: “Rather than asking whether there is such a thing as core academic vocabulary it might be more appropriate to ask: For any given word, in which set of circumstances is the word academic?”. This echoes the notion of discipline-dependent polysemy set forth in Section 2.3.1.

Another critique set forth by Hyland and Tse (2007, p. 247) is that word lists such as the Academic Word List by focusing on single-word units do not take into account the important role that

collocations play in academic discourse in expressing discipline-specific meanings (cf. Henriksen & Westbrook, 2017).

In this section, I have defined what academic vocabulary is and cemented it as one of the macro types of vocabulary included in the vocabulary circle presented in Figure 2.1. Academic vocabulary separates from general and technical vocabulary in that it is used for describing academic activities and language functions shared across disciplines which is why it occurs more in academic texts than in non-academic texts and is encountered across a broad range of disciplines. But as outlined above and as illustrated in figure 2.1., overlapping zones of polysemous meaning and use can be found between academic words and general words, and academic words and technical words. This overlap will be expanded on in Study 1 and Study 4 respectively. Apparent from this section is also that the concept of academic vocabulary is closely connected to the teaching of English as a second and foreign language. Also in the Nordic context, the teaching and learning of words in academic discourse that are neither specifically technical nor obviously general, to paraphrase Baker (1988), have received attention.

While technical words have a clear function of referring to concepts, entities, activities, and processes central to disciplinary content, the functions of the non-technical but still specialised inventory of academic lexis are less clearly defined. In the next section, I give an account of how previous research has addressed the functions of academic vocabulary.

## 2.4. Functions of academic vocabulary

Just as researchers have been preoccupied with defining and identifying academic vocabulary, the functions of this lexis have also received some attention. For example, studies of multi-word units in academic discourse typically classify these according to pragmatic-functional categories (e.g. Biber et al., 2004; Hyland, 2008; Chen & Baker, 2010; Simpson-Vlach & Ellis, 2010; Ädel & Erman, 2012). Similarly, Hirsh (2004, 2010) analysed the words of the Academic Word List (Coxhead, 2000) according to the functions they performed in written academic texts. In this section, I will describe different functional classifications of academic vocabulary, both taxonomies with broad categories, and the more fine-grained categorisation offered by Hirsh (2004, 2010). These categorisations will form the basis of functional descriptions of the Danish academic words as described in Study 3.

In Danish, Stray Jørgensen (2004, 2007) has proposed a framework of three word categories comprising the different words that university students should use when writing academic papers. The basis of this framework is the expectations connected to the academic texts and is based on his

experience with working with teaching academic writing. As Stray Jørgensen (2007) explains, the expectations of the academic text, be it a student paper or a research article, is that it 1) **investigates** a problem with the purpose of creating new knowledge, 2) that it **uses** the theories, concepts, and methods of the discipline to analyse, explain, interpret, assess the problem in question, and 3) that the academic text **documents** information, sources, the methodology, and argumentation in a precise and unambiguous way. Consequently, the first word category is **investigation words** and comprises words that correspond to what Stray Jørgensen terms academic language acts such as analyse, interpret, characterise, categorise, and assess. A primary function of the words in this category is to show “the academic writer as an active researcher in the text” (Stray Jørgensen, 2007, p. 164, my translation), and, therefore, verbs such as ‘choose’, ‘delineate’, and ‘conclude’, and their nominal counterparts are included in this category. The second category consists of what is called *vidensbrugeord* which can be translated to **use of knowledge words**. This category comprises words that refer to the concepts of the disciplines theories and methods and are as such often discipline-specific words, not academic words. However, the category also includes what Stray Jørgensen terms academic metacommunication words (Stray Jørgensen, 2004, p. 189) which are used to explain how theories and methods are used in the research and analyses. The third word category is **document words**, words used to show what is investigated and how. These words have defining and specifying functions as well as functioning as expressions of coherence, reasoning, and argumentation. Document words are also used to show who is saying what in terms of references. In the descriptions of these word categories, as given in Stray Jørgensen (2004, 2007) and in Rienecker and Stray Jørgensen’s (2012) book on academic writing for university students, several examples of words belonging to each category are given. No doubt, an empirically-based identification of the words of each category would substantiate Stray Jørgensen’s argument that these word categories constitute a system of concepts which can be used in the teaching of academic writing and by academic supervisors to give explicit linguistic guidance to students. The extraction of words for a Danish academic word List, the DAWL, described in Study 2 and the functional description of these academic words presented in Study 3 provide this empirical basis, and thus increase our knowledge of academic vocabulary in Danish academic writing.

In the selection of English academic words to teach international students, Martin (1976) proposes three categories of academic vocabulary which she defines as “high-frequency context independent words occurring across disciplines” and which are words that share “a focus on research, analysis, and evaluation – those activities which characterize academic work.” (Martin, 1976, p. 92). The three

categories of academic vocabulary consist of the words related to 1) **the research process**, 2) **analysis**, and 3) **evaluation**. The vocabulary of the research process corresponds to academic language functions (cf. Bailey et al., 2007) such as analyse, define, examine, interpret, combine, and categorise. They are used to describe the research process and what it is the researcher does in process of exploring an issue or carrying out an experiment. As such, this category of academic words is similar to the category of investigation words proposed by Stray Jørgensen. The second category, the vocabulary of analysis, comprises, according to Martin, high-frequency verbs and two-word verbs that are necessary in order to present and organise the academic discourse. To some degree, they correspond to Stray-Jørgensen’s category of document words. The last category is the vocabulary of evaluation which in Martin’s definition includes adverbs and adjectives that “often add an evaluative, subjective tone.” (Martin, 1976, p. 95). Examples of words in the three categories are given in Table 2.7.

*Table 2.7. Martin’s (1976) categorisation of academic vocabulary with examples*

<b>1. Vocabulary of the research process</b>
Formulate, analyse, categorise, investigate, study, examine
<b>2. Vocabulary of analysis</b>
Consist of, contain, comprise, base on
<b>3. Vocabulary of evaluation</b>
Comprehensive, pervasive, rigorous

In line with the categorisations suggested by Stray-Jørgensen (2004, 2007), the taxonomy proposed by Martin is primarily based on experience with teaching academic writing to students.

Another categorisation of academic words are given in Nation (2013, pp. 300–301) who outlines a classification presented by Meyer (1990). Similarly to Martin (1976), Meyer’s classification operates with three categories, and the focus is in fact on all words in academic discourse. However, because it primarily centres on the delexicalised words of English used in academic texts (Nation, 2013, p. 300), Meyer’s classification in some ways resembles both Martin’s (1976) classification and also the classifications applied in functional analyses of academic multi-word units, an issue which I will return to below. The first category of Meyer’s is **vocabulary relating to the domain of the text and the linguistic acts performed in it**. Specifically, these are words that convey what the authors are doing in the text and what they ascribe to other authors. The latter function of this category is comparable to Stray-Jørgensen’s document words. The second category comprises **vocabulary**

**describing scientific activities**, and can be compared to Martin's vocabulary of the research process and to Stray Jørgensen's two categories of investigation words and knowledge use words. The third category includes technical words as it consists of **vocabulary referring to the subject matter of scientific activities**, but also general and academic words referring to tense, aspect, modality, etc., e.g. current, present, recent, ability, and likely. In addition, this category includes words conveying classification of states of affairs, e.g. change, development, process, structure, quality. Also, relations between states of affairs are expressed by words of this category.

As mentioned above, functional analyses have also been carried out in relation to multi-word units in academic discourse and I will briefly introduce those now. Typically, multi-word units in academic language in the form of lexical bundles (Biber et al., 2004; Chen & Baker, 2010; Ädel & Erman, 2012) or academic formulas (Simpson-Vlach & Ellis, 2010) are categorised into three overall categories of **referential** multi-word units, **text or discourse organising** multi-word units, and **stance** multi-word units each with a number of sub-categories. Table 2.8 provides the three categories including sub-categories as applied in the study of academic formulas by Simpson-Vlach and Ellis (2010) which is based on the pragmatic-functional taxonomy set forth by Biber, Conrad, and Cortes (2004). The sub-categories in bold were added to the taxonomy of Biber, Conrad, and Cortes (2004) by Simpson-Vlach and Ellis (2010).

*Table 2.8. A pragmatic-functional taxonomy of academic formulas (Simpson-Vlach & Ellis, 2010)*

Referential					
Specification of attributes	Identification and focus	Contrast and Comparison		Deictics and locatives	Vagueness markers
Stance					
Hedges	Epistemic stance	Obligation and directive	Ability and possibility	Evaluation	Intention/volition
Discourse organisation					
Metadiscourse and textual reference		Topic elaboration	Non-causal	Cause and effect	Discourse markers

The point of departure for these classifications or taxonomies are Halliday's (1976) metafunctional text model of **ideational**, **interpersonal**, and **textual** metafunctions. These metafunctions can be explained in relation to three **functional layers** of the academic text (Hirsh, 2004, p. 73). The first

functional layer of the academic text relates to the ideational metafunction as in this layer the writer expresses topic-specific ideas and concepts. Its correspondence with Halliday’s ideational, also termed experiential, metafunction is derived from how this metafunction is used for conveying the writer’s experiences with the content proper of the text. The second functional layer of the text is concerned with the relationship between writer and reader and corresponds to the interpersonal metafunction in that this metafunction centres on expressing the relations between writers and readers. The third layer of the academic text is related to how the text functions as a coherent text and corresponds to the textual metafunction of Halliday’s. The taxonomies for categorising academic formulas and lexical bundles are aligned with these three metafunctions in that the category of referential is ideational, whereas the discourse or text organising category is textual. The category of stance is interpersonal. Central to understanding Halliday’s three metafunctions is the contextual categories of **field**, **mode**, and **tenor** which relates to the concepts of **register** and **genre**. The register of a text is expressed through these three concepts in the following way: The topic of the text is represented through the field which in turn is expressed via the ideational metafunction. The mode is represented by the rhetorical and discourse-organising elements of the text, and thus relates to the textual metafunction. The tenor is represented by the relationship between writer and reader expressed through the interpersonal metafunction. The way these three concepts of field, mode, and tenor are expressed in the text via the metafunctions, constitutes the register of the text which in turn is determined by the genre which again is constructed by the social context of the text, e.g. the academic community in the case of research articles. Formulated more simply by Hirsh (2004, p. 77), “(...) genre affects register, which in turn affects language use.”

Metafunctions	Subfunctions
Textual	Metatextual, extratextual, intratextual
Ideational	Scholarly process, states of affairs, relations between entities
Interpersonal	Authoritative

Figure 2.2. Hirsh’s (2004) framework for functional analysis of academic words

In his functional analysis of the Academic Word List (Coxhead, 2000), Hirsh (2004) developed a functional analytical framework based on the functional layering of the academic texts. The framework is illustrated in Figure 2.2 which is adapted from Hirsh (2004, p. 96). As can be seen from Figure 2.2, Halliday’s three metafunctions of textual, ideational, and interpersonal form three overall



categories each with a number of sub-categories. This framework for a functional analysis of academic lexis is applied in Study 3 of this thesis and will be explained further in Chapter 7.

## 2.5. Summary

In this chapter, I have outlined the concept of a word with a focus on how to count words (unit of counting). Moreover, I have given an account of the most central elements of word knowledge, and of the linkage between vocabulary knowledge and size and comprehension. The majority of the chapter has centered on categorisations of vocabulary particularly in academic texts. The concept of academic vocabulary has been delineated, and research on vocabulary related to education in the Nordic context has been reported. The usage-based division of words in academic language into general, academic and technical has been complemented by the divisions established by especially Swedish researchers into topic-neutral (general and academic words) and topic-related vocabulary (technical and pre-technical words). Based on especially the Nordic research reviewed, it is clear that words in academic texts are to a high degree polysemous, and the category of pre-technical words has proven to be especially challenging for L2 students (overlap zone III in the vocabulary circle). In contrast to technical and pre-technical words, which are essential for the comprehension of the topic of the text, academic words are defined as overall topic-neutral and context-independent, even though these can also be polysemous and take on technical senses, as shown by Hyland and Tse (2007), and as illustrated by overlap zone II in the vocabulary circle. Finally, I have introduced different pragmatic-functional frameworks for analysing the functions of academic vocabulary. Study 3 of this thesis provides a functional analysis of the words identified as academic in Study 2, thus increasing our knowledge of the functions of Danish academic words, and thus pointing at a categorisation that may be useful for pedagogical uses of a Danish academic word list (the DAWL). In the next chapter of this thesis, Chapter 3, general and academic vocabularies are revisited in relation to development and use of word lists for research and language teaching.

## Chapter 3. Word lists

### 3.1. Introduction

A considerable portion of the previous chapter was spent on how words are categorised into general and specialised uses. In this chapter, I address a number of issues in relation to word lists as this concept plays an important role in especially Study 1, 2 and 4 of this thesis. First, I introduce what a word list is and what they are used for (Section 3.2). Then, I give an account of general vocabulary word lists developed in the Danish context (Section 3.3). The second half of the chapter (Section 3.4) focuses on pedagogical word lists by first describing some important principles and methods used in the development of pedagogical word lists and then by detailing seven studies on academic word lists. Even though the primary aims of the studies of this thesis are to provide linguistic description of a lexical inventory only explored to a limited degree in Danish, the majority of this chapter focuses on the use and development of pedagogical word lists. The motivation for that is that it is primarily this type of word lists that has inspired the studies of this thesis and in particular the methodologies of Study 1 and Study 2.

### 3.2. Word lists in general

Lists of words are developed for research purposes as well as for pedagogical purposes. The advent of large, electronically stored corpora that can be used to investigate how words behave frequency-wise and in terms of co-occurrence with other words in different language types has increased our understanding of these issues. In addition, it has also enabled applied linguists to develop pedagogical word lists of both general and specialised language based on objective criteria. Such lists are used in second and foreign language teaching and especially in the teaching of English as a second language. Frequency-based word lists which are extracted in corpora representing different text types can be used to show how words, both single-word and multi-word units, behave frequency-wise in these text types. Research on terminology also employs corpus-based methods for exploring and describing terminologies related to specialised fields of knowledge in order to develop lists of terms, among other things.

Scott and Tribble (2006, pp. 11–12) discern between two principles behind the creation of word lists: *Transformation*, which involves taking a number of texts and ordering all lexical items of these texts into an alphabetised or frequency-ordered list, and *selection*, which involves as the term signals, selecting a certain set of lexical items. Word lists developed for pedagogical use, such as academic word lists, but also dictionaries, are developed based on the principle of selection. The Danish

Language and Literature Society's list of the 10,000 most used words in Danish (Det Danske Sprog- og Litteraturselskab, n.d.-d) can be said to be an example of a word list developed on the basis of the principle of transformation even though the list does not contain all the lemmas of the underlying corpus. In Section 3.3, I describe this word list and two other Danish general word lists<sup>9</sup> in detail. All three lists are examples of word lists primarily developed with the aim of exploring the nature of a certain type of vocabulary in Danish.

It should be noted that the distinction set forth above between word lists created for pedagogical vs. research purposes is a somewhat artificial one set up here for the sake of clarity. In fact, pedagogical word lists, such as the BNC/COCA lists (Nation, 2006, 2012) and the General Service List (West, 1953), are also used for research purposes such as investigating the lexical profiles of different text types (e.g. Dang & Webb, 2014; Bardel & Lindqvist, 2011; Cobb & Horst, 2004) and as a basis for describing a lexical inventory. For example, Hirsh (2004, 2010) used the Academic Word List (Coxhead, 2000) for exploring the functions of academic vocabulary in academic writing. In Study 1 of this thesis, a general high frequency word list is used to explore the nature of academic Danish language use in comparison with general Danish language use.

### 3.3. Danish word lists of general high frequency vocabulary

In this section, I review three investigations of Danish vocabulary based on primarily frequency-based criteria.

#### 3.3.1. Most frequently used lemmas in Danish

The most recent research on Danish general high frequency vocabulary is the lemma-based frequency list developed by the Danish Language and Literature Society called the "Most frequently used lemmas in Danish" (Det Danske Sprog- og Litteraturselskab, n.d.-d). This list comprises 10,000 lemmas currently<sup>10</sup>. The DSL list is derived from a corpus of 880 million running words comprising 80 percent written texts and 20 percent spoken texts from 1983 until 2016. As can be surmised from the title of the list, the unit of counting is the lemma, and the definition of a lemma as a headword and its inflections of the same part of speech as given by Francis and Kučera (1982) is followed. Thus, the item *om* (about) occurs twice on the list because it can both be an adverb and a preposition. It should be noted that the DSL list only contains headwords of the lemmas and not inflections. In total, the list ranks 10,000 lemmas according to relative frequency (number of occurrences of all forms of

---

<sup>9</sup> I view the results of all three investigations as word lists even though Bergenholtz (1992) is called a frequency dictionary.

<sup>10</sup> When Study 1 was carried out, the DSL list comprised 5,000 lemmas.

the lemma divided by the size of the corpus) which is also listed together with part of speech information for each lemma as can be seen in Table 3.1 which depicts the first 10 lemmas of the list.

*Table 3.1. The first ten lemmas in the DSL list*

<b>Part of speech</b>	<b>Lemma</b>	<b>Relative frequency</b>
T	<i>i</i> (in)	0.032249628510297
V	<i>være</i> (be)	0.0309023882233708
C	<i>og</i> (and)	0.029584070147617
P	<i>en</i> (a)	0.0253413695013101
P	<i>den</i> (it)	0.0248728148892572
T	<i>på</i> (on)	0.015317332743123
T	<i>til</i> (to)	0.0152345449047462
P	<i>det</i> (it)	0.0147142978135353
U	<i>at</i> (that)	0.0144963754376622
T	<i>af</i> (of)	0.014170235977948

The fact that the DSL list contains 10,000 lemmas means that it does not only represent general high frequency words but also mid- and low frequency words similarly to the BNC/COCA lists developed by Nation (2006, 2012) and based on the British National Corpus (BNC) (BNC Consortium, 2007) and the Corpus of Contemporary American English (COCA) (Davies, 2008). The BNC/COCA lists<sup>11</sup> represent 28 frequency levels of English vocabulary ranging from high to low frequency vocabulary (Nation, 2016, p. 132). If we compare the DSL list to the BNC/COCA lists, there are some obvious limitations of the DSL list. Most importantly, the DSL list is a pure frequency list created, as mentioned, on the basis of transformation in that the lemmas are ranked according to their total relative frequency in the corpus. The BNC/COCA lists, on the other hand, were developed using criteria of both frequency and range, and for the high frequency words also subjective judgement (Nation, 2016, p. 133). Moreover, two corpora representing different varieties of both written and spoken English were used in the creation of the BNC/COCA lists. The DSL list is based on only one, albeit rather large, corpus of primarily written material and created using only one criterion, frequency, and it can be questioned how representative the DSL list is of Danish general high frequency vocabulary. However, given that it is the most recent derived list of Danish general

<sup>11</sup> The BNC/COCA comprises 28 lists of each a 1,000 word families plus six lists of nonsense words, proper nouns, marginal words, transparent compounds, and acronyms (Nation, 2016, p. 132).

vocabulary and based on a rather large corpus also used for developing and updating the Danish Dictionary, the first 2,000 lemmas of the list are used in Study 1 of this thesis as a representation of Danish general high frequency vocabulary.

### 3.3.2. Danske kerneord (Danish core words)

Another study of Danish general high frequency vocabulary was carried out in the early 1990's by Ruus (1995) and a replication study of this is underway by the Danish Language Council (Dansk Sprognævn, n.d.). Ruus' study of what she terms Danish core words (*Danske Kerneord*) is an example of a word list created on the basis of the principle of selection as described above and was carried out using principled criteria for the identification of core words. The foundation for the study was previous frequency-based word form lists from the 1980's, extracted from a corpus of 1,250,000 million running words comprising five text types: newspaper texts, professional journals, magazines, novels, and children's books. In Ruus' study, the word form lists were converted into a dictionary of lemmas through a comprehensive lemmatisation procedure using principled and reasoned decisions. One of these decisions was to list orthographic double forms (for example, the word 'look' can in Danish be spelled *kigge* and *kikke*), and homographs (words spelled the same but pronounced differently) as one lemma (Ruus, 1995, p. 21). Conversely, homonyms with more than one part of speech were listed as separate lemmas as in the DSL list. Thus, as in the DSL list, the lemma definition given by Francis and Kučera (1982) was used. Ruus (1995) applied two criteria for extracting the core words from this lemma dictionary. The first one was that for a lemma to qualify as a core word it had to have "great theoretical frequency" (Ruus, 1995, p. 39). This means that a lemma had to be much more frequent than another lemma occurring very frequently in a text type. Much more frequent in this context means that a lemma had to occur with a frequency of at least 20 in one or more of the involved text types. The cut-off of 20 was reached by a standard deviation measure that assumes that "[a] lemma is more frequent than another, if it is possible to retract two standard deviations (i.e. two times the square root of the number) from the first lemma's frequency count without reaching the second lemma's frequency count." (Ruus, 1995, p. 38, my translation).

The second criterion was that a lemma had to occur in the word form list with the fewest marked (*påfaldende*) words. Occurrence in this list should ensure that the lemma could be found in all text types using the first criterion (Ruus, 1995, pp. 38–39). Here 'marked' means that a word has a very limited range. In other words, it only occurs in one of the five word form lists. The word list of magazines was gauged the word list with the fewest striking words. The last step of identifying the core words of Danish was to control the list of lemmas extracted using the two criteria described

above. Specifically, the extracted lemmas had to occur more than 19 times in the magazine text type to be included in the final core word list (Ruus, 1995, pp. 39–40). For this control, Ruus used concordance data to ascertain that all occurrences of a lemma did indeed pertain to the assumed part of speech of the lemma. As an example, the lexical items *august/August* occurred 28 times in the magazine corpus, but only 17 times as a noun (the remaining 11 occurrences were the proper name *August*). Therefore, the lemma *august* was not included as a core word. The final pool of identified core words consisted of 1,117 lemmas which were analysed further in relation to antonymy, hyponymy, and meronymy. Ruus argues that these lemmas are part of the Danish lexical norm necessary for all users of Danish to know in order to produce and understand Danish (Ruus, 1995, p. 192). Table 3.2 shows how a core lemma is listed in the published book reporting on the investigation and is adapted from Ruus (1995, p. 59). The core lemmas are listed according to part of speech with the part of speech with the most lemmas at the top of the list. The # sign after the frequency count for *dør* signals that this is a polylemmatic form meaning that it can be found in more than one lemma. The Danish word for ‘die’ is *dø* and the present tense is *dør*.

Table 3.2. Excerpt from *Danske Kerneord* (Ruus, 1995)

døgn ( <i>sb</i> ) (24-hours)	
døgn	10
døgnet	12
dør ( <i>sb</i> ) (door)	
dør#	#14
døre	9
døren	36
efterår ( <i>sb</i> ) (autumn)	
efterår	4
efteråret	20
eksempel ( <i>sb</i> ) (example)	
eksempel	41
eks.	38
eksempler	4

### 3.3.3. Dansk frekvensordbog (Danish frequency dictionary)

The last study I will review here, does, in line with the DSL word list, not only involve general high frequency vocabulary, but is a frequency dictionary created by Bergenholtz and published in 1992. The dictionary does, however, contain a list of the 5,000 most frequent Danish words besides its approximately 183,000 entries which are listed alphabetically. The corpus on which the dictionary is based is a corpus of 4 million words consisting of four text types: fiction (50%), newspapers (25%), and weekly magazines (25%). The corpus is thus a written corpus similar to the corpora used for Ruus' Kerneord and the DSL list. The word form or what the author refers to as 'orthographic words' is used as the unit of counting and for each word form, the frequency in each of the four text types is given (see Table 3.3). The use of the word form partly explains the high number of entries (182,860 word forms). The other explanation is that all word forms are included even if they only occur once which may make the dictionary difficult to navigate in (Ruus, 1992, p. 155). Bergenholtz states in the introduction to the dictionary that it can be used by people working professionally with Danish from researchers to teachers, but it is apparent from the same introduction that the purpose of the dictionary is not pedagogical but to provide descriptive information about frequency, for example in relation to adjusting the orthographical norm (p. xvi-xvii). Table 3.3 shows an excerpt from Bergenholtz's Danish Frequency Dictionary (1992, p. 517). Frequency counts are given for the entire corpus (87-90) and for the three text types (newspapers, magazines, and books).

Table 3.3. Extract from *Dansk frekvensordbog* (Bergenholtz, 1992)

	87-90	Aviser (newspapers)	Blade (magazines)	Bøger (books)
HVÆS (hiss)	3	-	-	3
hvæs	6	-	2	4
hvæse (hiss)	3	-	-	3
hvæsede	33	1	3	29
hvæsen	5	1	-	4
hvæsende	11	1	-	10
hvæser	19	-	4	15
hvæssede (whet)	2	-	2	-
hvæsser	3	-	1	2
hvæsset	2	-	-	2
hvæste	1	-	-	1
H.W. (H.W.)	2	2	2	-
Hyacint (hyacinth)	4	-	-	4
hyacint	1	-	-	1

Having reviewed three Danish studies concerning general high frequency vocabulary which can also be categorised as word lists developed primarily for linguistic description, I will now turn to word lists that are created for primarily pedagogical purposes and how these are developed. As outlined below, many of these lists may also be used for research purposes and give us a general understanding of the vocabulary profile in a given language, e.g. English. I also discuss the pedagogical relevance of the three Danish word lists discussed above in the following section.

### 3.4. Pedagogical word lists

Word lists are widely used pedagogically, for example in the teaching of English as a second or foreign language. As Nation (2016, p. i) argues, word lists are central to vocabulary teaching and testing and to course and material design. Pedagogical word lists are created to represent the different types of vocabulary that the learners need to know in different contexts. For example, Coxhead's Academic Word List (AWL) (2000) was developed to aid learners of English at university level in their reading of academic texts. It has, however, also been used in primary and secondary school



contexts especially in the U.S. (cf. Nagy & Townsend, 2012). Most pedagogical lists in English are developed for receptive purposes, i.e. for reading and comprehending language, but pedagogical word lists can also be developed for productive purposes such as Dang, Coxhead and Webb's (2017) Academic Spoken Word List (ASWL). Moreover, word lists such as Coxhead's Academic Word List and West's General Service List (GSL) (1953) have also been used for vocabulary testing. For example, the Vocabulary Levels Test (Nation, 1983, 1990; N. Schmitt, Schmitt, & Clapham, 2001) measures how many words the learner knows from four different frequency levels using West's General Service List (1953) to select words representing the first two frequency levels.

In the Danish context, *Dansk frekvensordbog* (Danish frequency dictionary) (Bergenholtz, 1992) has been used for developing a Danish version of the Vocabulary Levels Test (Albrechtsen, Haastrup, & Henriksen, 2008) as mentioned above. In contrast to the research literature related to the issue of word lists in English language teaching and learning, the literature on the use and development of word lists for Danish as a second and foreign language is limited. Word lists of different kinds are used in the teaching of Danish, but these are primarily teacher-created and developed for very specific contexts, i.e. talking about the weather or words related to the body. If we consider the pedagogical use of e.g. the BNC/COCA word lists described above or the General Service List (West, 1953), the question is if the three Danish investigations discussed in the previous section can be used as pedagogical word lists of general high frequency vocabulary. Firstly, while the corpora they are based on are created to cover a range of text types, they primarily represent written language and as such they provide little or no information on which words are highly frequent across both spoken and written genres. This may not be a fair criticism with regards to the two last investigations considering they were carried out in the early 1990's at a point when spoken corpora were not widespread. Secondly, it is questionable if any of the lists are organised in a way that allows for direct use in the classroom. The DSL list does provide a qualified overview of the most frequent words as it is a frequency-ranked list and one can easily isolate e.g. the first 100 words for teaching purposes. Also, the frequency dictionary provides a frequency ranked list of the most frequent 5,000 word forms which can also be lifted into teaching materials. As such, both of these lists could be used for testing purposes (cf. Albrechtsen et al., 2008). One could argue, however, given the age of Bergenholtz (1992) frequency dictionary and of Ruus' (1995) investigation, that they are not representative of the most frequent words in modern Danish. Compared to the General Service List, which is still widely used, these lists, however, are quite new and it is doubtful how much the general high frequency vocabulary of Danish has changed over the last 30 years. The fact is, however, that none of these lists,

to my knowledge, have been applied to pedagogical purposes. While all three studies on Danish general high frequency vocabulary advance our knowledge of this lexical inventory, a step towards finding out more about vocabulary knowledge in Danish is to develop word lists of general language that can be used in testing, language learning, and materials development. In the next section, I look more closely at how pedagogical word lists are developed with a focus on academic word lists. After I have outlined some central issues in relation to word list development, I report on seven different studies on academic word lists.

### 3.4.1. Developing pedagogical word lists

According to Richards “the particular circumstances under which the language is to be taught and used” (1974, p. 70) should guide word selection for a pedagogical list. This means that developers of word lists should take into account what kind of vocabulary the learner needs to know and for what purpose. In particular, the developers for example need to consider who the word lists are developed for: L1 or L2 learners, for children or adults, for general language purposes or related to specific language use. So basically, word list development should ideally be based on a needs analysis for the specific user group in question. Closely related to these concerns, is the issue of unit of counting in word list development as described in Chapter 2. For example, word families are seen as a suitable unit of counting for word lists to be used for receptive purposes. In contrast, lemmas are commonly seen as appropriate for non-advanced learners as acquisition of lemmas only requires inflectional knowledge and not derivational knowledge as word families do. As with word lists developed for research purposes, the type and composition of the corpus used is of central concern. Using a corpus of formal written language for the creation of a word list for young children is not advisable. Similarly, spoken academic word lists are developed since the academic word lists developed on the basis of corpora of written academic language may not be suitable in aiding the comprehension of spoken academic discourse. Finally, equally important in relation to pedagogical word lists is the criteria used for word selection, and in the section below I will focus on what these involve.

#### 3.4.1.1. *Methods and criteria used in word list development*

The clear-cut distinction between transformation and selection in word list creation given above in Section 3.2 may not hold entirely when we look at development of pedagogically oriented general high frequency lists such as the General Service List (West, 1953) and the BNC/COCA lists (Nation, 2006, 2012). In the development of such lists, a first step has been to rank all the lexical items of a corpus according to frequency and then carry out a number of selection measures to make the list pedagogically useful. These selection measures may include other objective, quantitative measures

such as range and dispersion but also subjective criteria such as ease of learning and necessity as well as semantic and functional criteria. The afore-mentioned General Service List (West, 1953) is an important example of a general high frequency word list developed using a range of both objective and subjective selection criteria. This list, commonly referred to as the first 2,000 words of English, has had a massive impact on vocabulary instruction in English as a second and foreign language. The use of subjective criteria along with objective criteria is, among other things, motivated by issues related to corpus composition. As pointed out by Nation (2004) in his study on the most frequent word families in the BNC Corpus, purely frequency-based word lists may exclude useful vocabulary for non-advanced learners as the corpus behind may be based on primarily written language used by adult speakers in more or less formal settings. Similarly, the first 2,000 lemmas of the DSL list described above reflect rather formal written Danish. For example, words like *diskussion* (discussion), *undersøgelse* (investigation), and *analyse* (analysis) occur within the first 2,000 lemmas probably due to the prominence of newspaper texts in the underlying corpus.

An example of a word list developed using both objective and subjective criteria in order to ensure a suitable basis vocabulary of relevance to beginners is the *Geirfa Graidd* (Morris, 2010; Morris & Meara, 2014) which is a word list of Welsh core vocabulary aimed at adult learners. Due to the lack of a general language corpus of Welsh, the developers used semantic fields, *centres of interest*, to form a corpus for the creation of a word list. These centres of interest were established by asking teachers of Welsh to provide a number of words for each centre. The provided words were ranked alphabetically and after frequency of the responses, i.e. words were ranked after how many times they occurred in the number of words provided by the teachers. In addition, 400 high frequency words from a lexical database of Welsh were added to ensure that basic function words were included. In total, the resulting list of these procedures contained 1,900 items. This list was compared to course materials and divided into two lists adjusted to the two beginner's levels of the Common European Framework of Reference ((CEFR) Council of Europe, n.d.), A1 and A2. The A1 list contains 616 items, and the A2 list 515. The procedures adopted by Morris were also used in the development of a pedagogical word list of English for Danish learners in the first grades of primary school (Croy, 2016). Vocabulary selection for word lists thus comprises both objective and subjective methods which can be used separately or together. Objective methods for creating word lists consist of extracting lexical items from one or more corpora using measures of frequency, range, dispersion, and in the case of using more than one corpus, distribution across corpora. Especially for word lists of general vocabulary, the use of more than one corpus is common. Such quantitative extraction

procedures have become increasingly easier and more feasible with the advances in corpus linguistics and computational linguistics during the last 30 years.

Johansson, Hagen and Johannesen (2017, pp. 150–151) list three general principles for creating academic word lists. The first principle concern the elimination of general high frequency words without academic senses. The second principle concerns the elimination of technical vocabulary, and the third principle concerns how to capture the words that occur in most disciplines and sub-disciplines, that is the academic words. These principles are realised through the use of the measures of frequency, range, and dispersion which I will introduce and discuss in the following two sections.

### *Frequency*

The objective measure of frequency ensures that the words selected are the most frequent words within the language type being studied. The fact that the frequency-based word selection for the General Service List (West, 1953) was carried out before the Second World War shows that even without computers frequency-based vocabulary selection has been an important approach for creating word lists (see Gilner, 2011 for a review of the history of West's General Service List (1953)). Frequency is also an important criterion in the creation of academic word lists since academic vocabulary is defined as being highly frequent in academic language use. There are essentially two ways of ensuring the selected words are indeed frequently occurring in for example academic language: using a frequency cut-off point or a comparative frequency measure. These two ways are related to how the relationship between general high frequency vocabulary and academic vocabulary is viewed. According to Dang, Coxhead and Webb (2017), this relationship can be viewed in relation to how academic vocabulary is defined:

According to the first approach, academic words are defined as items that fall outside general high-frequency words bands but that have a wide range and high frequency in academic texts. (...) The second approach considers academic vocabulary as a separate kind of vocabulary cutting across different 1,000 levels of general vocabulary (Gardner and Davies, 2014). (Dang et al., 2017, p. 6)

Both in Coxhead's (2000) The Academic Word List as well as in The Spoken Academic Word List (Dang, 2017; Dang et al., 2017), a frequency cut-off point was used as one of the criteria for selecting words for the word list. I will return to why Coxhead chose to exclude the General Service List words from her lists when I discuss this and other academic word lists in detail in Section 3.4.2. Conversely, other academic word list developers have argued for not excluding any form of high-frequency

vocabulary as there may be important academic words among them (Paquot, 2010, p. 45) (Gardner & Davies, 2014, p. 309). For vocabulary selection in the Academic Vocabulary List, Gardner and Davies (2014) applied a relative frequency ratio (Gries, 2010) measure to compare the frequencies of the words in an academic corpus and in a comparison corpus. This measure was also used by Hagen, Johannesen, and Saidi (2016) in their development of a Norwegian list of academic vocabulary. Using the relative frequency measures implies setting a cut-off value. Hagen et al. (2016) had as their cut-off that an item had to have a ratio value between 2.2 to 2.6 to be included. Gardner & Davies (2014), on the other hand, set the cut-off to 1.5 meaning that an item had to occur 50 percent more frequently in the academic corpus than in the non-academic corpus. Determining the cut-off value for this ratio measure involves experimenting with different values and looking at which words are left out at the different cut-off points. Paquot (2010) in her word selection for the Academic Keyword List (AKL) used keyness analysis (Scott, 1997) which is a statistical measure that ensures the extraction of so-called distinctive words. That is, words that clearly characterise a given text or a corpus. This measure was also used in the development of a Swedish academic word list (Ribeck et al., 2014). Specifically, the keyness method makes use of the log-likelihood ratio test to compare the frequencies of words across e.g. corpora and to determine if the difference in frequencies is statistically significant. The log-likelihood statistic has also been applied in the development of other academic word lists. Vlach-Simpson and Ellis (2010) used the measure to separate general language phrases from academic phrases in their Academic Formulas List. The strength of the log-likelihood ratio test in contrast to the relative frequency ratio measure is that it gives a strong indication of how distinctive the extracted words are to a given corpus. The relative frequency ratio measure only shows the differences in corpus frequencies, not what these differences say about the corpus in focus. For this reason, in Study 2, the log-likelihood ratio test was used for the frequency criterion in the word identification and selection for a Danish academic word list.

### *Range and dispersion*

Another important objective measure for developing word lists is that of range. This is commonly considered to take priority over frequency for word lists of academic vocabulary as a defining trait of academic words is that they occur across a wide range of disciplines (Coxhead, 2000). Consequently, in selecting words for an academic word list, the words need to occur in as many academic disciplines as possible. As will be shown in the section on academic word lists, there are different approaches to carrying out the range criterion when developing academic word lists, but the crucial issue is to ensure that the selected words occur in as many disciplines and sub-disciplines as possible. Overall, three

approaches emerge from the research literature. The first approach comprises what I will here term “simple occurrence”. This means that the researchers establish how many disciplines or sub-disciplines a word has to occur in to be selected for the word lists. As will be shown in Section 3.4.2 on different academic word lists, a simple occurrence range criterion could be, as in Dang, Coxhead, and Webb (2017), that a word has to occur in all academic disciplines and in half of the sub-disciplines. Another way of applying a range criterion is to use expected frequency as in both Gardner and Davies (2014) and in Hagen et al. (2016). Expected frequency is calculated in relation to the size of a sub-corpus. If a word occurs 500 times in a corpus of 500,000 words, its relative frequency is 0.001. The expected frequency of this word in a sub-corpus of e.g. 50,000 words is  $0.001 * 50,000 = 50$ . An expected frequency range criterion is set so that a selected word has to have a frequency that is equal to or above a certain percentage of the expected frequency. Again, occurrence in as many disciplinary sub-corpora as possible is involved. In Gardner and Davies (2014), the range criterion was thus that a word was selected if it occurred with at least 20 percent of its expected frequency in at least seven out of nine academic disciplines. Also, with regards to general high frequency word lists, range is an important measure for selecting words objectively, but it requires one or more corpora comprising different general language registers. Both in Study 1 in which academic words were identified among the most frequently used lemmas of Danish, and in Study 2, a more simple range criterion was applied which will be justified and explained in Chapters 5 and 6 of this thesis.

While frequency and range indicate how many times a word occurs and where in the corpus, these measures do not tell us how evenly these occurrences are distributed in the corpus. A word can be highly frequent in a given corpus, yet only occur within a small section of said corpus. This is especially the case with technical words. Academic words, conversely, should occur with a more even distribution across the corpora and the disciplines and subject areas contained in it because academic words by definition occur across different disciplines. General high frequency words also have an even distribution due to their ubiquity in all text types. A number of different dispersion measures is used in the development of word lists, but one of the most common ones, used particularly in academic word lists, is the statistical coefficient of Juilland and Chang-Rodríguez’s D commonly referred to as Juilland’s D (Gries, 2008; Juilland & Chang-Rodríguez, 1964). This measure returns a value between 0.01 and 1.0. The closer to 1.0, the more evenly the item is dispersed in the corpus. If an item has a value close to 0.01 it means it only occurs in a very small part of the corpus, e.g. one text or subsection depending on which kind of subdivision of the corpus the measure is calculated. Using a dispersion measure such as Juilland’s D as part of the objective vocabulary selection for a

word list involves setting a cut-off point. The used cut-off values in the academic word list research are: 0.60 (Dang, 2017; Dang et al., 2017; Hagen et al., 2016), and 0.80 (Gardner & Davies, 2014; Paquot, 2010). As Gardner & Davies (2014, p. 316) point out, these cut-off values for dispersion are reached by experimenting with different cut-off points and evaluating the outcome either subjectively or by lexical coverage as Dang, et al. (2017) did. In the identification and word selection carried out in the studies of this thesis, the Juilland's D measure was used for dispersion analysis.

Having described how both objective and subjective criteria form the basis of pedagogical word list development, I will in the next section introduce six lists of academic vocabulary and a study on the need for one in French.

### 3.4.2. Academic word lists

Academic vocabulary not only includes words highly frequent in academic discourse but it can also include items that are also highly frequent in general language. Word list research is one of the main avenues of research into the relationship between general high frequency and academic vocabulary. Studies in academic vocabulary lists in English have approached general high frequency vocabulary in different ways, as our understanding of this field has developed. Coxhead (2000) excluded the first 2,000 words of English in the form of West's General Service List (1953) from the Academic Word List on the assumption that the learners would already know these items. Gardner and Davis (2014, pp. 308–310), on the other hand, point out that academic vocabulary includes general high frequency items, and in line with Paquot (2010), they suggest that methods for developing academic word lists need to be able to include general high frequency words. This can be done by using comparative quantitative measures to extract items that are more frequent in academic language than in non-academic language. Both Paquot (2010) and Gardner and Davies (2014) use such measures in their academic word lists which will be detailed below. The question of high frequency lexical items also influences the development of academic word lists in languages other than English. Asking if there is room for an academic word list, Cobb and Horst (2004) found that the first 3,000 words of French covered large proportions of academic (81.27%) and general (83.88%) texts in the language, and that these general high frequency words of French are used for general and academic purposes. Similar findings of high frequency words occurring in academic texts in English can be found in Coxhead's (2000) analysis of academic written texts in English using West's General Service List (1953), Chung and Nation's (2004) study of technical vocabulary in Anatomy and Applied Linguistics textbooks, and Nation's (2016) finding that a high proportion of Gardner and Davies' (2014) Academic Vocabulary List is high frequency words in English. This research points to the importance of

identifying the amount of academic vocabulary in high frequency words in Danish, as a way to support L2 as well as L1 students studying in Danish higher education, i.e. not assuming that these words will be known in their academic sense and function.

In the following sections, I will review four academic word lists in English. The two lists introduced first are Coxhead's Academic Word List (AWL) (2000, 2011, 2016), and Gardner and Davies' Academic Vocabulary List (AVL) (2014). Paquot's Academic Keyword List (AKL) (2010) and Dang, Coxhead, and Webb's (2017) Academic Spoken Word List (ASWL) are then reviewed.

Information on corpus, word selection criteria, unit of counting, and coverage for each of the reviewed list is summarised in Table 3.4 which is placed after the review of the academic word lists in English.

#### 3.4.2.1 *The Academic Word List (AWL)*

Coxhead's AWL (2000) was not the first academic word list in English (see Champion & Elley, 1971; Ghadessy, 1979; Lynn, 1973; Praninskas, 1972; Xue & Nation, 1984), but it was the first list that used a large corpus representative of academic language and it is still the most widely used list of academic vocabulary in English language teaching (e.g. D. Schmitt & Schmitt, 2005; D. Schmitt, Schmitt, & Mann, 2011). As mentioned earlier, the purpose of the AWL was to assist learners of English in the reading of academic texts at university level and accordingly the corpus behind the list comprises written texts covering the different genres that students read at university: university textbooks, journal articles, book chapters, and laboratory manuals. In addition, Coxhead aimed at a balanced representation of academic disciplines and subject areas in the corpus used. The corpus of 3.5 million running words contains four equally-sized sub-corpora corresponding to the four academic disciplines of Arts, Commerce, Law, and Science. Each sub-corpus covers seven subject areas each of 875,000 words. The unit of counting for the AWL is the word family which is motivated by Bauer and Nation's position that "[c]omprehending regularly inflected or derived members of a family does not require much more effort by learners if they know the base word and if they have control of basic word-building processes (Bauer & Nation, 1993, p. 253)" (Coxhead, 2000, p. 218).

The criteria used for vocabulary selection for the AWL were *specialised occurrence*, *frequency*, and *range*. These criteria ensured that the word families selected were highly frequent in academic language occurring across a wide range of academic disciplines. In particular, the special occurrence criterion entailed the exclusion of the 2,000 words of the General Service List (West, 1953) since a central assumption behind the AWL was that the learners would already know these words.



Table 3.4 provides information about the frequency and range criteria applied by Coxhead (2000). The lexical coverage of the AWL was measured both in the corpus it was based on and in a separate corpus of academic texts in order to ascertain if the criteria for word selection had indeed managed to select frequently occurring words with a wide range in the academic disciplines and sub-disciplines represented in the AWL corpus. The coverage in the AWL corpus was ten percent and in the separate academic corpus about eight percent. These coverage numbers were compared to the coverage of the General Service List in the two academic corpora (76.1% and 70.6%, respectively) and to the AWL's coverage in a corpus of fiction texts (1.4%). Especially the notably low coverage by the AWL in the non-academic corpus led Coxhead to argue that "the majority of the word families in the AWL are associated particularly with academic writing (...)." (Coxhead, 2000, p. 225).

As reported above, the AWL has been criticised for excluding the word families of the General Service List, but also the AWL corpus in itself has been criticised by Hyland and Tse (2007) for being skewed towards law and commerce. Based on their criticism of the skewness of the corpus and coverage analyses, Hyland and Tse (2007) question the construct of a general academic vocabulary claiming that many of the words in the AWL are, in fact, discipline-specific and not generally used across disciplines. Despite its limitations, the AWL is still a valid representation of academic vocabulary as its coverage of academic texts has proven to be consistent regardless of academic discipline of the texts analysed (Coxhead, 2011, p. 357).

#### *3.4.2.2. The Academic Vocabulary List (AVL)*

Gardner and Davies' (2014) motivation for developing an academic vocabulary list comes from the criticism that Coxhead's Academic Word List (2000) had been subjected to. In particular, the authors address the use of word families as the unit of counting and the Academic Word List's relationship to the General Service List (West, 1953). They fully acknowledge the existence of a general academic vocabulary in contrast to Hyland and Tse (2007) (Gardner & Davies, 2014, p. 310). However, they argue 1) that the Academic Word List does not properly represent such a vocabulary especially due to its relationship with the General Service List, and 2) that the Academic Word List is not suitable for especially non-advanced learners due to its use of the word family. This critique should be seen in the light of the fact that the Academic Word List, as also reported by Nagy and Townsend (2012), has been widely used in both primary and secondary education for both L1 and L2 learners in the North-American context.

Consequently, Gardner and Davies's (2014) approach to word selection differs in important ways from that of Coxhead's. Additionally, they took a different approach to corpus compilation. The AVL

is based on an academic sub-corpus of the COCA-corpus which at the time of the study contained 425 million running words (Davies, 2008). This already lemmatised sub-corpus of 120 million running words contains approximately 70 percent academic journal articles from nine academic disciplines of different sizes:

- 1) Education (8,30,324)
- 2) Humanities (11,111,225)
- 3) History (14,289,007)
- 4) Social Science (16,720,729)
- 5) Philosophy, Religion, and Psychology (12,463,471)
- 6) Law and Political Science (12,154,568)
- 7) Science and Technology (22,777,656)
- 8) Medicine and Health (9,660,630)
- 9) Business and Finance (12,824,831)

(Gardner & Davies, 2014, p. 314)

The remainder of the texts are from newspapers and magazines. These text types (about 31.5 million words) were added to all disciplines except Humanities and Education. In addition, the sub-corpus of Business and Finance included newspaper texts because of difficulties of retrieving texts from “formula- and table-laden academic journals dealing with topics such as economics and finance.” (Gardner & Davies, 2014, p. 313). The chosen unit of counting for the AVL is the lemma. Part of the motivation for using the lemma and not the word family as the unit of counting was that it allows for distinguishing between parts of speech and as such makes for “an accurate assessment of word forms, functions, and meanings.” (Gardner & Davies, 2014, p. 313). Moreover, as outlined Chapter 2, a further motivation for using lemmas instead of word families was that derivational word knowledge is developed later than inflectional knowledge, which makes the word family too complex, especially for non-advanced learners (Gardner, 2007; Gardner & Davies, 2014, pp. 307–308).

For vocabulary selection for the AVL, measures of *ratio*, *range*, *dispersion*, and discipline specificity (*discipline measure*) were used. The ratio criterion selects words occurring 50 percent more frequently in the AVL corpus compared to a general language corpus (the remainder part of the COCA corpus). This measure of relative frequency ratio (Gries, 2010) also ensures that general high frequency vocabulary could be included in the list. The range criterion was carried out using a measure of expected frequency: a lemma had to occur with minimum 20 percent of the expected

frequency in at least seven out of the nine discipline sub-corpora. For the dispersion measure, Juilland's D was used with a cut-off value of 0.80. Together with the last criterion, the discipline measure, the purpose of the range and dispersion criteria was to exclude discipline-specific vocabulary from the AVL. The used cut-off points for these three criteria were based on experiments with different ratio values, expected frequencies and D values. The resulting list of academic vocabulary contains 3,000 lemmas which were transformed into 2,000 word families to make the provided coverage comparable to the coverage of the Academic Word List (2000). Looking at the coverage provided by the AVL it provides almost the same coverages over the academic sub-corpora of the COCA (13.8%) and the BNC (13.7%). Further, the coverages of the AVL over other genres are lower which makes the authors argue that the AVL is representative of academic vocabulary. In comparison, the Academic Word List covers these two corpora with 7.2% and 6.9%, respectively. Gardner and Davies contribute the high coverage of the AVL to the fact that the AVL contains more general high frequency vocabulary than the Academic Word List does because of the powerful statistics used in the word selection. Further, they argue that being able to include high frequency vocabulary in their list leads to a new conceptualisation of what academic vocabulary is: that it is not only found outside the general high frequency vocabulary of English represented by the General Service List (West, 1953). This conceptualisation has certainly had influence on later developments of academic word lists such as the one carried out in Study 2 of this thesis. Moreover, many of the methodological choices made by Gardner and Davies have been followed in Study 2.

#### *3.4.2.3. The Academic Keyword List (AKL)*

In methodology, the AKL (Paquot, 2010) is similar to Gardner and Davies' Academic Vocabulary List (2014) in that it makes use of a comparative frequency criterion and used Juilland's D as the dispersion measure. However, the purpose of the word selection for the AKL was first and foremost to operationalise a functional definition of academic vocabulary as the lexical items used "to refer to those activities that characterize academic work, organize scientific discourse and build the rhetoric of academic texts (...)." (Paquot, 2010, p. 29). The AKL words are as such not academic words according to Paquot (2010, p. 29). Rather, they are potential academic words that need to be analysed corpus-linguistically to ascertain their status as academic words (Paquot, 2010, p. 63). In contrast to the corpora used for the other academic word lists reviewed here, the corpus of the AKL comprises both student and professional writing. The argument for including student writing is that academic writing not only involves professional writing such as textbooks and journal articles. Student writing is also a central part of academic writing and, as such, it should be represented in a corpus of academic

writing according to Paquot (2010, p. 31). The corpus contains three million running words and is divided into 15 sub-disciplinary sub-corpora of unequal sizes. These 15 sub-disciplines comprise texts from Science, Humanities, and Social Science, but apparently not from the Health and Medical Sciences. Paquot states that her corpus, though only in relation to the student writing part of the corpus, is skewed towards Humanities and Social Science, but that the applied statistical measures described below are strong enough to extract words that are more frequently occurring in academic texts than in non-academic texts and that are evenly dispersed in the texts of the academic language corpus: “the procedure used to extract potential academic words largely overcome this limitation.” (Paquot, 2010, p. 33). The corpus was lemmatised and annotated automatically with part of speech tags and semantical tags. The latter annotation was done to be able to supplement the word selection described below with semantically related words. The semantic tagging comprised 21 semantic fields. Each lexical item in the corpus was assigned a semantic field category except for closed word classes and proper names. The unit of counting for the AKL seems to be the lemma although it is not explicitly stated.

For word selection, Paquot used keyness (Scott, 1997; Scott & Tribble, 2006) to extract words that were more frequent in the AKL corpus than in a general language corpus. As mentioned in Section 3.4.1 on word list development, keyness and keyword analyses ensure that not only are the words more frequent in academic language, they are also distinctive of academic language because the difference in frequency is statistically significant. The range criterion was a rather simple one: only words occurring in all 15 subject areas were included. To avoid uneven distribution of the selected words, a cut-off dispersion value of 0.80 was chosen. The resulting list contained 599 potential academic words. This list was expanded using the semantic tagging of the corpus to include words that fulfilled the criteria of keyness and range, and were semantically related to the 599 potential academic words. The final AKL contains 930 potential academic words. Paquot reports that a limitation of the AKL is the arbitrary cut-off points used for word selection. A notable thing about the AKL is that it also comprises a small number of multiword units such as ‘according to’, ‘at best’, ‘depending on’, and ‘rather than’. No lexical coverage analyses were carried out, but the overlaps between the AKL and the Academic Word List (Coxhead, 2000) as well as the General Service List were measured. These analyses showed that 40 percent of the AKL words occurred in the Academic Word List while 57 percent of them occurred in the General Service List (West, 1953). According to Paquot (2010, p. 60), this emphasises the important fact that academic vocabulary

comprises general high frequency vocabulary and it justifies the use of comparative frequency criteria such as keyness analysis.

#### *3.4.2.4. The Academic Spoken Word List (ASWL)*

The last English academic word list reviewed, the ASWL (Dang, 2017; Dang et al., 2017), here differs from the other three in that the focus is on spoken academic discourse, a topic which has received somewhat limited attention in the word list research according to Dang, Coxhead and Webb (2017). The purpose of the ASWL is to aid learners of English at university understand academic discourse of the spoken variant. However, in the development of the ASWL, Dang, Coxhead and Webb (2017) acknowledged that English learners in higher education have varying levels of English vocabulary knowledge as shown by e.g. Webb and Chang (2012), Henriksen and Danelund (2015), Matthews and Cheng (2015), and Nguyen and Webb (2017). Therefore, a central principle for the ASWL development was that it should be adaptable to different learner levels, and the resulting list was analysed to determine what level of coverage learners with different levels of English would reach after having learned the ASWL words. For the development of the ASWL, a corpus of spoken academic discourse of about 13 million running words was compiled covering four academic genres: lectures, seminars, laboratory teaching, and tutorials. Discourse was sampled from universities and publishers in different English-speaking countries (the UK, the US, Ireland, Australia, New Zealand, Canada, Hong Kong) thus resulting in a corpus representing a wide variety of spoken academic English. The corpus was divided into the four sub-corpora corresponding to the academic disciplines of hard pure, hard applied, soft pure, and soft applied (Becher, 1989; Becher & Trowler, 2001). In turn, these four sub-corpora each comprised six equal-sized sub-disciplines. The criteria of frequency, range, and dispersion were used for word selection. The frequency criterion is carried out using a cut-off point similar to the AWL, i.e. a member of a word family had to occur a certain number of times in the corpus, but in contrast to the Academic Word List (Coxhead, 2000), the ASWL includes general high frequency words. This results in a rather large number of word families in the final list, many of which belong to general high frequency vocabulary. The inclusion of these general high frequency words is in part what makes the ASWL adaptable to learners with different proficiencies. Learners of English can reach up to 93% coverage of not only spoken academic discourse, but also general discourse if they know the words in the ASWL. For that reason, Dang et al. argue that the ASWL, with its high coverage of both academic and general spoken English, can serve as “a shortcut for low-level learners to achieve basic comprehension of academic speech while still allowing them to enhance their knowledge of general high-frequency vocabulary.” (Dang et al., 2017, p. 26).

#### *3.4.2.5. Summary*

The four academic word lists discussed above are summarised in Table 3.4. In the next sections I will turn to academic word lists in other languages: the Swedish Academic Word List (Ribeck et al., 2014) and the Norwegian Academic Word List (Hagen et al., 2016) as well as Cobb and Horst's (2004) investigation of the necessity of a French academic word list.

Table 3.4. Four academic word lists in English

	<b>The Academic Word List (Coxhead, 2000)</b>	<b>The Academic Vocabulary List (Gardner and Davies, 2014)</b>	<b>The Academic Keyword List (Paquot, 2010)</b>	<b>The Academic Spoken Word List (Dang, Coxhead, and Webb, 2017)</b>
<b>Corpus (size and composition)</b>	<ul style="list-style-type: none"> <li>- 3.5 million words</li> <li>- Professional academic writing</li> <li>- four academic disciplines (arts, commerce, law, and science) and 28 subject areas</li> </ul>	<ul style="list-style-type: none"> <li>- 120 million words</li> <li>- Professional academic writing and newspapers and magazines</li> <li>- nine academic disciplines</li> </ul>	<ul style="list-style-type: none"> <li>- 3 million words</li> <li>- student and professional writing</li> <li>- 15 subject areas</li> </ul>	<ul style="list-style-type: none"> <li>- 13 million words</li> <li>- Spoken academic discourse</li> <li>- four academic disciplines (hard pure, hard applied, soft pure, and soft applied) + 24 subject areas</li> </ul>
<b>Unit of counting</b>	Word family (Level 6)	Lemma	Lemma	Word family (Level 6)
<b>Criteria for vocabulary selection</b>	<ul style="list-style-type: none"> <li>- Special occurrence (exclusion of the GSL words)</li> <li>- Frequency cut-off (100)</li> <li>- Range (“a member of a word family had to occur at least 10 times in each of the four main sections of the corpus and in 15 or more of the 28 subject areas.” (p. 221))</li> </ul>	<ul style="list-style-type: none"> <li>- Relative frequency ratio (1.5)</li> <li>- Range (20% of expected frequency in seven out of nine disciplines)</li> <li>- Dispersion Juilland’s D (0.80)</li> <li>- Discipline measure (exclusion of words occurring three times more than the expected frequency)</li> </ul>	<ul style="list-style-type: none"> <li>- Keyness</li> <li>- Range (occurrence in all 15 sub-corpora)</li> <li>- Dispersion Juilland’s D (0.80)</li> <li>- Semantic relatedness</li> </ul>	<ul style="list-style-type: none"> <li>- Frequency cut-off (350)</li> <li>- Range (occurrence in all four disciplinary sub-corpora and in 50% of the 24 subject areas (p.17))</li> <li>- Dispersion Juilland’s D (0.60)</li> </ul>
<b>General high frequency vocabulary</b>	Excluded in the form of the General Service List (West, 1953)	Included if they were more frequent in the AVL corpus than in a reference corpus (relative frequency ratio)	Included if they were distinctive of the AKL corpus compared to a reference corpus (keyness). 57 percent of the AKL words are also in the General Service List (West, 1953)	Included if they meet the frequency cut off.
<b>Size of the lists</b>	570 word families	3,000 lemmas/2,000 word families	903 lemmas	1,741 word families
<b>Coverage over academic language</b>	8.5-10%	Almost 14%	No coverage analysis	90%
<b>Coverage over general language</b>	1.4% (fiction)	7-8% (newspaper) 3.4% (fiction)	No coverage analysis	87%

#### 3.4.2.6. The Swedish Academic Word List (SAWL)

The purpose of the SAWL (Ribeck et al., 2014; Sköldberg & Johansson Kokkinakis, 2012; Carlund, Jansson, Kokkinakis, Prentice, & Ribeck, 2012; Jansson, Kokkinakis, Ribeck, & Sköldberg, 2012) is to aid advanced learners of Swedish as well as L1 Swedish students who are not acquainted with academic language in their academic writing (Ribeck et al., 2014, p. 370). However, the list can also be used for receptive purposes in that at least the first 100 word of the list are accompanied with meaning, examples of use, and English translation in the online version (Ribeck et al., 2014, p. 371). The corpus behind the list comprises texts automatically compiled from a Swedish research publication database ('SwePub', n.d.). The developers decided only to include texts from Humanities and Social Science as they define Swedish academic written language as "consisting of texts written of and for academics in Swedish" (Ribeck et al., 2014, p. 375, my translation). Using this definition allows for excluding the disciplines of Science and Medical Science as the publication language for these disciplines is primarily English (see Chapter 4 on the issue of publication language). The unit of counting is the lemma and the corpus is annotated accordingly. The developers argue for the use of the lemma as the unit of counting because it allows for closely related items such as *bedöma* (assess) and *bedömning* (assessment) to be listed as separate entries and described as separate entities. Moreover, according to Ribeck et al. (2014, pp. 378–379) this unit of counting is more suitable for productive purposes, which is in line with the purpose of the SAWL. Since academic words are defined as being outside the basic vocabulary of Swedish (Ribeck et al., 2014, p. 376), a stop list of the most frequent 1,000 words in a corpus of easily read texts were used to eliminate these words from the SAWL. This was the first criterion in the word selection procedure and it is similar to Coxhead's special occurrence criterion described above. In addition to this criterion, the developers used *reduced frequency* (Savický & Hlaváčová, 2002) which is a frequency measure that includes dispersion. As such this measure operates with cut-off points and extracts lemmas that occur at a certain frequency and with certain dispersion in the corpus. In particular, to be selected for the word list, a lemma had to have a reduced frequency of at least 15 per million words in all of the 15 subject areas. The number of 15 hits per million was heuristically based by experimenting with different cut-off values as in Gardner and Davies' (2014) Academic Vocabulary List (Ribeck et al., 2014, p. 377). The words selected by this reduced frequency criterion were analysed in relation to keyness (Scott, 1997; Scott & Tribble, 2006) by comparing their reduced frequencies with their reduced frequency in a 2.5 million corpus of fiction texts. Those words that met the set keyness value of 1.1 were included in the final SAWL. After a manual check of the list to remove non-words and English words,



the SAWL contained 655 lemmas. Even though the lemma is the unit of counting for the SAWL, the developers do point out the need to investigate whether the word family in Swedish is a suitable conceptualisation of words in relation to vocabulary acquisition, and to develop a word family word list if that is the case (Ribeck et al., 2014, p. 381). Table 3.5. provides a summary of the SAWL.

*Table 3.5. Overview of the Swedish Academic Word List*

Corpus composition	- Academic writing (dissertations and journal articles) - 15 subject areas from the disciplines of Humanities and Social Sciences
Size of corpus	- 25.4 million running words
Unit of counting	Lemma
Criteria for vocabulary selection	- Reduced frequency - Keyness
General high frequency vocabulary	The 1,000 most frequent words in a corpus of easily read texts were eliminated.
Size of the list	655 lemmas

#### *3.4.2.7. The Norwegian Academic Word List (NAWL)*

The purpose of the NAWL<sup>12</sup> (Hagen et al., 2016) is similar to that of the Swedish (and other academic word lists) in that it should help learners of Norwegian comprehend and produce Norwegian academic texts. In particular, on the website for the word list (Universitetet i Oslo, n.d.), it says that users of the list are learners for Norwegian as a second language and students with non-academic backgrounds. Teachers and material designers are also encouraged to consult the list. After having experimented with the method for developing the SAWL (Ribeck et al., 2014) including the use of stoplists of general high frequency vocabulary, the “Gothenburg method” (because the SAWL was developed at the University of Gothenburg), Hagen, Johannesen and Saidi (2016) decided to follow the methodology of Gardner and Davies’ (2014). The reason for this was that the coverage measured of the list that was generated using the “Gothenburg method” was notably lower than that of the list developed by following the methodology of Gardner and Davies (2014).

The corpus behind the NAWL, the DUO Corpus, consists of a 100 million running words with texts from eight academic disciplines corresponding to the eight faculties of the University of Oslo as all

<sup>12</sup> NAWL in English, in Norwegian NBAO: Norsk Bokmål Akademisk Ordliste.

the texts were collected through this university's digital publication archive. In the word selection for the NAWL, Gardner and Davies' (2014) criteria for their Academic Vocabulary List were used but with other cut-off points. For the relative frequency ratio measure, a cut-off point of ratio values between 2.2-2.6 was chosen based on experiments. That means that in order to be selected, a word had to occur more than twice as frequent in the DUO corpus as in a general language corpus. The chosen ratio values for the cut-off point are relatively high compared to the ratio value (1.5) chosen as cut-off value employed for the Academic Vocabulary List. Hagen et al. argues (2016, p. 1459) that it shows "that these measures are language and culture specific". Also, in the range criterion, the cut-off point was different from that of the Academic Vocabulary List which had 20 percent as the cut-off. For the NAWL, all words occurring with a frequency 30 to 60 percent of the expected frequency in six out of the eight disciplines met the range criterion. The dispersion criterion of the NAWL was that all words with a Juilland's D of at least 0.60 were selected. The last measure of Gardner and Davies, the discipline measure, eliminated all words occurring more than 3 to 3.2 times their expected frequency in the eight disciplines. The use of more than one cut-off point for the criteria of ratio (2.2.-2.6), range (30-60%), and for the discipline measure (3-3.2) naturally resulted in a number of different lists that the developers compared to each other using coverage as the primary comparison measure. Two lists were found to represent academic vocabulary the best. The first list contained words with a ratio value of at least 2.2, a frequency of 40 to 60 percent of expected frequency in six out of eight sub-corpora (range), and a frequency lower than 3 times the expected frequency in any of the eight disciplines (discipline measure). The second list contained words with a ratio value of at least 2.8, a frequency of 30 to 60 percent of expected frequency in six out of eight sub-corpora (range), and a frequency lower than 3.2 times the expected frequency in any of the eight disciplines (discipline measure). These two lists were merged manually into the final academic word list resulting in a Norwegian Academic Word List of 750 lemmas. These different cut-off points can of course be language-related, but nonetheless it is not surprising that different developers use different cut-off points when the choice of these are heuristically based/based on experiment, i.e. manual check. Table 3.6 provides an overview of the NAWL.

*Table 3.6. Overview of the Norwegian Academic Word List*

Corpus composition	<ul style="list-style-type: none"> <li>- Academic writing (master's theses, doctoral dissertations, and journal articles)</li> <li>- Eight academic disciplines (Humanities, Educational Sciences, Medicine, Social Sciences, Natural Sciences, Theology, Law, and Dentistry).</li> </ul>
Size	- 100 million running words
Unit of counting	Lemma
Criteria for vocabulary selection	<ul style="list-style-type: none"> <li>- Relative frequency ratio (2.2-2.6)</li> <li>- Range (30-40% of expected frequency in six out of eight disciplines)</li> <li>- Dispersion Juilland's D (0.60)</li> <li>- Discipline measure (exclusion of words occurring 3-3.2 times more than the expected frequency)</li> </ul>
General high frequency vocabulary	Included if they were more frequent in the academic corpus than in a reference corpus (relative frequency ratio)
Size of the list	750 lemmas

#### *3.4.2.8. Is there room for an academic word list in French?*

The study of Cobb and Horst (2004) does not per se involve the development of an academic word list. Instead it explores, as the title suggests, if a French academic word list similar in design and purpose to the Academic Word List (Coxhead, 2000) is feasible and necessary in French by measuring the lexical coverage of the 3,000 most frequent lemmas in French. In measuring the lexical coverage of French, Cobb and Horst (2004) applied the Lexical Frequency Profiling (LFP) framework (Laufer & Nation, 1995a) developed for English and used in numerous English vocabulary studies since the 1990's (see Chapter 5 for more on LFP). As Cobb and Horst (2004, p. 16) point out, the LFP framework has only been applied to other languages to a limited extent, and their study is an attempt to apply the framework to French. To ascertain if there is room for an academic word list in French, the authors took a list of the 3,000 most frequent lemmas in French and divided it into three 1000-items levels: The first 1,000 (equal to 1000 most frequent lemmas in French), the second 1,000 (equal to second 1000 most frequent lemmas in French), and a third level termed the AWL. Thus, Cobb and Horst equalled these third 1,000 most frequent lemmas to Coxhead's Academic Word List (2000) since this list excludes the first 2,000 words of in English in the form of the General Service

List (West, 1953), as mentioned above. The authors acknowledge that the Coxhead's Academic Word List comprises words from other frequency levels than the third 1,000 and emphasise that the French third 1,000 list is only a hypothesised French Academic Word List. Using the *Vocabprofile* programme, they uploaded the three frequency lists to the programme and a range of academic and non-academic texts. The results showed that the third 1,000 lemmas of French, or what the authors called "the experimental French AWL" (Cobb & Horst, 2004, p. 31) provided almost the same coverage of academic and non-academic texts (3.57% and 3.30%, respectively). Cobb and Horst also measured the coverage of English academic and non-academic texts using the General Service List and the Academic Word List as frequency level lists. This analysis showed that the first 2,000 lemmas of French provided a higher coverage of French academic texts (81.27%) than the General Service List over English academic texts (70.42%).

The results of Cobb and Horst (2004) suggest that French does not have a separate academic vocabulary. Instead, French academic words are also general high frequency words. In other words, for learners of French to reach a 90% coverage of academic texts, they need to know only the first 2,000 words of French which allows for a more efficient naturalistic acquisition process (Cobb & Horst, 2004, p. 36). In contrast, findings for Dutch suggest that it requires up to 10,000 words to read and understand an academic text (Hazenberg & Hulstijn, 1996). For English, the General Service List plus the Academic Word List provide a combined coverage of 90 percent of academic texts (Coxhead, 2000). The explanation for this minimal number of words required to understand academic texts in French is, according to Cobb and Horst (2004, p. 35), that French academic language does not make use of a distinct vocabulary in academic discourse as English to some degree does (cf. Nagy & Townsend, 2012). An additional analysis of a group of 42 French words showed that 63 percent of these occurred in translation in the Academic Word List and 56 percent among the first 2,000 lemmas of French. This analysis further supports the conclusion drawn by Cobb and Horst that academic words in French are also high frequency words. Cobb and Horst's study rests on the assumption that academic vocabulary is a distinct vocabulary from general vocabulary and that the Academic Word List represents this vocabulary. However, it is important to note that the exclusion of the General Service List (West, 1953) from the Academic Word List may not entirely be based on a perception of academic vocabulary being outside the general high frequency vocabulary as often assumed. Rather, the word selection for the Academic Word List was based on the assumption that English learners on their way to university would already know the first 2,000 words of English as represented by the General Service List. As mentioned in relation to the Academic Spoken Word List (Dang et

al., 2017), this assumption has been questioned by research that finds that even relatively advanced learners of English are not stable within the first 2,000 words of English. However, including general high frequency vocabulary in academic word lists does not resolve this problem entirely since the occurrence of general high frequency words in academic vocabulary is also a question of general words' polysemous nature. What Cobb and Horst's study highlights is the importance of investigating general high frequency vocabulary in relation to academic vocabulary and development of academic word lists to reach an understanding of the function and meaning of general high frequency words in academic discourse. As Cobb and Horst (2004, p. 36) conclude: "(...) the challenge of learning academic uses of common words is probably just as great as learning new academic words." In Study 1, such an investigation is carried out in relation to Danish general high frequency vocabulary.

### 3.5. Rationale for the studies of this thesis

In this chapter on word lists, I have described what word lists are used for and how they are developed. As mentioned in the introduction of this chapter, the focus on pedagogical word lists is motivated by the fact that even though the primary aim of the studies of this thesis is to identify and describe academic vocabulary in Danish, they are very much inspired by research carried out in other languages, especially in English, which has focused on creating lists of academic words for learning purposes. These lists have, however, been used for research describing and analysing academic language use and vocabulary which emphasises the applied nature of much academic vocabulary research. A part of this chapter has focused on Danish research on general high frequency words and the issue of general vocabulary constitutes an important discussion across academic word list development as shown in this chapter. Also, in Chapter 2 on vocabulary divisions, particularly in the Nordic context, this issue was shown to play a significant role in relation to the vocabulary of education and research. Because of this, an identification analysis of Danish academic vocabulary must involve an investigation of the nature of general high frequency vocabulary in relation to academic language which is what Study 1 (Chapter 5) centres on. In order to identify and describe Danish academic vocabulary, a word list approach similar to the ones described in this chapter is taken in Study 2 (Chapter 6). The motivation for doing so rests on the assumption that identifying and selecting lexical items for a word list through corpus-based, quantitative, and objective measures, primarily, forms a solid foundation for both pedagogical developments and for further research into Danish academic vocabulary. In Study 2, the resulting word list is described in relation to a number of issues such as frequency, part of speech, and overlap with general high frequency vocabulary. Study 3 (Chapter 7) continues the description of Danish academic vocabulary by analysing the words

in relation to their function in academic writing. Study 4 (Chapter 8) investigates the boundaries of the Danish academic word list by analysing the words that did not meet the criteria for vocabulary selection for the word list. These words were analysed according to their morphological and semantical relatedness to the DAWL words in order to determine if they could be added to the DAWL. Before moving on to the four studies, the next chapter of this thesis, Chapter 4, describes the data in the form of corpora used for the various analyses carried out in the four studies.

## Chapter 4. Corpora

### 4.1. Introduction

The aim of this chapter is to give an overview of the corpora used in the four studies of this thesis focusing on design and compilation. Particular attention is given to the design and compilation of the AcaDan corpus used in all three studies. At the onset of the project, no large collection of Danish academic texts had been compiled, so the AcaDan corpus is the first Danish corpus of Danish academic language ever established. Coxhead's seminal work on developing the Academic Word List (2000) established corpus-linguistics as a primary method for developing word lists of academic vocabulary and exploring academic vocabulary. Likewise, the three studies in this thesis all employ corpus-linguistic methods. The corpus-linguistic method consists of a corpus, which is to be understood as an electronically stored collection of authentic texts representing a certain type of language use, and of corpus tools (e.g. concordance, frequency counts, word list.) to be used for the exploration of the corpus (McEnery & Hardie, 2011, p. 1). While a corpus represents a particular use of language whether it is written, oral, specialised or general language use, it is worth noting that a corpus is only a sample of a given language type. Thus, when we analyse corpus data, we are still analysing language out of context, and our results will not provide us with a full picture of the context to which the language use belongs (Hunston, 2002, pp. 22–23). On the other hand, a corpus does allow linguistic analyses of vast amounts of data in an efficient and speedy manner. In other words, it enables researchers to find support for theoretical assumptions with the exploration of not just a few texts but a wide range of texts in both quantitative and qualitative ways.

This chapter begins with a brief overview of the corpora used in the three studies of this thesis (Section 4.2) and continues with an introduction to corpus design (Section 4.3). After an account of designing and compiling a corpus representative of Danish academic language use, the AcaDan corpus is described (Section 4.4). Then, the other corpora used in three studies are outlined (Sections 4.5 and 4.6). These corpora are independent of the AcaDan corpus.

### 4.2. Overview of corpora used in this thesis

Before going into detail about the design and compilation of the corpora used in the three studies, an overview of these corpora is given in Table 4.1. Each of these corpora is described in the succeeding sections.

Table 4.1. Overview of corpora used in this thesis

Corpus	Number of running words	Content	Use
<i>Specialised corpora</i>			
The AcaDan Corpus	Approximately 3.3 million	Professional written academic Danish: Research articles and reports (479 texts) from the four academic disciplines of: Humanities (156 texts, 1,091,674), Social Science (160 texts, 1,065,346), Natural Science (80 texts, 585,369), and Health Science (83 texts, 595,600)	Studies 1, 2, and 3
Second Academic Language Corpus	Approximately 298,000	Professional written academic Danish: 7 monographs from Humanities (243,411), 4 research articles from Social Science (23,732), and 6 research articles from Natural and Health Sciences (30,659)	Study 2
<i>General corpora</i>			
Journalisten.dk (a sub-corpus in the Danish Web 2017 corpus)	Approximately 4.6 million	Content from the website <a href="http://www.journalisten.dk">www.journalisten.dk</a>	Study 1
Journalisten.dk (a sub-corpus in the Danish Web 2014 corpus)	Approximately 6.2 million	Content from the website <a href="http://www.journalisten.dk">www.journalisten.dk</a>	Study 2
General language Corpus	Approximately 330,000	Professional written general Danish 265 texts: 240 feature and news articles, 20 leisure magazine articles, and 5 sets of teaching materials	Studies 1 and 2



### 4.3. Corpus design

In the studies of this thesis, two types of corpora are used in the analyses carried out: general and specialised corpora. According to McEnery, Xiao, and Tono (2006), a general corpus is used for describing language or language varieties in general. An example of a general corpus is the Danish KorpusDK (Det Danske Sprog- og Litteraturselskab, n.d.-c) which can be used to describe general written Danish from 1990 to today as it consists of a broad range of primarily general language texts such as newspaper articles and fiction texts. Similarly, the British National Corpus (BNC) (BNC Consortium, 2007) comprises a variety of text types, both written and spoken, and has been used in numerous studies of English vocabulary (e.g. Kennedy, 2003; Leech & Rayson, 2014; Nation, 2004). A specialised corpus, on the other hand, is used for studying a particular domain and consists of texts representative of this domain (McEnery et al., 2006). The AcaDan corpus is an example of a specialised corpus to be used for a particular purpose, namely that of developing an academic word list for Danish. Since a corpus-based word list will automatically reflect the corpus it is based on (Nation & Sorell, 2016), it is pivotal to consider the kind of language to be included in such a corpus. Here the interrelated concepts of **representativeness** and **balance** are key (McEnery et al., 2006). In the case of developing a corpus of academic language for word list development, representativeness refers to which kind of academic language use the corpus should include. An academic language corpus may contain spoken and/or written academic language and it may contain different academic genres such as monographs, journal articles, theses, and textbooks. The concept of balance is closely related to representativeness, since a corpus must represent the language use in question, ensuring that the different registers, modes, and genres of a particular language type are proportionally balanced. An example of a specialised corpus compiled for investigating academic spoken language use is the corpus used for developing the Academic Spoken Word List (Dang, 2017; Dang et al., 2017), and subsequently for developing more discipline-specific word lists for spoken academic discourse (Dang, 2018a, 2018b). This corpus comprised lectures, seminars, labs, and tutorials from a variety of subject areas and from different varieties of English in balanced proportions, thus representing English academic spoken discourse in a balanced way.

### 4.4. The AcaDan corpus – design and compilation

As can be seen in Table 4.1, the AcaDan corpus consists of 3.3 million words of professional academic writing from the disciplines of Humanities, Social Science, Natural Science, and Health Science. In this section the process of designing and compiling the AcaDan corpus is described in detail. This process entailed a range of decisions related to issues such as language type, academic

disciplines, metadata, corpus software, corpus mark-up and annotation, copyright, and corpus size. Many of these decisions are interrelated. Moreover, even if Figure 4.1 conveys the development of the AcaDan corpus as a stepwise process, it is in many ways better described as an interactive circular process in which the developer goes back and forth between the three steps, as a decision for a component of one step may affect a component of another step. However, in the succeeding sections an attempt has been made to describe these decisions and steps as separate issues to make the development of the AcaDan corpus as transparent as possible to the reader.

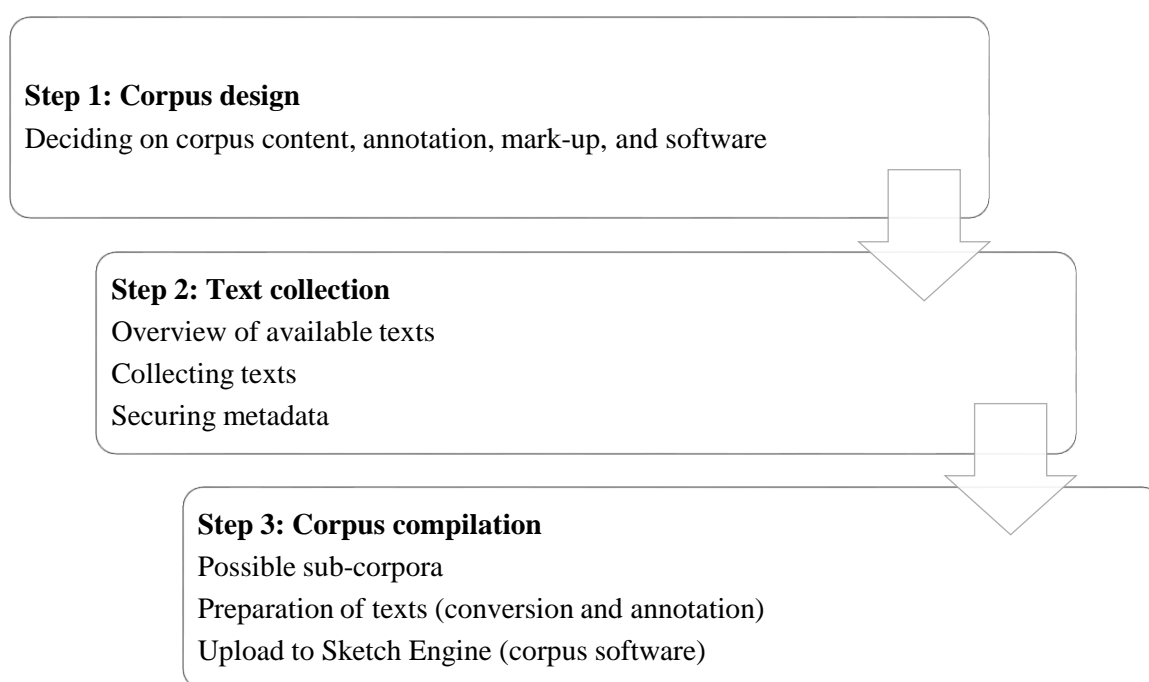


Figure 4.1. Overview of the development of the AcaDan corpus

#### 4.4.1. Language represented: Professional academic writing

The AcaDan corpus represents *professional academic writing*, i.e. only journal articles and reports written by professionals and not students were included as text sources. The academic, peer-reviewed journal article was chosen as the primary text source for the corpus for two related reasons: First, the academic journal article is a result of not only the author's writing, but also of the peer-review and editing processes. As such, it is expected that the language is of high quality and reflects shared norms of academic writing. The same cannot be said of student academic writing (e.g. BA and MA theses) which is why this form of writing was not included in the corpus even if other studies on academic vocabulary include such writing (e.g. Hagen et al., 2016; Paquot, 2010). The second reason for choosing the journal article as representative of professional academic writing is that the academic

article can be said to represent archetypal or prototypical academic language use as it has become one of the most common ways of publishing research. Moreover, in the journal article, researchers in communicating their research to their peers build up an argument for their analysis and findings using different kind of linguistic and rhetorical resources. As Hyland (2004) argues, it is through published texts that the researchers establish themselves as belonging to the discipline they write within, i.e. the shared “community of practice” (Wenger, 1998). As such, the research article as a published text not only reflects the researcher’s academic membership, but it also linguistically establishes this membership. However, there was also a practical side to why the journal article was chosen as a primary text source for the corpus. The open access principle makes it, to some extent, easy to access articles on the Internet and download the journal articles. Professional academic writing, however, does not only include journal articles. It also includes other forms of writings such as book chapters, monographs, and scientific reports.

Taking into consideration that more than 80 percent of all academic publications in Denmark are published in English (Hultgren, 2013), the mere act of compiling Danish academic texts for a large-sized and well-balanced corpus can constitute a significant challenge. Especially, collecting journal articles from the disciplines of Natural and Health Sciences proved rather difficult. It was therefore necessary to consider other publication types for these two disciplines. I decided to include the scientific report as a text type as this is a common form of publication within Natural and Health Science. A search in the Danish National Research Database (Uddannelses- og Forskningsministeriet, n.d.) shows that it contains circa 11,500 scientific reports written in Danish from Natural and Health Sciences and only around 5,000 from Humanities and Social Science. It should be noted that the target group for this type of publication is different from that of journal articles as scientific reports are often written for the authorities. Moreover, the purpose is to inform and substantiate in a research-based and empirical way the decisions of said authorities. On the other hand, the scientific report as it appears in the Danish context in many ways reflects the structure of the research articles with sections on previous research, methodology, results, discussion, limitations, conclusions and implications. The large number of such reports written in Danish within the hard sciences suggests that the scientific report can be considered a prototypical form of Danish academic writing within these disciplines. Still, it could be argued that these reports might be less academic in their language use because their audience is not academic peers but authorities and practitioners. The language of such report has, to my knowledge, not been explored in previous research, but a reading of a sample of these reports did not reveal a language use markedly different from that of the academic journal

article. It would be interesting to get further insight into the particular language use of these reports as they constitute a significant part of what is written in Danish within the disciplines of Natural and Health Science. This is, however, not within the scope of this research project.

To summarise, the AcaDan corpus comprises professional academic writing in the form of academic journal articles and scientific reports. The language represented by these two genres can be considered a form of prototypical academic language which makes the AcaDan corpus especially suitable for the development of a Danish academic word list and for investigating Danish academic words in general. In the next section, I focus on the importance of having a broad representation of academic disciplines in a corpus used for investigation of academic vocabulary and developing word lists.

#### 4.4.2. Academic disciplines

When developing academic word lists, the corpus representation of disciplines and subject areas is an essential methodological concern. Given that academic vocabulary is defined as occurring across a broad range of academic disciplines, the corpus used for the development of an academic word list must be divided into multiple sub-corpora according to the disciplines and/or subject areas represented in the corpus. A sub-corpus is a component of a corpus and consists of a particular subset of the genres, modes, register, and/or domains present overall in the corpus of which it is a sub-corpus (McEnery et al., 2006, p. 350). Defining what constitutes a discipline and distinguishing between these branches of knowledge is important for organising an academic language corpus into disciplinary division. As accounted for in Chapter 3 on academic word list studies, different approaches have been taken to divide the corpus content into disciplinary sub-corpora. One of these approaches is Becher's (1989) and Becher and Trowler's (2001) classification of academic disciplines into hard-pure, hard-applied, soft-pure, and soft-applied. According to this classification framework, disciplines are categorised as hard because of their adherence to a paradigm. On the contrary, soft disciplines adhere to a lesser degree to a fixed paradigm. The pure-applied distinction is linked to whether the discipline applies its research findings to the solution of practical problems. The classification has been used for organising a number of academic language corpora of English such as the British Academic Written English Corpus (BAWE), the British Academic Spoken Corpus (BASE), and the academic corpus used in the development of the Academic Spoken Word List (Dang et al., 2017).

Other academic word lists studies have taken a more data-driven approach to the disciplinary division issue letting the sources of the corpus decide the disciplinary division. For example, the corpus behind

the Norwegian Academic Word List, the Academic DUO Corpus, is divided into eight sub-corpora based on the eight faculties of the University of Oslo as the texts were all collected from this university (Hagen et al., 2016, p. 1457). The same seems, at least to some degree, to be the case for the corpus of the Swedish Academic Word List which used the Swedish national research database, SwePub ('SwePub', n.d.), as their source. The content in SwePub is organised according to a national standard based on the OECD classification Field of Science and Technology (Organisation for Economic Co-operation and Development, 2007) (henceforth the OECD classification). The authors argue that using the OECD classification ensures an unbiased disciplinary division as it is "an official typology of Academic subjects" (Carlund et al., 2012, p. 22). Undoubtedly, the OECD classification offers a more objective categorisation of disciplines than using the faculty division of an institution which, at least to some degree, can seem to be guided by organisational rather than scientific motives. The OECD classification, which was originally developed for the purpose of collecting statistics on research and development ('Revision of the Frascati Manual - OECD', n.d.), operates with six research domains as outlined in Table 4.2. As can be seen, this classification does not operate with the dimensions of hard-soft or pure-applied. However, it does offer a fine-grained division of what in Becker's classification are the hard sciences into four separate categories presumably based on the volume of these disciplines. These categories are marked with an asterisk in Table 4.2.

*Table 4.2. The six research domains as outlined in the OECD classification*

<b>Natural Sciences *</b>	<b>Engineering and Technology *</b>	<b>Medical and Health Sciences *</b>	<b>Agricultural Sciences *</b>	<b>Social Sciences</b>	<b>Humanities</b>
-------------------------------	---	--	------------------------------------	----------------------------	-------------------

The Danish National Research Database (DNRD) divides research into four main areas: Humanities, Social Sciences, Medical Science, and Science/Technology. While the DNRD does not offer any justification for this division, these four areas or domains are similar to the traditional division of sciences into hard and soft sciences given in Becher's (1989) and Becher and Trowler's (2001) classification. Moreover, it reflects a bipartite division common in Danish Higher Education into so-called *wet* and *dry* sciences. The wet sciences comprise Medical Science and Science/Technology, and the dry sciences comprise Humanities and Social Science as outlined in Table 4.3. This colloquial and not empirical attested wet/dry dichotomy assumedly reflects the nature of the work carried out in the respective sciences. The contention is that wet sciences use laboratory work whereas the dry sciences use books and other kind of dry material to do research. This dichotomy, however, does

reflect the occasionally sharp distinctions between these four main research areas. In addition, the dichotomy is commonly used in the debates on language policy and choice in higher education, as there is a somewhat clear pattern when it comes to the choice of language for teaching, research and publication with especially the wet sciences increasingly preferring English to Danish (Hultgren, 2013). It is not within the scope of this thesis to explore or discuss the language choices of the academic disciplines in Denmark, but, as mentioned above, the increasing use of English for publication within the fields of Science/Technology and Medical Science makes it particularly difficult to find peer-reviewed publications from these academic disciplines. It can even be argued that it is not necessary to include the academic Danish language use from these disciplines in a corpus representing Danish academic language use, exactly because of the decreasing use of Danish. This approach was taken in the development of the Swedish Academic Word List (Ribeck et al., 2014). The corpus behind this list thus only contained texts from the Humanities and Social Sciences.

*Table 4.3. Classifications of academic disciplines*

<b>Classification</b>	<b>Disciplines</b>					
Danish National Research Database	Humanities	Social Science	Medical Science	Science/Technology		
OECD classification	Humanities	Social Science	Medical and Health Sciences	Agricultural Sciences	Engineering and Technology	Natural Sciences
	“dry sciences”			“wet sciences”		
Becher’s (1989) and Becher and Trowler’s (2001) classification	Soft-pure and soft-applied sciences			Hard-pure and hard-applied sciences		

Above, I have introduced four approaches to categorising academic disciplines. These are outlined in Table 4.3. Strong arguments can be made for using either Becher’s (1989) and Becher and Trowler’s (2001) classification or the OECD classification in the organisation of the content of the AcaDan corpus, as both classification schemes are highly valid and transferable across languages. A possible criticism of any disciplinary classification framework, however, is that they run the risk of not in a

sufficient way reflecting the interdisciplinary nature of much research (Nesi, 2002, p. 354). For instance, the AcaDan corpus contains a number of texts from sports science which is a highly interdisciplinary area of research using methodologies from both social science as well as from natural and health sciences. In order to present the disciplinary divisions of the AcaDan Corpus in a way that reflects the way the disciplines and sub-disciplines see themselves, I decided to use the main research area division of the Danish National Research Database for organising the content into disciplines. This disciplinary division is well-known in the Danish context and will therefore be recognisable divisions for the users of the word list and for dissemination of the results in Denmark. I will in the remainder of the thesis refer to them as *academic disciplines* and not main research areas, and I have adjusted the names of the two hard science disciplines as displayed in Table 4.4.

*Table 4.4. The four academic disciplines represented in the AcaDan corpus*

Humanities	Social Science	Health Science	Natural Science
Soft sciences/dry sciences		Hard sciences/wet sciences	

#### 4.4.3. Corpus platform

The corpus software programme Sketch Engine (Kilgarriff et al., 2014; Kilgarriff, Rychlý, Smrž, & Tugwell, 2004) was chosen for storage of the AcaDan corpus and as the primary tool for working with the AcaDan corpus because the platform offers both storage and tools for working with corpora. The Sketch Engine platform contains already established corpora such as the British National Corpus, a 100-million word corpus of both written and spoken British English (BNC Consortium, 2007) and the Ten Ten corpus family (Jakubíček, Kilgarriff, Kovář, Rychlý, & Suchomel, 2013), which consists of very large corpora of different languages automatically created from web texts. The Danish version of the Ten Ten corpus family was used in Studies 1 and 2 of this thesis and it is described below in Section 4.6.

#### 4.4.4. Text collection

This section begins with an account of how texts were collected for the AcaDan corpus including a description of the resources used in this process. A sub-section of this section details the sources of the collected texts.

Two sources were used for collecting texts for the AcaDan corpus: The University of Copenhagen (UCPH) CURIS database and the Danish National Research Database (henceforth DNRD). UCPH

requires that its researchers register their research publications and similar academic work in the CURIS database in order for UCPH to be able to report all publications to the DNRD. From the CURIS database, six reports were generated<sup>13</sup> that each included an overview of the research publications registered by the UCPH researchers in the form of journal articles, books, anthologies, doctoral dissertations, reports, and contributions to books, anthologies, and reports. The six reports corresponded to the six faculties of UCPH: Humanities, Social Science, Law, Theology, Natural Sciences, and Health Sciences. Each report contained four categories: publication type (i.e. journal articles, books, anthologies, dissertations, reports, and contributions to books, anthologies and reports), ID number (each registered publication has its own unique ID number), original title of the publication, and electronic version of the publication (i.e. a hyperlink). The publications were grouped according to publication type and ordered alphabetically. The format of the reports was an MS Word file, and the content of each file was in the form of a table containing the stated categories as headlines for each column. The tables of each report were copied into individual MS Excel files one for each faculty. Since I had decided to use only journal articles and scientific reports for my corpus, I removed all other kinds of publication thus ending up with a list of articles and reports from each faculty to download. The next step was to start downloading the provided texts in PDF format. I soon discovered, however, that even though it had been a criterion for the generation of the reports that the publications were accessible online, this was not always the case. Often the provided link would not lead me to a full version of the article or report in question, and sometimes the link was not working. Moreover, Danish is the default publication language in the CURIS database, and the researcher must therefore actively choose English as publication language when registering publications. Consequently, some of the articles were in fact written in English and not in Danish even though the publication language was stated as Danish. A more effective way of accessing the publications listed in CURIS reports was to search the title of the article or report in the Danish National Research Database. The DNRD divides publications according to main research area (which I refer to as academic disciplines), review type (peer review or undetermined), type of publication (journal article, monographs, etc.) and scientific level (scientific, educational, or popular), and provides this information together with information on author(s), their affiliation(s), and when (year) and where the publication is published (source). Moreover, the DNRD indicates if there is open access to the publication and provides a link to where to access it. Only publications categorised in the DNRD as peer-reviewed were collected for the AcaDan Corpus with a few notable exceptions. These exceptions

---

<sup>13</sup> Communication officer Suzanne Løje from the Faculty of Humanities helped me generate the reports.



were made for articles from the Natural Science and the Health Science disciplines. Here a few articles of the undetermined review type in the DNRD were included, as collecting peer-reviewed articles from these domains proved rather difficult in comparison with collecting articles from the Humanities and the Social Science disciplines. The undetermined review type covers publications that have not undergone traditional peer review but may have undergone some kind of academic review by an editor or a group of editors. Therefore, in a few incidences, I chose to include articles from this category if I could tell via the source journal or magazine that they had been reviewed by e.g. editors and were aimed at fellow researchers. Approximately ten articles from the Natural Science and the Health Science disciplines are not peer-reviewed in the traditional sense of the word.

Given that it is desirable to have a balanced representation of disciplines in an academic corpus to be used for word list development, additional text collection was necessary for the Natural Science and the Health Science disciplines. The text collection detailed above did not yield a sufficient number of articles from these two hard disciplines compared to the two soft disciplines. For the additional text collection, the DNRD was used as the primary source. In the DNRD, it is possible to choose different search parameters in terms of publication type, publication year, author, research institution, journal title, submission year, language, scientific level, publication status, review type, and main research area (Science/Technology, Medical Science, Humanities, and Social Science). The research database includes links to full online versions if available, and I primarily chose the results in which a full version link was given. From there, I accessed the article or report and downloaded it. In addition, I used the website [www.tidsskrift.dk](http://www.tidsskrift.dk) hosted by the Royal Library (Det Kongelige Bibliotek, n.d.) and a number of Danish and Nordic academic journals' websites.

#### *4.4.4.1. Result of the text collection – texts, disciplines, and sub-disciplines*

In total 479 texts, 448 articles and 31 scientific reports, were collected for the AcaDan Corpus resulting in a corpus of 3.3 million running words. Table 4.5 shows how many texts each of the four academic disciplines contain and the size of each discipline (number of reports are given in brackets). As can be seen, the soft or dry disciplines of Humanities and Social Science comprise roughly 65 percent of the running words and a little above 65 percent of the number of texts. The reports, which as mentioned above are found in the two hard disciplines, comprise almost seven percent of the texts in the corpus and 23 percent of the running words.

Table 4.5. Texts and running words in the AcaDan corpus

Discipline	Number of texts	Percentage of texts	Number of running words	Percentage of running words
Health Science	83 (14)	17.33	594,600	17.54
Humanities	156 (0)	32.57	1,091,674	32.71
Natural Science	80 (17)	16.70	585,369	17.81
Social Science	160 (0)	33.40	1,065,345	31.92
<b>Total</b>	479 texts	100	3,336,988	100

The reason for the unequal sizes of the four disciplines is that, as mentioned earlier, the hard sciences publish far less in Danish than soft sciences do. In the amount of time available for developing the AcaDan corpus, it was not possible to gain an equal number of texts for each discipline. To compensate for this, for some of the analyses carried out in Studies 1 and 2, the two hard science disciplines of Health Science and of Natural Science were merged into one hard science sub-corpus titled Natural and Health Sciences. This merger resulted in three more or less equal-sized sub-corpora as displayed in Table 4.6.

Table 4.6. Three sub-corpora of the AcaDan corpus used in Studies 1 and 2

Sub-corpus	Texts	Number of running words	Percentage of running words
Humanities	156	1,091,674	32.71
Natural and Health Sciences	163	1,179,969	35.36
Social Science	160	1,065,345	31.92
<b>Total</b>	479	3,336,988	100

In addition, in Study 2, all four academic disciplines were divided into two sub-corpora each as shown in Table 4.7. The details of this division are given in Study 2, Chapter 6.

Table 4.7. Six sub-corpora of the AcaDan corpus used in Study 2

Sub-corpus	Texts	Number of running words	Percentage of running words
Health Science	83	594,600	17.82
Humanities_A	86	565,974	16.96
Humanities_B	70	525,700	15.75
Natural Science	80	585,369	17.54
Social Science_A	80	545,686	16.35
Social Science_B	80	519,659	15.57

### *Sub-disciplines*

In Table 4.7, the 479 texts are divided according to 37 sub-disciplines of the four academic disciplines. Table 4.7 also displays how the texts are distributed according to the six OECD categories outlined in Section 4.4.2. As can be seen, the sub-disciplines are not equally represented since the collection of texts was guided by what could be found within each of the four academic disciplines. In all, the AcaDan corpus contains 37 sub-disciplines, and even if they are not of equal sizes, the sub-discipline categorisation of the 479 texts allows for future comparative corpus-linguistic analyses of these sub-disciplines.

*Table 4.7. Sub-disciplines and OECD disciplines in the AcaDan corpus*

<b>Academic disciplines</b>	<b>Sub-disciplines</b>	<b>OECD disciplines</b>
Health Science (83)		Medical and Health Sciences (83)
	Dentistry (29)	
	Medicine (28)	
	Public Health sciences (15)	
	Sport and Fitness sciences (11)	
Humanities (156)		Humanities (156)
	Archaeology (5)	
	Classical philology (11)	
	Cultural sciences (6)	
	Educational science (13)	
	History (12)	
	History of art (5)	
	Linguistics (30)	
	Literature (20)	
	Media science (15)	
	Musicology (6)	
	Philosophy (5)	
	Religion (11)	
	Rhetoric (10)	
	Theology (7)	
Natural Science (80)		Agricultural Sciences (31)
	Agriculture (5)	
	Food and Resource Economics (7)	
	Food sciences (8)	
	Veterinary sciences (11)	
		Engineering and Technology (11)
	Engineering (11)	
		Natural Science (38)

	Mathematics (2)	
	Physics (5)	
	Chemistry (1)	
	Geography (10)	
	Geology (6)	
	Biology (14)	
Social Science (160)		Social Science (160)
	Anthropology (20)	
	Economics (5)	
	Ethnology (11)	
	Information science (40)	
	Law (7)	
	Political science (22)	
	Psychology (6)	
	Sociology (49)	

Compared to contemporary academic language corpora in English, Swedish, and Norwegian (see below), the number of texts collected for the AcaDan corpus resulted in a somewhat limited corpus size (see Table 4.5). Therefore, the next section discusses the size of the AcaDan corpus in comparison with similar academic vocabulary studies.

#### 4.4.5. Size of the AcaDan Corpus

Technological advances in the past twenty years have made it easier in many respects to compile very large corpora. This is especially the case for general language corpora. For example, the monitor corpus used for updating the Danish Dictionary (Det Danske Sprog- og Litteraturselskab, n.d.-a) contains around 880 million tokens. Also, newly established specialised corpora of academic language are of considerable sizes. The corpora used for developing the Norwegian and Swedish Academic Word Lists (Johansson et al., 2017) are of more than 100 million words. Dang, Coxhead and Webb (2017) used a corpus of about 13 million words to develop the Academic Spoken Word List, and Gardner and Davies' Academic Vocabulary List (2014) was based on the academic section of the 120 million sized Corpus of Contemporary American English (Davies, 2008). In comparison, the AcaDan corpus of 3.3 million words seems rather small in size, but its size is comparable to corpora used in other academic word list studies such as Simpson-Vlach and Ellis' (2010) Academic Formulas List (see also Coxhead, 2000; Paquot, 2010). Many of these English studies on academic vocabulary, with the notable exception of Coxhead (2000), make use of already developed corpora of academic English. While the researchers still had to make decisions concerning the overall design

of the corpus in relation to the purpose of their investigations, they did not have to go through the time-consuming process of collecting and preparing texts. As no corpus of Danish academic language exists, it has been crucial for this study to carefully collect and prepare texts in a manner consistent with the study's research aims and questions. This first and very laborious task of compiling the AcaDan corpus was therefore a pre-requisite for taking on the central research aims of the project outlined in relation to Studies 1, 2, and 3. In this respect, the size of the corpus reflects the time resources available in the present project. Future studies on academic Danish can expand the corpus if needed in relation to research aims specified.

#### 4.4.6. Authors, publications, and time of publication in the AcaDan Corpus

In the following section, the number of authors and publications together with time of publication are described and how these are distributed according to the four academic disciplines of the corpus. The purpose of this description is to address the issue of representativeness and balance as discussed in Section 4.3 on corpus design. A high number of individual authors represented in a corpus minimises the risk of having language use dominated by one or a few authors' idiosyncratic writing style and word choice. The same can be said about having as many different text sources, i.e. journals, represented. The issue of publication date also refers to the issue of representativeness as the aim of the AcaDan corpus is to represent contemporary written academic language use.

The 479 texts in AcaDan are written by 684 individual authors. Table 4.8 provides the number of individual authors and how many authors a text has in average in each of the four disciplines.

*Table 4.8. Authors in the AcaDan corpus*

<b>Discipline</b>	<b>Authors</b>	<b>Average number of authors per text</b>
Health Science (83 texts)	164	2.0
Humanities (156 (texts)	143	.92
Natural Science (80 texts)	185	2.3
Social Science (160 texts)	192	1.2
Total	684	1.4

As suggested by the figures in Table 4.8, the AcaDan corpus comprises a high number of individual authors which helps to avoid idiosyncrasies of one or a few authors influencing the results of analyses carried out via the corpus. Table 4.8 also shows that the tendency to co-author articles in the hard sciences is greater than it is in the soft sciences with 2.3 authors per text in average in Natural Science

and .92 authors per text in the Humanities. The average number of authors in the Humanities reveals that some authors are represented more than once in this sub-corpus.

Turning to the issue of the sources of the 479 texts of the corpus, the 448 articles stem from 183 different journals. The 31 reports are published by 23 different research and government institutions such as the National Board of Health, the Danish Environmental Protection Agency, and the National Research Centre for the Working Environment. Table 4.9 suggests that the disciplines of Health and Natural Sciences draw on a limited number of journals when publishing in Danish in contrast to the two other disciplines which have more opportunities of publishing in Danish (see also Hultgren, 2013).

*Table 4.9. Overview of the number of sources for the texts in the AcaDan corpus*

<b>Discipline</b>	<b>Number of sources</b>
Health Science (69 articles)	24 journals
Health Science (14 reports)	10 research and government institutions
Humanities (156 articles)	83 journals
Natural Science (63 articles)	28 journals
Natural Science (17 reports)	13 research and government institutions
Social Science (160 articles)	48 journals

Table 4.10 shows the earliest and latest publication years for the articles in each of the four disciplines represented in the corpus. Only one percent of the 479 texts was published before the year 2000. I may have been able to retrieve more articles from the Natural and Health Science disciplines had I chosen to search for articles from before 2000. Such articles, however, would not reflect current language usage in the hard sciences, and I would risk giving an outdated depiction of academic language usage in the hard sciences. As discussed earlier in this chapter, the last 15 years have seen a change in the way the hard sciences publish, choosing to focus primarily on English language research publications. Conversely, the softer sciences of Humanities and Social Science still publish to a large degree in Danish, though it is changing especially for Social Science, and the few articles from before 2000 do not in the same way influence the representativeness of the AcaDan corpus, due to the fact that the Humanities and Social Science subsections of the corpus are larger.

*Table 4.10. Time of publication for the texts in the AcaDan corpus*

<b>Discipline</b>	<b>Time of publication</b>
Health Science	2004-2016
Humanities	1997-2015
Natural Science	2006-2016
Social Sciences	1998-2015

#### 4.4.7. Metadata in the AcaDan Corpus

The particulars of the 479 texts given above can be regarded as substantial parts of the metadata of the AcaDan corpus. Metadata can prove essential for the subsequent corpus analyses as it informs the researchers about the context of the investigated language. This is especially the case if the metadata is contained within the corpus as mark-up (McEnery et al., 2006, p. 22). Metadata mark-up comprises annotating each text with information about for example author, source, year of publication, genre, and/or discipline. As such, the texts of the AcaDan corpus could have been annotated with information about which academic disciplines they belong to. It was, however, decided, early on in the process of establishing the AcaDan corpus that the corpus texts would not be marked-up with metadata due to time restraints. The metadata for the AcaDan corpus texts are instead kept in a MS Excel book. A sheet for each discipline is kept listing the matching texts and the relevant metadata for each text including the file name of each text. For all texts, the following metadata were collected:

- Reference, i.e. name of author (s), year of publication, journal title, volume, year, pages
- Discipline and sub-discipline (see Table 4.7)

Discipline and sub-discipline were decided upon by considering in which DNRD research domain the text was listed in, the journal in which the text was published, and the affiliation of the author(s). Each text was named after the assigned discipline and sub-discipline and given a number, e.g. ‘hum\_linguistics\_30’ with ‘hum’ denoting academic discipline, ‘linguistics’ sub-discipline, and ‘30’ signalling that it was the 30<sup>th</sup> article downloaded within the discipline of Humanities. Appendix A provides the references of all 479 texts of the AcaDan corpus.

#### 4.4.8. Copyright of the AcaDan Corpus texts

The issue of copyright is an important issue to consider when compiling a corpus (McEnery et al., 2006, pp. 77–79) and ideally one should ask for permission of copyright from the authors of the texts that comprises a given corpus even if the texts are freely available online. Considering today’s giant

corpora created from the Internet, this, however, may seem an insurmountable task. Even for a relatively small corpus like the AcaDan corpus, it would be a very time-consuming task, which I have chosen not to undertake because of time restraints. The AcaDan corpus is a research corpus designed and compiled for the studies of this thesis alone. The corpus is only available to me, the researcher, and it will not be made publicly accessible in its present form. Any examples from the corpus that I supply in my analyses are referred to appropriately. However, given that the AcaDan corpus is the only one of its kind, for future research projects the issue of copyright in relation to the corpus should be dealt with.

#### 4.4.9. Text preparation, corpus upload, and sub-corpora

Preparing the 479 collected texts for upload to the Sketch Engine platform (see Section 4.4.3) comprised two steps. In the first step, English abstracts, front pages, bibliographies, endnotes, and whole-paged pictures and tables were removed from the texts. However, it was not always possible to remove all of English abstracts and bibliographies (in which English titles occur frequently) which means that the AcaDan corpus does contain some English language. The purpose of the second step was to make the texts ready for analyses when uploaded to the corpus. This step involved a series of procedures outlined below. The CLARIN-DK platform (CLARIN\_DK, n.d.-a) offers a tool (CLARIN\_DK, n.d.-b) that in one workflow can both convert from PDF to raw text format and tokenise, lemmatise and part-of-speech-tag the texts as well as convert them to a readable file format for Sketch Engine. Moreover, this tool can process multiple texts together. The most important steps in relation to making the texts ready for linguistic analyses within the corpus platform are the automatized procedures of tokenisation, lemmatisation, and POS-tagging. While tokenisation allows for the counting of tokens, i.e. occurrences of any given word form (McEnery et al., 2006, p. 350), lemmatisation (the reduction of inflectional forms of a lexical item to a lemma), and POS-tagging (the assignment of part-of-speech to each word) can be considered as linguistic analyses embedded in the corpus data which can be extracted when analysing the corpus data (McEnery et al., 2006, pp. 29–36). The lemmatisation of the AcaDan Corpus texts follows the definition of a lemma given by Bowker and Pearson (2002) and Francis and Kücera (1982, p. 1) among others as “a set of lexical forms having the same stem and belonging to the same major word class differing only in inflection and/or spelling.”

The output format of the CLARIN-DK tool is a vertical file format called CONLL2009 (Institute of Formal and Applied Linguistics, n.d.), which is a readable vertical file format for the corpus platform



Sketch Engine. The verticality of the file format means that each file contains a number of columns to fit each annotation made by the CLARIN-DK tool. This enables analysis on the level of token (word form), lemma, and part of speech in the corpus platform. The 479 files kept the names they were given in the text collection phase.

After the 479 texts in the file format of CONLL2009 were uploaded to Sketch Engine to a corpus template designed to fit the vertical input format of the CONLL2009-files, four sub-corpora were defined according to the four academic disciplines. The size and content of these are given above in Table 4.5. Specifically, the creation of these sub-corpora was carried out using the file names which contain information about discipline and sub-disciplines. This means that it is possible to create sub-corpora based on sub-disciplines as well. In addition, the Sketch Engine platform also allows for defining sub-corpora based on word counts. If the files had been annotated for metadata, even more advanced opportunities for defining sub-corpora would exist.

#### 4.4.10. Summary: Description of the AcaDan Corpus

In the preceding sections, I have introduced and described the decisions and choices made in the development of the AcaDan corpus used for the studies of this thesis. Before I turn to describing the other corpora used, I provide a summary of the development and the content of the AcaDan corpus in Table 4.11.

*Table 4.11. Development and content of the AcaDan corpus*

<b>Language represented by the corpus</b>	<b>Professional written academic Danish from the disciplines of Humanities, Health Science, Natural Science, and Social Science from 1994 to 2016</b>
<b>Text types</b>	Academic journal articles (448) and scientific reports (31)
<b>Size</b>	3,336,988 running words: Humanities (156 texts, 1,091,674) Social Science (160 texts, 1,065,346) Natural Science (80 texts, 585,369) Health Science (83 texts, 595,600)
<b>Authors and sources</b>	684 individual authors and 206 separate sources. The Danish National Research Database was used as the primary resource for collecting texts.
<b>Metadata</b>	No metadata mark-up of the corpus, but references to each text are given in Appendix A
<b>Annotation</b>	Lemmatised, tokenised, and POS-tagged
<b>Access</b>	The corpus is stored in Sketch Engine but has not been made public.
<b>Sub-corpora</b>	The corpus content can be divided into sub-corpora based on disciplines, subject areas, text types, and word count via the Sketch Engine user interface.

#### 4.5. The second academic language corpus

The AcaDan corpus was created with the primary purpose of developing a word list of academic vocabulary (Study 2 of this thesis). The purpose of creating a second academic language corpus was twofold: 1) to be able to measure the lexical coverage of a Danish academic word list as developed in Study 2, and 2) to measure the lexical coverage of general high frequency words (Study 1). As the task of developing an academic language corpus from scratch is rather time consuming, the second academic language corpus consists of only 298,000 running words, and 82 percent of the running words derive from Humanities. This is due to the availability of texts in this discipline. Table 4.12 provides the details of this corpus. It should be noted that the texts from the Humanities are peer reviewed monographs and not journal articles.

Table 4.12. *The Second Academic Language Corpus*

Number of running words	Texts
297,802	7 monographs from Humanities (243,411). 4 journal articles from Social Science (23,732). 4 journal articles from Natural and Health Sciences (30,659).

Since this corpus was only used for lexical coverage, it was not annotated or marked up in any way, and it has not been uploaded to Sketch Engine. As such it consists of text files (.txt). Metadata for the texts are given in Appendix B.

#### 4.6. General language corpora

As outlined in Table 4.1 in this chapter, three general language corpora were used in the studies of this thesis. All three corpora represent written general Danish texts. The two Journalisten.dk corpora were created as sub-corpora in the Danish Ten Ten corpus (Jakubíček et al., 2013), which is a pre-loaded corpus in Sketch Engine. The two Journalisten.dk corpora were created by selecting all texts in the corpus from the website Journalisten.dk, which is the website of the Danish Association of Journalists (Dansk Journalistforbund, n.d.). The Danish Ten Ten corpus was altered and renamed from Danish Web 2014 to Danish Web 2017 by the research group behind it during the preparation of this thesis. This meant that I created the Journalisten.dk sub-corpus twice, once in the Danish Web 2014 and once in the Danish Web 2017 as the 2014 version became unavailable in Sketch Engine. The 2014 version was used in Study 1 and the 2017 version in Study 2. It should be noted that the year stated in the titles of the Danish Web corpora refer to when the corpora were compiled and not to the year of publication of the texts of these two corpora. In Table 4.13, the details of each version are given including their use in the two studies.

Table 4.13. *The Journalisten.dk corpora*

Content	Texts from the website <a href="http://www.journalisten.dk">www.journalisten.dk</a>	
<b>Name of corpus</b>	Sub-corpus in Danish Web 2014	Sub-corpus in Danish Web 2017
<b>Running words</b>	6,207,563	4,587,116
<b>Used in</b>	Study 1	Study 2

The third general language corpus used in this thesis was a corpus of general language that I created on my own. Table 4.14 gives the details of this corpus, which is called The General Language Corpus.

In Appendix C the sources of the texts are given. The purpose of this corpus, as with the second academic language corpus, was to use it for the lexical coverage analyses carried out in Studies 1 and 2. This corpus features newspaper and magazine articles collected manually using the Internet in May 2017 as well as a set of teaching materials for Danish as a second language. The majority of texts are from 2016 and 2017 with a few from 2014 and 2015 and one from 2013. The General Language Corpus is rather limited in size as it only contains about 330,000 running words. Parallel to the Second Academic Language Corpus, this corpus only exists as .txt files and has not been processed in any way.

*Table 4.14. The General Language Corpus*

<b>Content</b>	<b>Professional written general Danish</b> <b>265 texts:</b> <ul style="list-style-type: none"> <li>- <b>240 feature and news articles (269,983)</b></li> <li>- <b>20 leisure magazine articles (35,981)</b></li> <li>- <b>5 sets of teaching materials (23,569)</b></li> </ul>
<b>Running words (total)</b>	329,533
<b>Used in</b>	Study 1 for lexical coverage analysis

## 4.7. Summary

In this chapter, I have described the development of the primary corpus, the AcaDan corpus, used for the studies reported on in this thesis and given descriptions of the other corpora used. Limitations as related to size and composition have been briefly discussed throughout the chapter, but these issues will be revisited in connection with the studies in which the corpora are used.

## Chapter 5. Study 1 – General and academic high frequency vocabulary in Danish<sup>14</sup>

### 5.1. Introduction and research questions

This chapter presents Study 1 of this thesis. The focus of Study 1 was general high frequency vocabulary in written Danish and its relation to academic written language. Study 1 served as an important foundation for the development of a Danish Academic Word List (described in Study 2) by exploring the issue of general high frequency vocabulary in detail. This issue has, as shown in Chapter 3 on word lists, received increasing attention in the development of academic word lists (e.g. Dang et al., 2017; Gardner & Davies, 2014; Hagen et al., 2016; Paquot, 2010), but the precise overlap between general high frequency vocabulary and academic vocabulary has not been studied very extensively. In English, general high frequency vocabulary is commonly defined as the 2,000 most frequently used words, and several lists of general high frequency words exist developed primarily for pedagogical use such as the General Service List (West, 1953), the New General Service List (Brezina & Gablasova, 2015), and the BNC/COCA 2000 (Nation, 2012). In Danish, general high frequency vocabulary has been explored earlier (Bergenholtz, 1992; Ruus, 1995), but these studies aimed to be more linguistically oriented than pedagogical.

Study 1 investigated the 2,000 first items found in a list of the 5,000 most frequently used lemmas in Danish (Det Danske Sprog- og Litteraturselskab, n.d.-d) (henceforth the Top 5,000 list). These first 2,000 lemmas of the Top 5,000 list are thus conceptualised as Danish general high frequency vocabulary as these lemmas are derived from a large, representative corpus of general language use in Danish. An important assumption behind Study 1 was that general high frequency vocabulary is important for both general and academic language use. Thus, the motivation for Study 1 was to increase our knowledge of Danish general high frequency vocabulary in relation to academic language and academic vocabulary. This knowledge can serve as a basis for developing tools, learning materials, and teaching resources to support language and academic learning and teaching. In particular, Study 1 investigated the lexical coverage of Danish general high frequency vocabulary in general language and academic texts, and compared it with the findings for other languages. It also

---

<sup>14</sup> Study 1 is a revised version of Jakobsen, Coxhead, and Henriksen (2018). Co-author declaration can be found in Appendix D.

explored if there are items among the 2,000 most frequent words in Danish that are academic in nature.

This study was guided by the following research questions:

- 1) What is the lexical coverage of the most frequent 2,000 words of Danish in general written language?
- 2) What is the lexical coverage of the most frequent 2,000 words of Danish in academic written language?
- 3) How many of the most frequent 2,000 lexical items in Danish are academic in nature?

The presentation of Study 1 is divided into five sections: The first section reports on the methodology of the study (Section 5.2). This section is followed by a results section answering the three research questions outlined above (Section 5.3). A discussion section compares the findings of Study 1 with previous research and discusses what implications these findings have for our understanding of academic words and for Danish high frequency words in general (Section 5.4). The fourth section discusses the limitations of Study 1 (Section 5.5). The presentation of Study 1 ends with a summary and a section on the rationale for Study 2, which is presented in the next session of this thesis.

## 5.2. Methodology

In this section, I describe the procedures for measuring the lexical coverage of the 2,000 most frequent words in Danish (research questions 1 and 2) and for identifying academic words among these words (research question 3). The section begins with a brief description of the data collected for answering the three research questions.

### 5.2.1. Data

The data for Study 1 consisted of:

- A list of the 5,000 most frequently used lemmas of Danish (Top 5,000)
- Two general language corpora:
  - o Journalisten.dk, a sub-corpus in the Danish Web 2014 corpus, and
  - o The General Language Corpus
- The AcaDan corpus

I refer to Chapter 4 for details on the compilation of the two general corpora and the academic corpus. The list of the 5,000 most frequently used lemmas in Danish was developed by the Danish Society of Language and Literature (henceforth DSL). For developing this frequency-ranked word list, the DSL

used a corpus of 880 million tokens which contains text material collected during the period of 1983-2016 (Det Danske Sprog- og Litteraturselskab, n.d.-d). This corpus is developed for use in the preparation and updating of the Danish Dictionary and is here called the DDO Corpus. The AcaDan corpus was used for measuring the coverage of the most frequent 2,000 words in academic and general language, and for identifying academic vocabulary in the high frequency lists. The General Language Corpus was used to measure the coverage of the most frequent 2,000 words in general language. The Journalisten.dk corpus (a sub-corpus of the DaTenTen corpus from Jakubiček, Kilgarriff, Kovář, Rychlý, & Suchomel, 2013) available through Sketch Engine (Kilgarriff et al., 2014) was used in this study for identifying academic vocabulary among the 2,000 most used lemmas in Danish.

### 5.2.2. Data analysis

In this section, I describe the procedures for developing the first 2,000 lemmas of the most used 5,000 lemmas into a word list to be used in the vocabulary load programme AntWordProfiler (Anthony, 2014) to measure the lexical coverage of the most frequently occurring 2,000 lemmas of Danish. Moreover, I give an account of how academic words were identified among the 2,000 general high frequency items. I begin this section, however, with a brief overview of lexical coverage analysis as applied in this study.

#### 5.2.2.1. *Lexical coverage analysis*

Lexical coverage analysis comprises the measuring of the occurrence of words from different frequency bands in a text or a corpus. These frequency bands are represented by word lists, e.g. the first 2,000 words of English in the form of the General Service List (West, 1953), or other types of vocabulary, e.g. the Academic Word List (Coxhead, 2000). The method was originally developed by Laufer and Nation (1995b) to measure the lexical richness of learner texts and is referred to as lexical frequency profiling, but it can also be used for analysing the lexical load of texts in general and then it is often referred to as vocabulary load analysis. Finally, lexical coverage analysis is used for evaluating word lists, i.e. for measuring the coverage of a list in a corpus (cf. Coxhead, 2000; Dang et al., 2017; Dang & Webb, 2017; Durrant, 2013; Gardner & Davies, 2014; Hagen et al., 2016). In contrast to other lexical profiling measures, such as type-token analysis, which profiles the lexical inventory of a particular text, the LFP framework measures “the frequency of [a text’s] lexis with respect to the language large.” (Cobb & Horst, 2004, p. 25). In particular, a vocabulary load analysis classifies the words of a text in relation to frequency lists representing different frequency levels of

the language. This can be done via computer programmes like Range (Nation et al., 2002), or AntWordProfiler (Anthony, 2014) in which different word lists are preloaded. The rationale behind lexical coverage rests on the notion that frequency is a determining factor in vocabulary use and acquisition, and the use of this measurement is widespread in vocabulary studies, especially within English second and foreign language acquisition and teaching. Lexical coverage analysis can thus be used for measuring how many of the words in a given word list occur in a text or a corpus. The analysis, however, shows both how much the word list covers of a given text or corpus and gives us information about the vocabulary load of the text itself by showing how many and how often e.g. the General Service List words are used in an academic text. Thus, in Study 1, lexical coverage analysis is not only used to show how a word list of the 2,000 most frequently used words of Danish cover different text types. It is also used to show how different text types make use of general high frequency vocabulary.

Most studies employing lexical coverage analysis as described above have investigated English vocabulary, and the aforementioned programmes are all developed for English vocabulary studies even though they all can be used with other language too if the necessary resources are available (see Bardel, 2016 for an overview of the use of LFP in other languages than English). The necessary resources are frequency-based word lists representing e.g. high-, mid-, and low- frequency vocabulary. The unit of counting for these word lists is typically either word families or lemmas. The word family framework (Bauer & Nation, 1993) was developed to make the creation of reliable frequency-based word lists possible for use in vocabulary tests and in vocabulary load analysis programmes, but these programmes can just as well be run with lemma-based word lists (cf. Cobb & Horst, 2004). As discussed in the literature review of this thesis, only a few studies employing lexical coverage analysis in other language than English exist (e.g. Bardel & Lindqvist, 2011; Cobb & Horst, 2004), and the framework has not been employed in studies on Danish vocabulary so far. Other measures for determining the lexical richness or variability of a particular text exist, whereof the Type-Token Ratio measure is perhaps the most widely used independent of language. Unlike lexical coverage analysis, the type-token ratio measure is dependent on text length as it measures the occurrence of different words (types) in relation to a total number of words in a text. This limitation of the Type-Token Ratio and other limitations of other methods for measuring lexical richness lead Laufer and Nation (1995b) to argue that lexical frequency profiling is the most reliable method for measuring the use of vocabulary in a text. However, some points of critique of lexical coverage analysis can be raised especially related to using lexical coverage analysis for assessing learner texts



and by proxy learner vocabulary knowledge. Firstly, knowledge of one member of a word family or a lemma does not necessarily entail knowledge of the whole word family or lemma. So when one member of a lemma or a word family is shown to be used in a text, it does not necessarily mean that the writer of the text, i.e. the learner, knows all the members of the lemma or word family (Gardner, 2007; Treffers-Daller, Parslow, & Williams, 2018). Secondly, the measure only measures the use of a word, not whether that use is in fact correct or whether the word has different senses. As such, lexical coverage analysis is not an accurate measure of knowledge of word meaning, nor can it account for semantic variation or appropriate language use. On the other hand, via the use of word lists representative of general language or of different language types, lexical coverage analysis can provide us with knowledge of how different text types make use of vocabulary.

Having so far described lexical coverage analysis as a measure to be used for not only evaluating a word list's coverage across different text types but also for profiling the vocabulary of a given text, I will now move on to describing the measuring of the lexical coverage of Danish general high frequency word lists.

#### *5.2.2.2. Measuring the lexical coverage of the 2,000 list (working title)*

AntWordProfiler (Anthony, 2014) was used for analysing the lexical coverage of Danish general high frequency vocabulary in the form of the 2,000 most frequently used lemmas in Danish in academic language and in general language. AntWordProfiler handles the Danish letters of æ, ø and å without any difficulty. In the next section, I describe the development of two so-called base word lists to be used in the AntWordProfiler programme for lexical coverage analysis.

#### *5.2.2.3. Developing the Danish general high frequency word lists*

Det Danske Sprog-og Litteraturselskab/Danish Society of Language and Literature (henceforth DSL) has developed a list of the most used lemmas in Danish, extracted from the DSL corpus described above (Det Danske Sprog- og Litteraturselskab, n.d.-a). The lemma list, ranked according to relative frequency, now comprises the most frequent 10,000 lemmas of Danish, but when the present study was carried out, it comprised 5,000 lemmas. I refer to Chapter 3 for more details on the 10,000 list. The 5,000 lemma list, which only contained the baseform of the lemmas, had two versions: *brutto* and *netto*. Brutto contained proper nouns and numerals ('Top 5,000 B') and netto did not contain proper nouns and numerals ('Top 5,000 N'). The list without proper nouns and numerals, Top 5,000 N, was used to develop the general high frequency list. The aim was to create a list of 2,000 lemmas (baseform plus inflections) representing the general high frequency vocabulary of Danish to be used

for lexical coverage analysis as described above. In order to do so, the first 2,000 lemmas of the Top 5,000 N list needed to be expanded to include inflections. To guide this process, a list of inflections of 80,000 lemmas was used (Det Danske Sprog- og Litteraturselskab, n.d.-b). The 80,000 inflections are automatically derived from various lexical sources. The inflections of each lemma were added manually, and in order to make it readable for the vocabulary load analysis programme, a space and a zero were added to each baseform and its inflections as depicted in Figure 5.1.

VÆRE 0	HAVE 0
ER 0	HAR 0
VÆRES 0	HAVES 0
VAR 0	HAVDE 0
VÆRENDE 0	HAVDES 0
VÆRENDES 0	HAVENDE 0
VÆRET 0	HAVENDES 0
VÆRETS 0	HAFT 0
VÆR 0	#HAV 0

Figure 5.1. The lemmas *være* (to be) and *have* (to have)

Since the programme cannot count repeated items and therefore marks them as errors, repeated items in the list needed to be marked with a hashtag for the programme to ignore them. An example of an item in the list occurring more than once is the lemma *få* (get) which contains the same form twice, first as the baseform itself and in the imperative. To complicate matters further, the form *få* is a homonym as it can mean both ‘get’ and ‘few’, and it also occurs in the latter meaning within the first 1,000 lemmas of Danish. This occurrence was marked with a hashtag as it is less frequent than the ‘get’ meaning based on its rank in the Top 5,000 N list. Marking repeated items is a useful step for a later analysis of homoforms; an important step in word list development which awaits automation (Nation, 2016). To be able to mark all repeated items, a list with the first 2,000 lemmas including inflections was uploaded as a base word list to AntWordProfiler (Anthony, 2014). A lexical coverage analysis was run over a text, and the programme returned a list of repeated items as an error message. This list was used as a basis for marking repeated forms with hashtags. The decision of which forms to make unreadable was based on frequency. That is, the less frequent repeated form was marked with a hashtag. For example, the lemma *for* occurs three times in the top 5,000 list as different parts of speech (preposition, conjunction, and adverb). As a preposition, *for* occurs within the first 20 items on the list with a rank of 13. As a conjunction and as an adverb, it has a rank of 126 and 471, respectively. These two latter occurrences were made unreadable by hashtags. The removal of repeated forms meant that it was necessary to add more lemmas to the list to ensure that it comprised a total number of 2,000 lemmas. In total, 2,058 lemmas from the Top 5,000 N list were used for the

base word lists and 58 items were marked unreadable by hashtags. Once the general high frequency list contained in total 2,000 lemmas and their inflections, it was then divided into two lists to allow for analysing which frequency belt (the first most frequent 1,000 lemmas or the second most frequent 1,000 lemmas) a word belongs to. The two lists were saved as .txt files with UTF-8 encoding and named Dan1st1000 (the first 1,000 most frequent lemmas) and Dan2nd1000 (the second most frequent 1,000 lemmas).

Table 5.1 contains a summary of the specifications of the general high frequency list of Danish comprised by the two lists Dan1st1000 (the first 1,000 most frequent lemmas) and Dan2nd1000 (the second most frequent 1,000 lemmas), based on Nation's own critique of his BNC/COCA lists (2016) and a critique of the Danish lists.

*Table 5.1. Specifications for the lists of the 2,000 most frequently used words of Danish*

<b>Focus</b>	<b>The 2,000 most frequent words of Danish</b>
<b>Purpose</b>	Course design for Danish as a second or foreign language; research on lexical coverage of Danish texts
<b>Unit of counting</b>	Lemmas
<b>Corpus</b>	The DDO Corpus (880 million tokens)
<b>Main word lists</b>	Two frequency-based word lists
<b>Other lists</b>	No other lists were used, such as proper nouns, in the analysis
<b>Making the list</b>	Developed from the Danish Society of Language and Literature's (DSL) top 5,000 lemma lists; Used AntWordProfiler to troubleshoot the high frequency lists
<b>Possible criticisms</b>	Based on primarily written Danish, not developed from scratch, little knowledge of the corpus on which it is based
<b>Function words</b>	Included in the list
<b>Homoforms</b>	Identified but not taken into account in the lexical coverage reporting
<b>Size of the lists</b>	1,000 lemmas per list; around 7,000 types per list
<b>Other features</b>	402 lemmas in the list are identified as academic in nature

#### *5.2.2.4. Identifying academic words in Danish general high frequency vocabulary*

For identifying the proportion of academic words in the high frequency lists (research question 3), a quantitative approach was taken using measures of ratio, range and dispersion. These measures are also used in academic vocabulary list studies such as Gardner and Davies (2014) and Hagen et al. (2016) and are standard criteria for selecting academic vocabulary (see also Chapter 3). The basic principle of the ratio measure, which is in fact a corpus comparison, is that academic vocabulary will

occur more often in an academic text than in a general text. Range ensures that academic vocabulary occurs across disciplines, whereas dispersion ensures that the vocabulary is evenly distributed throughout the corpus. For the present study, frequencies per million were used to allow for comparison between corpora of different sizes. The AcaDan corpus representing academic language and a sub-corpus of the DaTenTen corpus, Journalisten.dk, representing general language use (see Table 5.1) were used. As described in Chapter 4, both the AcaDan Corpus and the Journalisten.dk Corpus are stored in Sketch Engine (Kilgarriff et al., 2014, 2004) and the word list tools of this software were used for extracting frequency counts used in the measures of ratio, range, and dispersion.

Experimentation with different ratio values from 1.5 to 2.6 suggested that the 1.5 ratio would best fit. Range was measured by checking the ratio value across all three sub-corpora representing Humanities, Social Science and Natural and Health Sciences. Dispersion was measured using Juilland's D (Juilland & Chang-Rodríguez, 1964). Experimentation with different dispersion values showed that a cut-off value of 0.60 captured best a range of words with apparent academic senses. A more thorough presentation of these measures can be found in Chapter 5 in the presentation of Study 2.

### 5.3. Results

In response to research question 1, Table 5.2 shows the coverage of the most frequent 2,000 words of Danish in the general language corpus and in different text types within the general language corpus. The 1,000 most frequent words of Danish covered over two thirds of the words in the general corpus (76.31%). The second 1,000-word list has a much lower coverage of almost 6%. This coverage pattern of the first 1,000 covering a much higher number than the second 1,000 is very similar to general word lists in English such as West's General Service List (1953; see Nation, 2013 for more).

Table 5.2. Coverage of the first and second 1,000-word lists of Danish in general language

<b>List</b>	<b>Language type</b>	<b>Text types</b>			
	<i>General language</i>	<i>Magazine articles</i>	<i>Feature articles</i>	<i>Newspaper articles</i>	<i>Teaching material</i>
<b>First 1,000</b>	76.31%	80.35%	77.48%	75.81%	73.45%
<b>Second 1,000</b>	5.90%	5.42%	6.07%	5.87%	6.18%
<b>First and second combined</b>	82.21%	85.77%	83.55%	81.68%	79.64%
<b>off-list</b>	17.79%	14.23%	16.45%	18.33%	20.30%
<b>Total coverage</b>	100%	100%	100%	100.01%	99.93%

When we look at the most frequent 2,000 words' coverage in different general language text types, we see that the coverage of the first 1,000 words in newspaper feature articles is 77.48% and the second 1,000 cover 6.07%, making a combined coverage of 83.55%. For magazine articles, the coverage is somewhat higher than that of newspaper articles and feature articles. Here, the first 1,000 words cover 80.35%, but the coverage of the second 1,000 words is similar to that in newspaper articles and feature articles. The words in the off-list are mid- and low frequency words including proper names.

For research question 2, as with the general corpus results, the first 1,000-word list of Danish covered a much higher percentage of the AcaDan corpus than the second 1,000-word list in academic texts. Table 5.3 provides the coverage figures of the academic corpus. The high frequency word lists provide a combined coverage of 70%, which leaves around 30% of the texts not covered by these lists. This large group of words needs further investigation and categorisation. These words are likely to include mid- and low frequency lexis, proper nouns, academic and technical words, and abbreviations.

Table 5.3. Coverage of the first and second 1,000-word lists of Danish over academic language

<b>List</b>	<b>Language type</b>	<b>Disciplines</b>			
	<i>Academic language</i>	<i>Social Science</i>	<i>Humanities</i>	<i>Natural Science</i>	<i>Health Science</i>
<b>First 1,000</b>	63.04%	66.5%	62.79%	60.71%	59.19%
<b>Second 1,000</b>	7.31%	7.79%	6.99%	6.97%	7.34%
<b>First and second combined</b>	70.35%	74.29%	69.78%	67.68%	66.53%
<b>Off-list</b>	29.65%	25.7%	30.22%	32.32%	33.48%
<b>Total coverage</b>	100%	99.99%	100%	100%	100.01%

The coverage figures for the four academic disciplines represented in the corpus (also in Table 5.3), differ noticeably across the disciplines. The highest coverage is in texts from Social Science (74.29%) while Health Science texts make less use of the first 2,000 words of Danish (66.53%). The off-lists figures suggest that Health and Natural Sciences make more use of a specialised, technical vocabulary than the disciplines of Humanities and Social Science.

To explore the relationship between Danish general high frequency vocabulary and academic language further, texts from the sub-disciplines of Medicine (Health Science) and Information Science (Social Science) were run through the vocabulary load analysis programme, AntWordProfiler (Anthony, 2014) for coverage figures (see Tables 5.4 and 5.5). This analysis highlighted the disciplinary differences in relation to the general high frequency word coverage, and words outside the most frequently used 2,000 words, as shown in the results for research question 2, are further highlighted. The most frequently used 2,000 words provide on average higher coverage in Information Science (73.46%) than in the academic corpus (70.35%) and Medicine (55.96%). The Medicine text contains more words outside the most frequently used 2,000 words than the Information Science text, i.e. a remarkably low coverage of the Danish general high frequency words are found in Medicine. Figure 2 shows excerpts from a Medicine article and an Information Science article

illustrating how the two sub-disciplines make use of the general high frequency words. As clearly illustrated, general high frequency items are used more extensively in the Information Science article.

*Table 5.4. Coverage and means of the first and second 1,000 word lists of Danish over Medical research articles (M)*

<b>Text (Medicine)</b>	<b>First 1,000 words</b>	<b>Second 1,000 words</b>	<b>First+second 1,000 words</b>	<b>Outside the first 2,000 lists</b>
<b>M1</b>	45.81	6.56	52.37	47.63
<b>M2</b>	40.38	6.17	46.55	53.44
<b>M3</b>	58.1	7.80	65.9	34.10
<b>M4</b>	38.51	8.28	46.79	53.22
<b>M5</b>	57.28	9.58	66.86	33.14
<b>M6</b>	40.17	4.27	44.44	55.56
<b>M7</b>	55.35	5.4	60.75	39.25
<b>M8</b>	51.62	6.12	57.74	42.26
<b>M9</b>	49.04	5.17	54.21	45.80
<b>M10</b>	55.8	8.14	63.94	36.06
<b>M</b>	49.21	6.75	55.96	44.05

*Table 5.5. Coverage and means of first and second 1,000 word lists of Danish over Information science research articles (IS)*

<b>Text (Information Science)</b>	<b>First 1,000 words</b>	<b>Second 1,000 words</b>	<b>First+second 1,000 words</b>	<b>Outside the first 2,000 lists</b>
<b>IS 1</b>	60.62	9.89	70.51	29.49
<b>IS 2</b>	63.85	9.07	72.92	27.09
<b>IS 3</b>	65.09	9.55	74.64	25.37
<b>IS 4</b>	59.94	11.80	71.74	28.26
<b>IS 5</b>	61.94	8.31	70.25	29.75
<b>IS 6</b>	70.09	7.88	77.97	22.03
<b>IS 7</b>	65.89	8.13	74.02	25.98
<b>IS 8</b>	63.82	7.90	71.72	28.28
<b>IS 9</b>	65.72	7.24	72.96	27.04
<b>IS 10</b>	69.70	8.16	77.86	22.14
<b>M</b>	64.67	8.79	73.46	26.54

The bold words are general high frequency words and the italicised are words outside the 2,000 most frequently used lemmas in Danish.

#### Medicine article

**Analyserne vil undersøge 1) de uafhængige effekter af og samspillet mellem sociale, kognitive og helbredsmæssige faktorer tidligere i livet på inflammation midt i livet og 2) den akkumulerede effekt af disse forskellige faktorer på inflammation midt i livet. Analyserne vil blive baseret på longitudinelle opfølgningsstudier, hvor information om individer er indsamlet på forskellige tidspunkter. Dette betyder, at det komplekse samspil mellem biologiske, psykiske og sociale variabler kan blive baseret på individuelle målinger og det vanskelige spørgsmål om årsagsretning vil blive støttet af den kronologiske rækkefølge af begivenhederne. Styrken ved de eksisterende dataer, at en stor del af de relevante prædikator- og udfaldsvariabler er målt på forskellige tidspunkter.**

English translation: **The analyses will investigate 1) the independent effects of and the correlation of social, cognitive and health-related factors earlier in life on inflammation mid-life, and 2) the accumulated effect of these different factors on mid-life inflammation. The analyses will be based on longitudinal follow-up studies in which information about individuals has been collected at different times. This means that the complex interplay between biological, mental, and social variables can be based on individual measurements and the difficult question regarding the direction of cause will be supported by the chronological sequence of events. The strength of the existing data is that a large part of the relevant predictor and result variables have been measured at different times.**

Source: Avlund, A. K., Bruunsgaard, H., Christensen, U., Fiehn, N. E., & Marie, Å. (2009). CAMB-Copenhagen Aging and Midlife Biobank. Perspektiver for fremtidig forskning. *Miljø og sundhed*, 15(suppl. nr. 1) 81-88.

#### Information science article

**Det vil være forkert her ikke at nævne med den store kultur- og fritidsaktivitetsundersøgelse (Bille et al., 2004). Den ligger ganske vist før den tidsmæssige afgrænsning, og biblioteksbenyttelsen udgør kun en mindre del af den. Der er imidlertid flere gode grunde til at medtage den. For det første er det en undersøgelse med videnskabelige ambitioner, som tilmed indfries. For det andet er det en metodisk meget velfungerende og velfunderet undersøgelse, der også indeholder en del avancerede analyser. For det tredje er der ikke mange nationale undersøgelser på markedet. Endelig er undersøgelsen én af en længere række af kultur- og fritidsundersøgelser, hvilket muliggør sammenligninger over tid.**

English translation: **It would be wrong here not to mention the large culture and leisure time study (Bille et al., 2004). It is true that it lies before the temporal delimitation, and the library use also comprises a smaller part of it. There are, however, several good reasons to include it. Firstly, it is a study with scientific ambitions that are even honoured. Secondly, it is a methodologically very well-functioning and well-founded study that contains a great deal of advanced analyses. Thirdly, there are few national studies available. Finally, the study is one out of a number of culture and leisure time studies which makes it possible to make comparisons over time.**

Source: Pors, N. O. (2011). Evidens om bibliotekernes brugere. *Dansk Biblioteksforskning*, 6(2/3), 65-81.

Figure 5.2. Text excerpts from Medicine and Information Science



Turning now to research question 3 (How many of the most frequent 2,000 lexical items in Danish are academic in nature?), a total of 574 lemmas met the ratio value of 1.5 or higher in the academic sub-corpus in comparison with the general Danish sub-corpus. A total of 569 lemmas met the range criterion (occurrence in all three sub-corpora). Four hundred lemmas met the dispersion value of 0.60. Examples of 20 of these 400 academic lemmas can be found in Table 5.6, with the Danish words translated into English. Items from the 400 academic lemmas that have the same form in Danish and English include *interview*, *central*, and *element*. Hagen et al. (2016), in their development of the Norwegian academic word list, used a cut-off value of 0.60, and it seems this value is reasonable for Danish as well. A more conservative set cut-off value of 0.80 as applied by Gardner and Davies (2014) would reduce the proportion of Danish academic words among the most frequent 2,000 lemmas to 189 lemmas and exclude words such as *begreb* (*concept*), *argument* (*argument*), and *fænomen* (*phenomenon*). These are all lexical items that we may expect to meet in academic texts. Compared to the Academic Vocabulary List (Gardner & Davies, 2014), which contains 3,000 lemmas, the results from the present study suggest that including high frequency vocabulary in the development of a Danish academic word list would produce a list with many more items than e.g. Coxhead's Academic Word List (2000), which contains 570 word families, or the Swedish or Norwegian academic word lists, which contain 652 and 750 lemmas respectively (Ribeck et al., 2014; Hagen et al., 2016). A qualitative approach (e.g. looking at the context of the words) is needed along with the quantitative identification of lemmas to ensure that words among the 400 lemmas that are discipline-specific or purely general are treated as such and not as general academic words. Moreover, two lemmas cover more than one part of speech due to the way the lemmas were identified in the AcaDan Corpus. When extracting lists of lemmas via the tools in Sketch Engine, the tools do not distinguish between parts of speech. This means that lexical items with more than one part of speech are extracted as one item, and as such the lemmas extracted for the DAWL are in fact flemmas (Pinchbeck, 2014 in Nation, 2016, p. 26), sets of lexical items sharing the same stem but with different parts of speech. The lemmas with more than one part of speech are listed as separate items and comprise the following items:

- The lemma *alternativ* (alternative) was identified as both a noun and as an adjective.
- The lemma *følge* was identified as both a noun (sequence) and as a verb (follow).

Consequently, the number of academic words among the general high frequency vocabulary is 402, and I will refer to this number in the remainder of this study. There are also other lemmas that could

potentially occur in more than one part of speech, but they occur in the alternative parts of speech with frequencies so low that I decided to list them only once and give the most frequent part of speech in Appendix E, which contains the 402 academic lemmas in the 2,000 word general high frequency list. Figure 5.3 contains examples of academic texts in which some of these academic lemmas occur.

Table 5.6. Twenty high frequency academic words in Danish

Danish academic word	English translation	Part of speech
<i>analyse</i>	analysis	noun
<i>basere</i>	base	verb
<i>begreb</i>	concept	noun
<i>dialog</i>	dialogue	noun
<i>effekt</i>	effect	noun
<i>eksempelvis</i>	as an example	adverb
<i>fænomen</i>	phenomenon	noun
<i>felt</i>	field	noun
<i>fokus</i>	focus	noun
<i>forudsætning</i>	assumption	noun
<i>gennemsnit</i>	mean	noun
<i>henholdsvis</i>	respectively	adverb
<i>ligeledes</i>	likewise	adverb
<i>perspektiv</i>	perspective	noun
<i>praksis</i>	practice	noun
<i>resultat</i>	result	noun
<i>sammenligne</i>	compare	verb
<i>tendens</i>	tendency	noun
<i>undersøge</i>	investigate	verb
<i>vurdere</i>	assess	verb

The words in bold italics are high frequency words that are also academic in nature.

#### Humanities

Det var ***tydeligt***, at den daværende chefredaktør Philippe Val var opsat på at engagere bladet i en ***afgørende*** ideologisk kamp for ytringsfriheden og retten til religionskritik. Ytringsfrihedens grænser i Danmark ***såvel*** som i Frankrig er defineret af lovgivningen. Man kan sige det meste, men der er ***enkelte*** undtagelser (injurier, racisme, opfordringer til ***vold*** osv.), men det korte af det lange er, at det er juraen, der bestemmer. I USA er ytringsfriheden uindskrænket, og man kan i ***princippet*** sige, hvad man vil.

English translation: It was ***clear*** that the then editor in chief Philippe Val was determined to engage the magazine in a ***decisively*** ideological battle for freedom of speech and the right to criticise religion. The limits of freedom of speech in Denmark ***as*** in France are defined by the legislation. You can say nearly everything, but there are a ***few*** exceptions (defamation, racism, call for ***violence***, etc.), but essentially it is the law that decides. In the U.S., the freedom of speech is absolute, and in ***principle*** you can say whatever you want.

Source: Boisen, J. (2015). "Vær frie – det er en ordre!" Verden vs. Charlie Hebdo. Fransk nyt, (268), 32-39.

#### Natural Science

Et problem er *desuden*, at ternekolonierne af *forskellige årsager* jævnligt flytter rundt mellem *forskellige* lokaliteter, *således* at en ny lokalitet, der ikke tidligere er blevet *betragtet* som *væsentlig* (og derfor er uden adgangsforbud), på et tidspunkt kan være en vigtig yngleplads, *eventuelt* blot i et eller nogle få år. Dette *forhold* betyder, at det vil være af *værdi* at sikre adgangsforbud i yngletiden på en *række* af de lokaliteter, som *udgør* potentielle ynglepladser men ikke *nødvendigvis aktuelt benyttes* som yngleplads.

English translation: *Moreover*, it is a problem that the colonies of terns for *different reasons* frequently move around between *different* locations *so* that a new location that earlier was not *considered* as *essential* (and therefore is without access ban) at some point can become an important breeding place, *perhaps* only in one or a few years. This *issue* entails that it will be *worth* while ensuring access bans in the breeding period in a *number* of locations that *comprise* potential breeding places, but which are not *necessarily* being *used* as breeding place *currently*.

Source: Bregnballe, T., & Jørgensen, H.E. (2013). Udvikling i ynglebestanden af Fjordterne i Danmark 1970-2012. Dansk Orn. Foren. Tidsskr, 107, 261-280.

#### Health Science

Figur 1 *viser*, at smerteintensitet for sårsmarter *umiddelbart* ( $p < 0,001$ ) og *generelt* under indlæggelsen ( $p < 0,001$ ) er signifikant *negativ* korreleret med *kvindens* alder – jo lavere alder, *desto* stærkere smerte og vice versa. *Undersøgelsen viser*, at efter at patienterne var kommet hjem, oplevede hhv. 25 % og 27 % af patienterne oplevede sårsmarter og smerter *ved bevægelse* målt på NRS 4+. Smerteintensitet *ved bevægelse* var *ligeledes* signifikant *negativt* korreleret til patientens alder ( $p = 0,048$ ). Tidligere *studier* viser *ligeledes* en *negativt* korreleret *sammenhæng mellem* alder og postoperativ smerte efter mastektomi (post mastectomy pain syndrome) (25).

English translation: Figure 1 *shows* that for pain intensity *immediately* ( $p < 0.001$ ) and *generally* during admission ( $p < 0.001$ ) is significantly *negatively* correlated with the age of the *woman* – *the* lower the age, the stronger the pain and vice versa. The *study shows* that after the patients were discharged, 25% and 27%, respectively, of the patients experiences wound pains and pains *when moving* measured at NRS 4+. Likewise, pain intensity when moving was significantly negatively correlated with the age of the patient ( $p = 0.048$ ). Earlier studies *also show* a *negatively* correlated *relation between* age and post-operative pain after mastectomy (post mastectomy pain syndrome) (25).

Source: From Rud, K., Egerod, I. E., & Brodersen, J. (2014). Patientoplevelse af accelererede brystkræftoperationer belyst ved spørgeskemaundersøgelse. Klinisk Sygepleje, 28 (1), 46-62.

Figure 5.3. Three examples of the 402 academic high frequency lemmas in Danish research articles

## 5.4. Discussion

The findings of Study 1 provide us with new knowledge about general high frequency vocabulary in general and academic written Danish. The lexical coverage findings of Study 1 shed light on the proportion of general high frequency vocabulary in general written Danish texts such as feature articles, newspapers, magazine articles, and teaching materials, as well as in written academic texts in the four academic disciplines of Humanities, Social Science, Natural Science, and Health Science.

Comparisons with research in English and French suggest that Danish high frequency vocabulary has similar patterns of coverage in general texts, but not in medical texts. These findings further our understanding of both Danish academic language and of Danish general vocabulary. In addition, the results of this study will be useful for second and foreign teachers and learners of Danish for general and academic purposes as they set goals for learning. They will also be useful for course and materials designers (both on paper and online) as they develop and evaluate resources for learning. Another finding in Study 1 was that 402 lemmas met the criteria for both high frequency and academic vocabulary, i.e. 20.1 % of the 2,000 high frequency lemmas investigated. These words thus fall into overlap zone I. in the vocabulary circle described in Chapter 2. This finding is important because few studies have investigated this overlap between general and academic vocabulary. Rather, they have focused on developing academic word lists by excluding the high frequency words or by using methods to include some high frequency words. As such, the results of this study emphasise the importance of carefully considering the issue of general high frequency vocabulary in the development of academic word lists and in research on academic language. Moreover, the results highlight the need to further investigate the issue of polysemy in high frequency Danish vocabulary, especially in relation to the 402 lemmas that have been identified in Study 1 as having both a general and an academic function, e.g. through a thorough study of concordance data.

As described in Chapter 3, Danish general high frequency vocabulary has been investigated prior to this study. Both Bergenholtz' frequency dictionary (1992) and the DSL lists (Det Danske Sprog- og Litteraturselskab, n.d.-d) shed light on which words occur most frequently in Danish. Ruus' (1995) study on Danish core vocabulary contributed to an even more thorough understanding of the general words of Danish and their relationships to each other through analyses of antonymy, hyponymy, and meronymy. As such, Study 1 continues this research focus and furthers our knowledge of Danish general high frequency vocabulary by moving beyond the words themselves, and exploring their occurrences in different text types using lexical coverage as the central method of analysis. In the following, I will discuss the results of Study 1 in comparison with English and French results with a particular focus on the pedagogical implications of the findings of Study 1.

Clearly, general high frequency vocabulary in Danish, as in other languages, covers a large proportion of the vocabulary of both general and academic language texts (76.31% and 63.04%, respectively). Further, the first 1,000 words cover a much larger amount of text than the second 1,000 words (which cover under 10% irrespective of language type), just as they do in English (cf. Nation, 2016). The coverage of the general high frequency word lists in Danish in general texts is less (76.31%) than

Cobb and Horst's (2004) coverage of first 2,000 words of French (86.24%) in newspaper texts. Cobb and Horst (2004) also measured the coverage of their French high frequency lists in another type of general language texts and compared it with the English high frequency lists' coverage of the same type of texts. The French high frequency coverage (83.88%) was higher than that of English (81.26%). It seems that the 2,000 most frequently used words in Danish provide less coverage in general language compared to French and English.

In relation to the coverage of general high frequency vocabulary in academic texts, the results showed that there is a noticeable difference between general and academic written Danish. This difference in lexical coverage of high frequency vocabulary between general and academic written Danish is even more noticeable when comparing the results with Cobb and Horst's (2004) results for French, which showed similar coverage numbers of the first 2,000 words of French in academic and general language (81.27% and 83.88%, respectively). The first and second 1,000 word lists coverage in Danish Humanities texts shows lower coverage (69.78%) than West's General Service List (1953) in English Humanities texts (77.4%), as shown by Coxhead (2000). The same pattern can be seen in Health Science (roughly 67%), and in Natural Science (approximately 68%) as compared to 70.7% in English (Coxhead, 2000). Further comparison with French shows that the first 2,000 words of French offer a coverage of 81.27% (Cobb & Horst, 2004, p. 31). Cobb and Horst (2004, p. 33) argue their results indicate that a high number of French high frequency words carry both academic and general senses, and together with the lexical coverage results for French outline above, they argue that there is no need to develop a French academic word list similar to Coxhead's Academic Word List. An alternative interpretation of their findings, however, could be that precisely because of the polysemy of French general high frequency vocabulary, it would be fruitful to further investigate the academic functions of French general words. This could be done by developing an academic word list that includes the first 2,000 words of French in line with what e.g. Gardner and Davies (cf. Gardner & Davies, 2014) did. In line with Cobb and Horst's argument, the results from Study 1's research questions 1 and 2 could be interpreted to suggest that in selecting words for a Danish academic word list, the most frequently used 2,000 words could be excluded. However, the findings in relation to research question 3 suggest that around 20 percent of the words among the first 2,000 words of Danish fulfil frequency-based criteria for being academic. Many of these 402 high frequency lemmas may be polysemous like the French general words, i.e. they may carry both a general and an academic meaning. For both L1 and L2 students in Danish higher education, these 402 lemmas may be worth focusing on, as the academic meaning of these general words may not be known to them (cf. Knudsen,

2009) and may lead to confusion or misunderstandings. However, these 402 words need to be analysed further with respect to their academic meaning before teaching them as academic vocabulary. Research using corpora from different disciplines and levels of education could shed light on the frequency of the general high frequency vocabulary, and whether the academic words part of the high frequency words occur in similar proportions or with similar frequencies in those texts, for example, in secondary school texts. Further research could also investigate the use of Danish high frequency and academic vocabulary in writing, such as Malmström, Pecorari, and Shaw's (2016) study of Swedish postgraduate students' use of Academic Vocabulary List (Gardner & Davies, 2014) words in their academic writing, or in speaking (Dang et al., 2017; Coxhead, Yen Dang, & Mukai, 2017).

The results of Study 1 also illustrate the importance of the most frequent 1,000 words of Danish in terms of coverage in general texts, academic texts, and disciplinary specific texts. In each analysis, these first 1,000 words cover the largest proportion of vocabulary in the texts, followed by the second 1,000 words. This coverage pattern is similar to English (see Nation, 2013; Coxhead, 2000). The coverage of the Danish first 1,000 in Medicine (49.21%) is similar to the 51.96% coverage of the first 1,000 of Nation's BNC/COCA lists (2012) in medical textbooks in English (Quero & Coxhead, 2018). That said, Quero and Coxhead (2018) reported that the second 1,000-word list in English covered 11.68%, giving a combined first and second 1,000 coverage of 63.64%. This means that the second 1,000 BNC/COCA list in English had higher coverage in medical texts than the Medicine coverage reported here in Danish. When comparing the Medicine coverage results with Information Science, it shows that the coverage of the first 1,000 words of Danish cover just over 15% more in Information Science. The coverage of the second 1,000 words of Danish is higher in Information Science than over Medicine. Taken together, these results demonstrate disciplinary differences in coverage of the high frequency words in Danish, just as Coxhead (2000) demonstrated in English. From a pedagogical perspective, this finding suggests that learners in these fields will possibly need to focus on discipline-specific vocabulary earlier on than learners in Humanities. Quero and Coxhead (2018) suggest that the first 3,000 words of English are important for students who are learning English for academic purposes as well as students of medical English. However, it appears that the same recommendation might not hold in Danish for medical purposes. Vocabulary and discourse analyses of texts from different academic disciplines might highlight differences in lexical use by writers in these fields, and much more research in this area is needed.

Nation (2016, p. 5) uses the ‘cost/benefit principle’ in relation to general high frequency vocabulary, stating ‘that learners should get the best return for their learning effort’. Based on this principle, the most frequent 1,000 words should be the main target for any learners of Danish. The general high frequency list of Danish will be useful for, amongst others, language learners, teachers, materials and course designers, learner dictionary publishers, and language learning website designers. That said, it is clear that this general high frequency word list is too large for learners and teachers to work with effectively in short language courses. Dang and Webb (2016) suggest that 800 words is a more realistic goal for language learners. The first 500 should be a useful goal for beginner learners, whether in Danish as a second or foreign language, and could be used to find out what vocabulary learners already know before a course begins. Moreover, the high frequency words of Danish could be used to develop vocabulary tests in Danish, as they have been in English (see Nation, 2016; Nation & Coxhead, 2014; Nation et al., 2016). These tests could then be used to find out more about the knowledge of Danish high frequency vocabulary of second and foreign language learners. Finally, the high frequency word lists could be consulted for developing graded readers in Danish for second and foreign language learners, such as those which can be found in English. Graded readers based on high frequency vocabulary would be more suitable for learners with small vocabulary sizes in Danish than, for example, children’s literature, which has been reported to have vastly different vocabulary loads, from over 4,000 word families in English (see McQuillan, 2016) for popular series, and approaching the vocabulary needed to read adult fiction (S. Webb & Macalister, 2013) for School Journals in New Zealand.

## 5.5. Limitations

Study 1 had a number of limitations. First of all, the extraction of lemmas from an existing list carries a risk that any errors in the initial study will have an effect on the current study. Secondly, this study was primarily based on written corpora, whereas Nation (2012) used spoken corpora to develop the first 2,000 words of the BNC/COCA lists. Thirdly, use of equal-sized corpora for the coverage analyses to answer research questions 1 and 2 would have added to the reliability and validity of the results, but it is difficult to obtain general language corpora with full text access in Danish, which is a serious limitation for replication (Miller & Biber, 2015). Finally, using the Juilland’s D for dispersion in the identification of academic words may be a limitation, because Biber, Reppen, Schnur, and Ghanem (2016) found it decreased in sensitivity when used on corpora with many parts. It could be that some of the words identified as academic are not as evenly distributed in the AcaDan Corpus as assumed.

Clearly, larger general language corpora, and in particular a corpus of general spoken Danish, would be extremely beneficial for future research on Danish high frequency vocabulary, and any overlaps with written Danish from the present study could be addressed. More research is also needed on larger corpora for further validation of the word lists and on different kinds of texts for lexical coverage purposes. It is also important to continue developing the 3<sup>rd</sup> 1,000 list of lemmas, and then proceeding on to the 4<sup>th</sup> and 5<sup>th</sup> lists. This can be seen in the light of the cut-off points between high, mid- and low frequency vocabulary suggested by Schmitt and Schmitt (2014). It would also be useful to investigate the lexical coverage of smaller groups of high frequency words, to follow the work by Dang and Webb (2016; 2017). They investigated the coverages of sets of each 100 headwords of lemmas to determine the number of items in a general word list for beginner learners. That way we can find out more about the nature and behaviour of very frequent words in comparison to less frequent but still fairly high frequency words.

## 5.6. Rationale for Study 2

With the new knowledge provided by Study 1 on the role of Danish general high frequency vocabulary in academic and general written language, Study 1 serves as a basis for Study 2 of this thesis. Study 2 reports on the development of a Danish Academic Word List, and in the vocabulary selection for this the issue of general high frequency vocabulary plays a significant role as the criteria applied for vocabulary selection are carefully chosen to be able to include this type of vocabulary.



## Chapter 6. Study 2 – A Danish Academic Word List

### 6.1. Introduction and research questions

In Study 1, Danish general high frequency vocabulary was explored in relation to academic language and academic vocabulary. Study 2 continued with this focus on academic vocabulary by investigating those lexical items occurring frequently and evenly across academic disciplines in an academic corpus, the AcaDan Corpus, in order to develop an academic word list. Thus, the primary aim of Study 2 was to select vocabulary for a Danish Academic Word List (DAWL), which can be of value in the development of pedagogical tools for strengthening academic language competence for users with Danish as the L1 as well as for L2 students.

Like many other studies that have developed academic and specialised word lists, Study 2 drew on corpus linguistics and primarily quantitative methods. The approach taken in Study 2 for selecting academic vocabulary was, with some notable exceptions, similar to that of Study 1: A **frequency** criterion was set to ensure the identification of lemmas that are more frequent in the academic corpus than in a comparison corpus representing non-academic language. A defining trait of academic words is that they occur in many different disciplines and subject areas. Therefore, two criteria of **range** and **dispersion** related to distribution were also applied to ensure the selection of vocabulary with an even distribution in the corpus. These criteria are further described and discussed in Section 6.2.

Study 2 had two aims: 1) to develop a Danish Academic Word List (the DAWL) and 2) to describe and evaluate the resulting list by answering the following five research questions:

- 1) Which lemmas occur with a higher frequency in academic written language than in non-academic written language and with an even distribution across academic written language, and may therefore be selected for a DAWL?
- 2) Which part of speech categories do the selected DAWL lemmas belong to and in what proportions?
- 3) What is the coverage of the DAWL in academic language, in different academic disciplines and in general language?
- 4) What is the overlap of the DAWL with Danish general high frequency vocabulary and with the 402 words identified as academic in Study 1?
- 5) What is the contribution of the general high frequency DAWL lemmas to the coverage provided by the DAWL in academic language, in different academic disciplines, and in general language?

As indicated by research questions 1 and 2, the resulting list of Danish academic vocabulary, the DAWL, was analysed in relation to frequency, dispersion, and part of speech. Moreover, in Study 3, as described in Chapter 7, the lemmas were described in relation to their function in academic language use.

Evaluation analyses are important to demonstrate the validity of a word list. Following other academic word list studies (Coxhead, 2000; Dang, 2018a; Dang et al., 2017; Gardner & Davies, 2014; Hagen et al., 2016), the DAWL was evaluated by measuring the lexical coverage in different corpora. This evaluation method is widely used in corpus-based word list studies (Miller & Biber, 2015, p. 48). It is relevant to measure the DAWL's coverage over academic as well as general language because it tells us how well the applied extraction methods were in extracting academic vocabulary. A high coverage over an academic language corpus and a low coverage over general language reveal that the vocabulary selection criteria have indeed selected academic vocabulary. In principle, the coverage of a word list should not be measured in the same corpus used for extracting the items for the list (Nation, 2016). Instead, an additional academic corpus should be used for evaluating the list, and, ideally, this corpus should be of the same size as the corpus that forms the basis for the word list (Nation & Webb, 2011). However, given that no prior corpus of Danish academic language use existed when this study was carried out, the evaluation of the DAWL was in fact carried out by using the AcaDan corpus, but also a smaller corpus of academic language. A general language corpus was also used for evaluating the list. Thus, the DAWL was tested against a general language corpus, the AcaDan corpus, and a second, smaller academic corpus in order to answer research question 3. The procedure for the lexical coverage analyses including the development of the DAWL into base word lists are detailed in Section 6.2.

As described in Chapter 5, Study 1 identified 402 lemmas in the general high frequency vocabulary of Danish as academic in nature. Therefore, the high-frequency vocabulary of Danish needed to be taken into consideration in the development of an academic word list in Danish. To be able to measure the contribution of general high frequency lemmas to the coverage of the DAWL, the overlap between general high frequency vocabulary in the form of the 402 general high frequency items identified in Study 1 and the DAWL was measured. Moreover, the overlap between the 2,000 most frequently used lemmas in Danish and the DAWL was measured. Finally, the DAWL was also analysed in relation to the whole list of the 10,000 most frequently used lemmas in Danish. These analyses were guided by research questions 4 and 5.

The presentation of Study 2 is divided into six sections. The first section, Section 6.2, centres on the methodology employed for the identification and selection of vocabulary for the DAWL. In Section 6.3, the DAWL is presented and described by answering the five research questions listed above. After a discussion and limitations section (Section 6.5), Chapter 6 ends with a summary and a rationale for Study 3.

## 6.2. Identifying academic words - vocabulary selection for the DAWL

This methodology section begins with an overview of the corpora used in the development of the DAWL. In the section Data analysis, I introduce and describe the criteria applied in the word selection, and I report on the procedure for preparing the DAWL for lexical coverage analysis.

### 6.2.1. Corpora data

Two corpora were used for vocabulary selection for the DAWL: A corpus of academic written language, the AcaDan Corpus, and a corpus of general language, Journalisten.dk. Both corpora are stored in the corpus programme Sketch Engine (Kilgarriff et al., 2014, 2004) and are described in detail in Chapter 4 of this thesis. It should be noted here that the version of the Journalisten.dk Corpus used in Study 2 is not the same as the version used in Study 1. Here in Study 2, the Journalisten.dk Corpus was set as a sub-corpus in the Danish Web 2017 (daTenTen17) (Jakubíček et al., 2013). The 2017 version (4.6 million running words) contains fewer words compared to the 2014 version (6.2 million running words) described in Chapter 4 and Study 1.

For evaluating the lexical coverage of the DAWL over academic and general language, two additional corpora were used: a corpus of academic language called the Second Academic Language Corpus and the General Language Corpus also used in Study 1 for lexical coverage analysis. Both corpora are described in detail in Chapter 4 of this thesis. Table 6.1 summarises the content and size of all four corpora used in Study 2.

Table 6.1. Overview of corpora used in Study 2

Corpus	Number of running words	Content	Use
<b>The AcaDan Corpus</b>	3,336,991	Professional written academic Danish: Research articles and reports (479 texts) from the four academic disciplines of: Humanities (156 texts, 1,091,674) Social Science (160 texts, 1,065,346) Natural Science (80 texts, 585,369) Health Science (83 texts, 595,600)	Selection of words for the DAWL Lexical coverage analysis
<b>Journalisten.dk (a sub-corpus in the Danish Web 2017 corpus)</b>	4,587,116	Content from the website <a href="http://www.journalisten.dk">www.journalisten.dk</a>	Selection of words for the DAWL
<b>Second Academic Language Corpus</b>	Approximately 298,000	Professional written academic Danish: 7 monographs from Humanities (243,411), 4 research articles from Social Science (23,732), 6 research articles from Natural and Health Sciences (30,659).	Lexical coverage analysis
<b>General language Corpus</b>	Approximately 320,000	Professional written general Danish 265 texts: 240 feature and news articles, 20 leisure magazine articles, 5 sets of teaching materials.	Lexical coverage analysis

For the purpose of selecting DAWL vocabulary, and in particular for the dispersion analysis, a number of sub-corpora in the AcaDan Corpus had to be defined based on academic disciplines. Table 6.2 displays the variety of sub-corpora used in the development of the DAWL.

Table 6.2. Sub-corpora of the AcaDan Corpus

Sub-corpora	Texts	Words	%
<b>Humanities</b>	156	1,091,674	32.71
<b>Humanities_A</b>	86	565,974	16.96
<b>Humanities_B</b>	70	525,700	15.75
<b>Social Science</b>	160	1,065,346	31.92
<b>Social Science_A</b>	80	545,686	16.35
<b>Social Science_B</b>	80	519,659	15.57
<b>Natural and Health Sciences</b>	163	1,179,969	35.36
<b>Natural Sciences</b>	80	585,369	17.54
<b>Health Sciences</b>	83	594,600	17.82

In dividing the corpus into the six sub-corpora (1a+b, 2a+b, and 3a+b), the following considerations were made: First of all, the division should follow the four major academic disciplines represented in the corpus. Because the two disciplines of Natural Sciences and Technology and Health Sciences were only represented with a little less than 600,000 words each, I decided to take the two other

disciplines, Humanities and Social Sciences, each of approximately 1 million words, and divide them into two sub-corpora of each around 500,000 words. This division was made by grouping sub-disciplines that, at least to some degree, resemble each other. For example, the sub-disciplines included in Humanities\_1 are all traditional sub-disciplines within the Humanities, e.g. Literature. The sub-disciplines in Humanities\_2 are, with some notable exceptions, more modern areas of research, e.g. Media Studies. The Social Science discipline was divided into a sub-corpus of anthropology, ethnology and sociology, and a less-related sub-corpus of sub-disciplines such as Information Science and Law. The different sub-disciplines in each of the six sub-corpora are displayed in Table 6.3.

*Table 6.3. An overview of the six sub-corpora in the AcaDan Corpus*

<b>Sub-corpus</b>	<b>Natural Science</b>	<b>Health Science</b>	<b>Humanities_1</b>	<b>Humanities_2</b>	<b>Social Science_1</b>	<b>Social Science_2</b>
<b>Sub-discipline</b>	Biology	Medicine	Literature	Educational	Anthropology	Information
	Physics	Dentistry	Art history	Science	Ethnology	Science
	Geography	Public	Classical	Linguistics	Sociology	Law
	Geology	Health	Studies	Cultural		Economics
	Mathematics	Sports	History	Studies		Psychology
	Chemistry	Science	Archaeology	Media		Political
	Veterinary		Rhetoric	Studies		Science
	Agronomy		Philosophy	Musicology		
	Food Science		Religious			
	Food and		Studies			
	Resource		Theology			
	Economics					
	Engineering					

### 6.2.2. Unit of counting

The vocabulary selection for the DAWL described below identifies the lemmas belonging to Danish academic vocabulary. As described in Chapter 2 in the section on unit of counting, the lemma is defined as a set of lexical items sharing the same stem and part of speech (Francis & Kučera, 1982). Even though the lemma is chosen as the unit of counting for the DAWL, the frequency counts done in the corpus tools of Sketch Engine counted all occurrences of a lexical items and its inflections irrespective of part of speech. This means that some of the identified lemmas in fact have more than one part of speech. As such a small number of the lemmas extracted for the DAWL are in fact flemmas (Pinchbeck, 2014 in Nation, 2016, p. 26), sets of lexical items sharing the same stem but with different parts of speech. This issue is detailed in Sections 6.2.3.7 and 6.4.2.

### 6.2.3. Data analysis

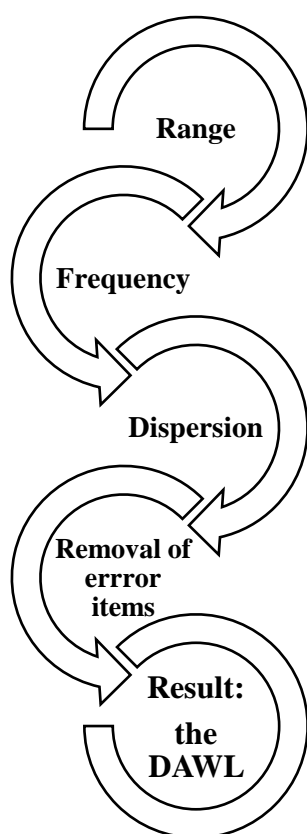
In this section, I give an account of the tools and measures I used for vocabulary selection for the Danish Academic Word List (DAWL).

The Word list function in Sketch Engine was used as the primary tool for the extraction of DAWL lemmas in the AcaDan Corpus. This function enables the extraction of frequency-ranked word lists. In addition, this function allows for frequency counts for a selected number of words via upload of a so-called whitelist containing the words in question to the Whitelist function. To include both initial small and capital letters, the lemma lowercase option was chosen for all steps in the vocabulary selection.

To select lemmas that occur frequently and evenly across disciplines in the AcaDan Corpus, these specific criteria were applied in the selection of the DAWL lemmas:

- 1) **Range:** A DAWL lemma has to occur at least three times in the AcaDan corpus and at least once in each of the three disciplines of Natural and Health Science, Social Science, and Humanities.
- 2) **Frequency:** A DAWL lemma has to occur significantly more frequently in the AcaDan Corpus than in the Journalisten.dk Corpus.
- 3) **Dispersion:** A DAWL lemma has to occur with a Juilland's D value of at least 0.80.

In the next three sections, I detail how these criteria were applied and I justify the methodological choices made in the development of the DAWL. The whole process is also depicted in Figure 6.1 below including the removal of error items which is described in Section 3.2.3.5.



*Figure 6.1. The vocabulary selection process of the DAWL*

#### *6.2.3.1. Range*

The process of applying the range criterion is depicted in Figure 6.2 and involves two steps. The first step was to extract all lemmas occurring at least three times in the AcaDan corpus. A minimum frequency of 3 was chosen to ensure that the selected lemmas in all three disciplinary sub-corpora of Humanities, Natural and Health Science, and Social Science as described in Table 6.2. A smaller minimum frequency of 1 or 2 would interfere with the requirement of a potential academic lemma occurring in all disciplines, i.e. across all three sub-corpora. The resulting list after the first step contained 40,162 lemmas. This list was used for the second step in the range analysis which involved extracting only those lemmas occurring in the three equal-sized sub-corpora of the AcaDan corpus described in Table 6.2: Humanities, Natural and Health Sciences, and Social Science. As such, the second step involved in total three sub-steps as depicted in Figure 6.2. A minimum frequency of 1 was chosen so that for a lemma to be extracted, it had to occur at least once in each of the three sub-corpora. Arguably, due to the merging of the two academic disciplines of Natural and Health Sciences into one sub-corpus, using a minimum frequency of one run the risk of extracting lemmas only occurring in only one of the two disciplines of Natural Science and Health Science. However, the

choice of one as the frequency cut-off was based on the fact that the subsequent application of a somewhat strict dispersion criterion would ensure that the extracted lemmas occur evenly across the corpus.

The Humanities sub-corpus was chosen as the starting point for the second step of the range analysis, but any of the three sub-corpora could have been chosen. 26,275 lemmas of the 40,162 lemmas extracted in step 1 occurred in the Humanities sub-corpus with a frequency of at least one. These 26,275 lemmas formed a whitelist for the extraction of lemmas occurring at least once in the Social Science sub-corpus. This resulted in a new whitelist of 17,909 lemmas to be used for extracting lemmas in the Natural and Health Sciences sub-corpus. The final result after step 2.3 was a list of 12,176 lemmas occurring in all three sub-corpora with a minimum frequency of one.

As a third and a final step, a list of the 12,176 lemmas was used as a whitelist to extract the frequency counts of each of these 12,176 lemmas in each of the three sub-corpora. This was done to make sure that the resulting list did indeed contain lemmas occurring in all three sub-corpora. This check showed that 11 lemmas had to be removed for different reasons. For example, the lemma *demokrat* (democrat) was removed because it did not occur in the Humanities sub-corpus. The lemma *sovjetunionen* (the Soviet Union) was removed because it occurred twice, perhaps due to an error in the annotation of the corpus. The final result was 12,165 lemmas that formed the basis for the next step in the selection of academic words in the AcaDan corpus: the frequency criterion which will be described in the next section.



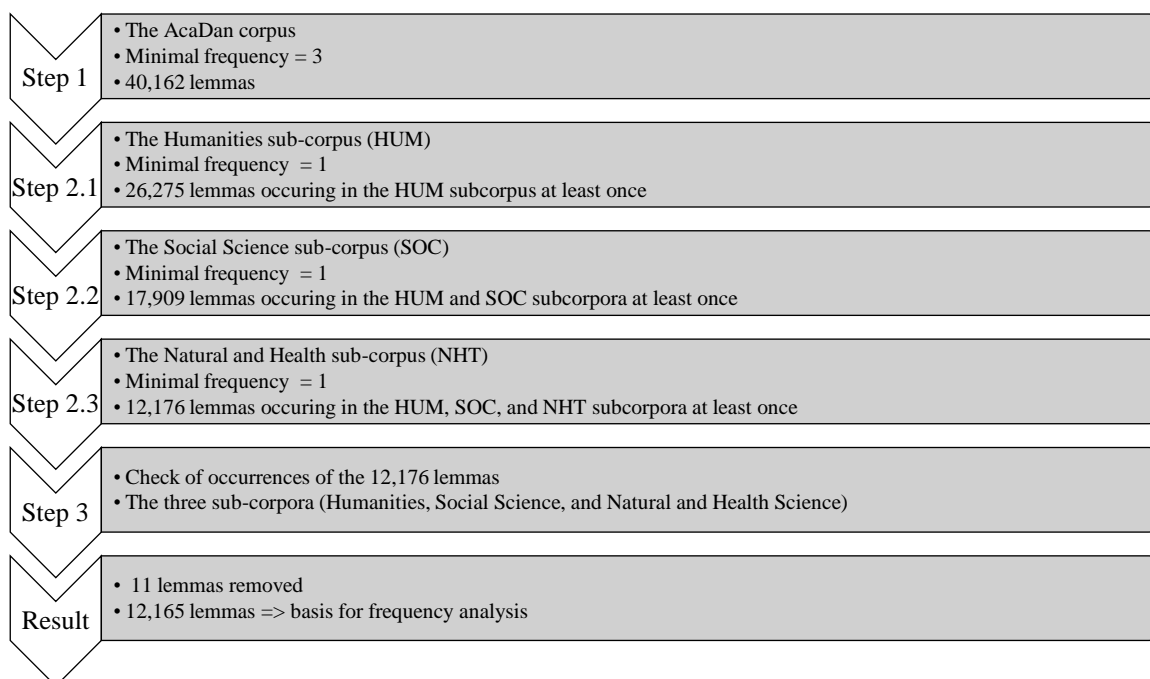


Figure 6.2. The process of the range analysis

### 6.2.3.2. Frequency

As accounted for in Chapter 3, in the vocabulary selection for academic word lists, the frequency criterion can be realised either as a comparison measure or as a pure frequency measure. The comparison measure extracts items that are more frequent in academic language than in non-academic language by comparing the frequency of a lexical item in an academic language corpus with the frequency of the same item in a general language corpus (cf. Gardner & Davies, 2014; Hagen et al., 2016; Paquot, 2010). The pure frequency measure, on the other hand, uses a cut-off point to extract lexical items occurring with a certain frequency in an academic language corpus, e.g. very high frequency vocabulary items (cf. Coxhead, 2000; Dang, 2017; Dang et al., 2017). The use of comparison measures is primarily motivated by the wish to include general high frequency words with academic uses in academic word lists. Admittedly, this can also be done using a pure frequency measure with a frequency cut-off point as done in Dang et al.'s (2017) Academic Spoken Word List in which all selected items had to occur with a frequency of minimum 350 in the academic corpus used for vocabulary selection. This meant that general high frequency items such as 'be', 'the', 'and', 'a', 'that', and 'to' were selected and ultimately included in the word list. It is questionable if these items are indeed academic words.

The frequency criterion for selection of words for the DAWL states that a lemma has to be significantly more frequent in the AcaDan Corpus than in the Journalisten.dk Corpus. To satisfy this

frequency criterion, a comparison measure was chosen in the present study. The reason for this choice has largely to do with the issue of general high frequency vocabulary in Danish explored in Study 1. The lexical coverage results from Study 1 showed that Danish academic language differs notably from Danish general language in that the 2,000 most frequent lemmas in Danish (referred to as general high frequency vocabulary) cover around 63 percent of academic language compared to 76 percent of general language. However, Study 1 also showed that 402 words in the general high frequency vocabulary of Danish can be defined as academic. This finding supports previous research on academic vocabulary (Dang et al., 2017; Gardner & Davies, 2014; Paquot, 2010) which have shown that general high frequency words can be academic by using comparison measures instead of a pure frequency measure with a cut-off point for selecting academic words. Thus, a comparison measure was applied to be able to capture the Danish general high frequency words that could also have an academic function.

As also discussed in Chapter 3, a comparison measure can be carried out in different ways. In Study 1, I calculated the **relative frequency ratio** (Gries, 2010) of the most frequent 2,000 lemmas in Danish and set a cut-off value of 1.5 based on experimentation. Together with range and dispersion measures, the relative frequency ratio measure identified 402 academic lemmas among the general high frequency vocabulary of Danish. In this study, the statistical test of **log-likelihood ratio** was chosen as the method for meeting the frequency criterion. The log-likelihood ratio test is similar to the keyness measure applied in Paquot's (2010) list of academic keywords and in the Swedish academic word list (Ribeck et al., 2014). A log-likelihood test was also used in the development of the Academic Formulas List by Simpson-Vlach and Ellis (2010). Using a log-likelihood ratio test ensures the extraction of words that are distinctive of the chosen corpus in contrast to the relative frequency ratio that only extracts items occurring more frequently in one corpus compared to another. The decision for using the log-likelihood ratio measure over the relative frequency ratio measure was reached after experimentation with both measures. This experimentation is reported on in the three succeeding sections. For both measures it was necessary to collect the frequencies of the 12,165 lemmas from the range analysis in both the AcaDan corpus and in the Journalisten.dk corpus using the Whitelist function in the Word list function in Sketch Engine.

### *Relative frequency ratio*

In applying the relative frequency ratio measure, the first step was to normalise the frequency figures from the two corpora, the AcaDan corpus and the Journalisten.dk corpus, to frequencies per million

because the two corpora are of unequal sizes. The relative frequency ratio was calculated for each lemma by dividing the AcaDan frequency per million figure with the Journalisten.dk frequency per million figure. The second step was to choose an appropriate cut-off value, which was done by experimenting with cut-off values used in previous research (Gardner & Davies, 2014; Hagen et al., 2016). The cut-off values (2.6 and 2.8) used by Hagen et al. (2016) resulted in a rather limited number of lemmas. Ultimately, the relative frequency ratio value of 1.5 was chosen. This value was also applied in Gardner and Davies' (2014) development of the Academic Vocabulary List. Out of the 12,165 lemmas satisfying the range criterion, 6,544 lemmas occurred at least 1.5 times more frequently in the AcaDan corpus than in the Journalisten.dk corpus. These 6,544 lemmas comprised what is referred to below as Result List 1.

#### *Log-likelihood ratio test*

To calculate the log-likelihood ratio of the 12,165 lemmas extracted through the range analysis, these lemmas and their frequencies were inserted into a log-likelihood ratio calculator (Xu, 2009), together with the sizes of the AcaDan Corpus and the Journalisten.dk Corpus. The calculator indicated which lemmas occurred significantly more in the academic corpus than in the comparison corpus. The significance level used in the calculator was  $p = 0.0001$  which means that there is less than 0.01 per cent chance of incorrectly claiming that the difference in frequency is significant. A total of 7,942 out of 12,165 lemmas occurred with a statistically higher frequency in the AcaDan corpus than in the Journalisten.dk corpus. These 7,942 lemmas comprised what is referred to below as Result list 2.

#### *Choosing the comparison measure*

To justify the choice of one comparison measure over the other, two overlap analyses were carried out. First, an overlap analysis was carried out to see how many lemmas occurred in both lists and which lemmas would be excluded dependent on the measure. Table 6.4 shows 1) how many lemmas each of the two comparison measures resulted in, and 2) how many lemmas occur in both results lists and how many do not. Included are also examples of shared and not-shared items.

*Table 6.4. Results of the overlap analysis*

<b>Result list 1 – Relative Frequency Ratio (RFR)</b>	<b>Result list 2 – Log-likelihood ratio test (LLR)</b>	<b>Number of lemmas shared between Result lists 1 and 2</b>	<b>Number of lemmas unique to Result list 1</b>	<b>Number of lemmas unique to Result list 2</b>
6,544	7,942	6,398	146	1,544

This analysis showed that 80 percent of the lemmas, 6,398 to be precise, were shared between the two lists, and that of the 7,942 lemmas from the log-likelihood ratio test, 1,544 of these did not occur in the list of 6,544 lemmas from the relative frequency ratio analysis. These 1,544 lemmas occurred less than 1.5 times more frequently in the academic corpus, yet the difference in frequencies between the two corpora is still significant. Examples of lemmas excluded by the relative frequency ratio measure were verbs such as *anslå* (estimate), *balancere* (balance), *identificere* (identify), *optimere* (optimise), *revidere* (revise), *tilhøre* (belong to), *udføre* (carry out), and *vægte* (weight), nouns such as *fejlslutning* (fallacy), *kendsgerning* (fact), and *problem* (problem), adverbs and adjectives such as *desuden* (moreover), *naturlig* (natural, naturally), *overvejende* (predominant, predominantly), *samlige* (all), and *øvrige* (other), and finally conjunctions such as *hverken* (neither), *hvortil* (how far), and *hvordan* (how). Conversely, 146 lemmas in result list 1 were not found in result list 2. Almost all of these were either abbreviations (56%) such as *dvs.* (i.e.), *ibid.* (ibid.), *pga.* (because of), *fx* (for example), *ph.d.* (PhD), or tokenisation errors (40%) such as ‘der|der’, ‘fra|fra|fra, or ‘befi’ (probably a clipping of the verb *befinde* (find). It can be argued, however, that abbreviations such as *dvs.*, *pga.*, and *ibid.* may be of relevance to L2 students as Danish make much use of these abbreviations in writing.

The second overlap analysis carried out to justify the choice of comparison measure was in fact an evaluation of the ability of the log-likelihood ratio test and the relative frequency ratio measure to extract general high frequency items. As mentioned above, Study 1 showed that many general high frequency items in Danish may have an academic function. Moreover, these items may be polysemous and may therefore need to be focussed on in a pedagogical academic word list. It was therefore important to find out how efficient the two measures were to extract general high frequency words. This was done by analysing the overlap between the two result lists and the most frequent 2,000 lemmas in Danish and the 402 words identified as academic in Study 1. The results of these analyses are given in Table 6.5.

Table 6.5. Results from the second overlap analysis

Overlap between the 2,000 most frequent lemmas and Result list 1(RFR)	Overlap between the 2,000 most frequent lemmas and Result list 2 (LLR)	Overlap between the 402 high frequency academic lemmas and Result list 1	Overlap between the 402 high frequency academic lemmas and Result list 2
472 lemmas	767 lemmas	244 lemmas	331 lemmas

As can be seen in Table 6.5, the log-likelihood ratio test (Result list 2) extracts a higher number of lemmas found in the 2,000 most frequent lemmas of Danish than the relative frequency ratio (Result list 1) does. Almost 40 percent of the most frequent 2,000 lemmas in Danish occurred in Result list 2 compared to only 24 percent in Result list 1. Of the 402 general high frequency academic lemmas from Study 1, 61 percent of them occurred in Result list 1, while almost 83 percent of them occur in Result list 2. These results suggest that a cut-off value of 1.5 might be too high for Danish if the aim is to capture items that are both general and academic in nature. Clearly, the log-likelihood ratio test did a better job of extracting general high frequency items occurring more frequently in academic language than in non-academic language than the relative frequency ratio measure.

Based on the results from the comparison of the two measures reported on above, the log-likelihood ratio test was chosen over the relative frequency ratio. The log-likelihood ratio test ensured that the difference in frequencies of the 7,942 extracted lemmas is statistically significant, thus emphasising that these lemmas are distinctive in academic language use, even if they are also general high frequency words in Danish.

#### *6.2.3.3. Dispersion*

The third and last vocabulary selection criterion is the dispersion criterion which is applied to ensure that the selected lemmas occur evenly across the AcaDan Corpus. Thus, the purpose of applying a dispersion criterion is to ensure that only items that are not too specific to certain disciplines or sub-disciplines are included in the identified pool of lemmas. As in Study 1, Juilland and Chang-Rodríguez's (1964) dispersion measure, Juilland's D, was chosen, as this has been widely used in similar studies on academic vocabulary (Dang et al., 2017; Gardner & Davies, 2014; Hagen et al., 2016; Paquot, 2010).

The 7,942 lemmas extracted by the log-likelihood ratio test formed the basis for calculating Juilland's D of each of these lemmas. The dispersion calculation was based on six sub-corpora (see Table 6.2). Only items with a Juilland's D value equal to or above 0.80 were selected for the DAWL. In the following two sections, I justify the number of sub-corpora used and the chosen cut-off value.

#### *Number of sub-corpora for dispersion measurement*

The calculation of the Juilland's D measure is contingent on the corpus being divided into a number of sections or sub-corpora (Gries, 2008). For the extraction of academic vocabulary, the divisions of the corpus should reflect the disciplines represented in the corpus used. The decision to use six sub-corpora instead of three sub-corpora as in the range analysis was based on experiments with different

numbers of sub-corpora. Biber, Reppen, Schnur, and Ghanem (2016) found that the sensitivity of the Juilland's D measure decreased the more sub-corpora the measure was calculated on. This could be seen as an argument for using as few sub-corpora as possible. An initial dispersion analysis using three sub-corpora (Humanities, Social Science, and Natural and Health Sciences, see Table 6.2 for details) for calculating the dispersion of the 7,942 lemmas resulted in a list of 1,569 lemmas with a D value of 0.80. A manual check of this list revealed apparent technical items such as *algebraisk* (algebraic), *artikulation* (articulation), *dissikere* (dissect), *identitetskrise* (identity crisis), and *kartografisk* (cartographic). In addition, this list contained a number of English words, formatting abbreviations, and incomplete lemmas (due to errors in the lemmatisation procedure). It thus seemed that calculating dispersion using the three sub-corpora did not manage to effectively ensure items from a broad range. Calculating the dispersion of the 7,942 lemmas using four sub-corpora (Social Science, Natural Science, Humanities, and Health Science) resulted in 860 items with a D value of 0.80. A manual check revealed, however, that this list also included a large number of English words, abbreviations, proper nouns, and error items. Removing these items would reduce the list notably. In comparison with these two experiments, using the six sub-corpora outlined in Table 6.2 resulted in a list of 903 items with a D value of 0.80. I will return to this list and why it constituted the best result, but, in order to do so, the chosen cut-off D value of 0.80 needs to be justified. The following section reports on experimentations with three different cut-off points which led to the choice of 0.80 as the cut-off point.

### *Setting the dispersion cut-off point*

The calculation of Juilland's D returns a number between zero and one. The closer an item's D value is to one, the more evenly the item is distributed in the corpus. In other words, setting the appropriate cut-off point is a question of how effective Juilland and Chang-Rodríguez's (1964) dispersion D is at eliminating items with an uneven distribution. It was thus a central concern in the development of the DAWL to set an appropriate cut-off point. As the aim of the dispersion measure is to extract items occurring as evenly as possible, it seems reasonable to set the cut-off point as high as possible so that only the items occurring very evenly would be identified as academic words. A heuristic approach was used to set the appropriate cut-off point as no Danish word list for lexical coverage analysis existed when the study was carried out (cf. Dang et al., 2017). The heuristic approach comprised comparing three lists of lemmas with D values between 1) 0.60-0.69, 2) 0.70-0.79, and 3) equal to and above 0.80. A manual inspection of each list was carried out to ascertain which cut-off value gave the most satisfying result, i.e. a list with as few technical items, English language items,

formatting abbreviations, and incomplete lemmas as possible. The manual inspection included going through the lists marking items with seemingly technical meanings and error items. Error items were pre-defined as English words, proper nouns, formatting items, and items lemmatised or tokenised incorrectly. In addition, selected items from the 0.60-0.69 and 0.70-0.79 lists were compared to the Academic Word List (Coxhead, 2000) and the Academic Vocabulary List (Gardner & Davies, 2014). The central question for the comparison analysis was if a D value below 0.80 could be considered as a suitable cut-off value. The 0.80 D value has proven to work well in the development of academic word lists such as Gardner and Davies's (2014) Academic Vocabulary List and Paquot's (2010) Academic Key Word List. Thus, this value was used as the baseline for comparisons with lists of lemmas with D values between 0.60 and 0.69 and between 0.70 and 0.79. The calculation of the D value of each lemma was done using six sub-corpora (see Table 6.2).

As can be seen from Table 6.6, the lower the cut-off value, the more items were extracted. Table 6.7 shows that the number of items decreases drastically from 3,836 at 0.60 and 2,182 at 0.70 to 903 items when applying 0.80 as the cut-off D value. 2,923 lemmas are excluded if a D value of 0.80 is used as the cut-off point.

*Table 6.6. Number of lemmas between different D values*

<b>D value</b>	<b>0.60-0.69</b>	<b>0.70-0.79</b>	<b>0.80-1.0</b>
<b>Number of lemmas</b>	1,644	1,279	903
<b>Percentage of error items</b>	29.38	25.25	15.84

*Table 6.7. Number of lemmas at different cut-off D values*

<b>Cut-off D value</b>	<b>0.60</b>	<b>0.70</b>	<b>0.80</b>
<b>Number of lemmas</b>	3,836	2,182	903

Table 6.6 also shows the percentage of error items in each list. What the percentage figures show is that the higher the D value, the lower the percentage of error items. Note that the removal of error items from the DAWL is described in Section 6.2.3.4 which also discusses the origin of these error items.

Randomly chosen examples of lemmas with D values between 0.70 and 0.79 are given in Table 6.8. A comparison of the English translations of these 24 lemmas with the Academic Word List (Coxhead, 2000) and the Academic Vocabulary List (Gardner & Davies, 2014) showed that eight of these 24 lemmas occurred in the Academic Word List and 20 in the Academic Vocabulary List (see Table

6.8). Since the translations of these examples can be found in English academic word lists, 21 out of the 24 lemmas given as examples can be claimed to be academic.

In Table 6.9, 20 randomly chosen lemmas from the 0.60-0.69 list are shown. These lemmas all seem to be rather general in meaning and not technical. When translated into English, 15 of them can be found in the Academic Word List (AWL) and/or the Academic Vocabulary List (AVL). Parts of compounds found in one or both of the two English lists are in bold. For example, the English translation of *sandsynliggøre* is ‘render probable’ and both of these items are found in the Academic Vocabulary List. The comparison of the 0.70-0.79 and 0.60-0.69 lists with the Academic Word List and the Academic Vocabulary List indicate that the 0.70-0.79 list contained more items shared with the Academic Vocabulary List and the Academic Word List than did the 0.60-0.69 list.



Table 6.8. Examples of lemmas with a D value between 0.70 and 0.79

<b>0.70-0.72</b>		<b>0.73-.0.75</b>		<b>0.76-0.77</b>		<b>0.78-0.79</b>	
Danish	English	Danish	English	Danish	English	Danish	English
<i>abstrakt</i>	abstract (AWL+AVL)	<i>afprøve</i>	test (AVL)	<i>anskueliggøre</i>	illustrate (AWL+AVL)	<i>afdækning</i>	uncovering
<i>akademisk</i>	academic	<i>argumentere</i>	argue (AVL)	<i>eksperimentere</i>	experiment (AVL)	<i>Anerkende</i>	recognise (AVL)
<i>anfægte</i>	affect (AWL+AVL)	<i>faglig</i>	technical (AWL+AVL)	<i>formindske</i>	reduce (AVL)	<i>forskning</i>	research (AWL+AVL)
<i>deskriptiv</i>	descriptive	<i>hypotese</i>	hypothesis (AWL+AVL)	<i>førstnævnte</i>	former	<i>identificere</i>	identify (AWL+AVL)
<i>italesætte</i>	articulate (AVL)	<i>indvirkning</i>	impact (AVL)	<i>medføre</i>	entail (AVL)	<i>overvejende</i>	predominantly (AVL)
<i>kortlægge</i>	survey (AVL)	<i>kritik</i>	critique (AVL)	<i>virkning</i>	effect (AVL)	<i>nuancere</i>	vary (AWL+AVL)

Table 6.9. Examples of lemmas with a dispersion value between 0.60 and 0.69

<b>0.60-0.62</b>		<b>0.63-.0.65</b>		<b>0.66-0.67</b>		<b>0.68-0.69</b>	
Danish	English	Danish	English	Danish	English	Danish	English
<i>evident</i>	evident (AWL+AVL)	<i>argument</i>	argument	<i>emneområde</i>	subject field (AVL)	<i>antagelig</i>	probable (AVL)
<i>forståelsesramme</i>	frame of <b>understanding</b> (AVL)	<i>konceptualisere</i>	conceptualise (AVL)	<i>forskningsområde</i>	research field (AWL+AVL)	<i>vurdere</i>	Estimate (AWL+AVL)
<i>konkretisere</i>	make concrete	<i>plausibel</i>	plausible (AVL)	<i>hovedresultat</i>	main <b>result</b> (AVL)	<i>dybdegående</i>	in-depth (AVL)
<i>grundprincip</i>	<b>basic principle</b> (AVL+AWL)	<i>sammenfatning</i>	Synthesis (AVL)	<i>kernebegreb</i>	<b>core concept</b> (AVL+AWL)	<i>førnævnt</i>	above- mentioned
<i>positionering</i>	positioning	<i>sandsynliggøre</i>	<b>render probable</b> (AVL)	<i>problematisering</i>	problematisation	<i>udlægning</i>	interpretation (AWL+AVL)

The 0.60-0.69 examples given in Table 6.9 also illustrate a difference between Danish and English related to the issue of compounds (marked in bold). In Danish, compounds consisting of independent words are normally joined together as in *emneområde* (subject area) or *forskningsområde* (research area). This is most often not the case in English where two-word collocations are often used. To further investigate the issues of compounds, all compounds in the three lists were marked. Only compounds consisting of independent elements such as *emneområde* which is made up of the two nouns *emne* (subject) and *område* (area) were marked. This analysis showed that the occurrence of compounds decreased the higher the D value was. In the 0.60-0.69 list, the percentage of compounds was 13 percent, in the 0.70-0.79 list it was five percent, and in the baseline list it was two percent. The fact that the baseline list contains so few compounds (16) suggests that the use of compounds is related to specialised language use in particular which explains the more uneven distribution of these in the corpus.

Table 6.10 shows the percentages and examples of lemmas in the three lists which are seemingly technical according to the manual inspection of these lists. A lemma was marked as seemingly technical if it was deemed to be a word only used by specialists within one or more disciplines. This marking was purely intuitive and was carried out to ascertain if, from a subjective vantage point, there were any items in the three lists that seemed irrelevant to an academic word list. As can be seen, the baseline list contained the fewest technical items while the 0.60-0.69 list has the highest percentage of seemingly technical words. If we look closer at the examples given in Table 6.10, we can expect to meet words like *kognitiv* and *modellering* in texts from the Humanities as well as from the Natural and Health Sciences:

Den **kognitive** lingvistik bygger sit syn på sprog på den antagelse, at sprog afspejler vores tankemønstre. (Humanities)<sup>15</sup>

English translation: **Cognitive** linguistics bases its language view on the assumption that language reflects our thoughts.

(...) at det overordnede formål med den **kognitive** testning er at etablere en baseline for senere studier af aldersrelaterede ændringer i **kognitiv** funktion. (Health Science)<sup>16</sup>

<sup>15</sup> Petersen, C., & Engerer, V. P. (2014). 'Den lange rejse ...' - metaforiske betydningslag og branding i filmmediet. *Mediekultur*, 30(57), 154–175.

<sup>16</sup> Avlund, K., Bruunsgaard, H., Christensen, U., Fiehn, N.-E., Hansen, Å. M., Holm-Pedersen, P., ... Lund, R. (2009). CAMB - Copenhagen Aging and Midlife Biobank. Perspektiver for fremtidig forskning. *Miljø Og Sundhed*, 15 (suppl. nr. 1), 81–88.

English translation: (...) that the primary purpose of the cognitive testing is to establish a baseline for later studies of age-related changes in cognitive function.

Derfor repræsenterer denne migrantkirke også en postwestfalsk modellering over det økonomiske funktionssystem. (Humanties)<sup>17</sup>

English translation: Therefore, this migrant church also represents a post-Westphalian modelling of the financial functional system.

Bioøkonomiske modellering er et veletableret felt inden for fiskeri- og naturressourceøkonomien og kan dateres tilbage til Warmings arbejde om optimalt fiskeri fra 1911(...). (Natural Science)<sup>18</sup>

English translation: Bio-economical modelling is a well-established field within fishing and natural resource economics, and can be dated back to Warming's work in optimal fishing from 1911 (...).

This highlights the fact that some academic words can have different senses according to the context they occur in while still fulfilling the quantitative definition of occurring with a high frequency and across many disciplines. It is, however, debatable if words such as *bakteriologi*, *kulturhistorisk læsefærdighed*, or *amme* are in fact used in all disciplines.

---

<sup>17</sup> Trolle, A. K. (2015). Migrantkirker. Religionsvidenskabeligt Tidsskrift, 62 (Temanummer Den danske religionsmodels grænseflade), 61–75.

<sup>18</sup> Ravensbeck, L., Frost, H., & Andersen, P. (2013). Fiskeri, økosystemtjenester og økonomi. Nationaløkonomisk Tidsskrift, 151, 259-277.

Table 6.10. Examples of technical lemmas with a *D* value between 0.70 and 0.79 and above 0.80

Technical lemmas in the 0.60-0.69 list	Technical lemmas in the 0.70-0.79 list	Technical lemmas in the baseline list
13 percent	8 percent	< 1%
<i>avantgarde</i> (avant-garde, noun)	<i>ahistorisk</i> (ahistorical, adj.)	<i>binær</i> (binary, adj.)
<i>arbejderklasse</i> (working class, noun)	<i>amme</i> (nurse, verb)	<i>cirkulation</i> (cirkulation, noun)
<i>bakteriologi</i> (bacteriology, noun)	<i>epistemologisk</i> (epistemological, adj.)	<i>heterogenitet</i> (heterogeneity, noun)
<i>formativ</i> (formative, adj.)	<i>fysiologisk</i> (physiological, adj.)	<i>indvandring</i> (immigration, noun)
<i>germansk</i> (germanic, adj.)	<i>industrialisering</i> (industrialisation, noun)	
<i>kognitiv</i> (cognitive, adj.)	<i>kulturhistorisk</i> (culture historical, adj.)	
<i>kønsspecifik</i> (gender-specific, adj.)	<i>landbefolkning</i> (rural population, noun)	
<i>lovfæste</i> (legalise, verb)	<i>markedsorienteret</i> (market-oriented, adj.)	
<i>læsefærdighed</i> (reading proficiency, noun)	<i>ontologisk</i> (ontological, adj.)	
<i>markedskraft</i> (market forces, noun)	<i>socialkonstruktivisme</i> (social constructivism, noun)	
<i>matematisk</i> (mathematical, adj.)		
<i>modellering</i> (modelling, noun)		
<i>postmoderne</i> (postmodern, adj.)		
<i>signifikans</i> (significance, noun)		
<i>sygehusvæsen</i> (hospital service, noun)		

While some of the items in the three lists can be regarded as technical words, such as *ontologisk*, *lovfæste*, and *postmoderne* listed in Table 6.10, others can be said to be technical words used in a figurative sense. An example from the baseline list is *binær* which in most of the concordance data in the AcaDan corpus is used as a technical word related to statistics and computer science:

Tabel 1 viser marginalfordelinger for de seks items i 1981, 1990, 1997, 1999 og 2005, samt fordelingen af de socioøkonomiske og demografiske kontrolvariable. I DDV er de seks items **binære** (valgt = 1 / ikke valgt = 0), mens de i AD er ordinale. (Social Science)<sup>19</sup>

English translation: Table 1 shows the marginal distribution for the six items in 1981, 1990, 1997, 1999, and 2005 together with the distribution of the socio-economic and demographic control variables. In DDV the six items are **binary** (chosen = 1 / not chosen = 0), while in AD they are ordinale.

<sup>19</sup> Holm, A., & Jæger, M. M. (2008). Livsformer i Danmark. Dansk Sociologi, 19(1), 31–53.

In two instances, however, *binær* is used in a more figurative sense:

(...) at queer-perspektivet helt på linje med de gamle seksualitetshistorier antænder en ny **binær** oppositionsdannelse mellem queer og straight/normal. (Social Science)<sup>20</sup>

English translation: (...) the fact that the queer perspective is completely in line with the old sexuality stories ignites a new **binary** creation of opposition between queer and straight/normal.

Den kan altså tolkes som en tekst, der selv undlader at bidrage til en reduktivt **binær** læsning, hvor der tilstræbes entydighed og et tydeligt budskab – eller som præsten siger i sin eksegese i Processen (...). (Humanities)<sup>21</sup>

English translation: It can thus be interpreted as a text that in itself omits participating to a reductive **binary** reading in which unambiguity and a clear message is sought after – or as the priest says in his exegesis in the Process (...).

This pattern of technical words used in figurative senses is also found among the items marked seemingly technical in the two lists of 0.70-0.79 and 0.60-0.69. The item *landvinding* occurs with a D value of 0.74 which is not surprising when looking at the concordance data for this word in the AcaDan corpus. In only one out of 15 concordances, does the word occur with its technical sense (land reclamation):

(...) fremkomst af nye yngleområder i forbindelse med kystsikring og **landvinding** og oprettelse af et stort antal reservater for ynglende kystfugle (...). (Natural Science)<sup>22</sup>

English translation: (...) the emergence of new breeding areas in connection with coast protection and **land reclamation** and the establishment of a large number of protected areas for breeding coast birds (...).

In the rest of the concordances, the word is used with the transferred sense of progress or conquest:

Men den herskende forståelse af den kulturelt aktive borger som modtager er også en følge af teknologiske **landvindinger**. (Social Science)<sup>23</sup>

English translation: But the dominant understanding of culturally active citizens as recipients is also a consequence of technological **progress**.

---

<sup>20</sup> Wøldike, M. E. (2007). Kvinders smag for mænd, mænds smag for kvinder. *Kvinder, Køn Og Forskning*, (4), 9–20.

<sup>21</sup> Johansen, M. B. (2015). ”Jeg har forstået den sådan, at den ikke skal forstås.” *Acta Didactica Norge - Nasjonalt Tidsskrift for Fagdidaktisk Forsknings- Og Utviklingsarbeid*, 9(1), 1–20.

<sup>22</sup> Bregnballe, T., Thorup, O., Jacobsen, L. B., Kjeldsen, J. P., & Hansen, M. (2015). Udviklingen i ynglebestanden af Klyder i Danmark 1970-2014. *Dansk Orn. Foren. Tidsskr.*, 109, 121-133.

<sup>23</sup> Rasmussen, C. H. (2015). Brugerinddragelse og kulturpolitisk kvalitet. *Nordisk Kulturpolitisk Tidsskrift*, 18(1), 76–95.

As Table 6.10 shows, only four items in the baseline list were deemed technical. From the concordance data from the AcaDan corpus, it is apparent that the core meaning of the lemma *cirkulation*, which means the movement or spreading of something especially within a confined area, is the same across disciplines and subject areas.

The comparison analyses carried out here suggest that the three D values of 0.60, 0.70, and 0.80 differ in their ability to eliminate error items and technical words. The percentage of seemingly technical items decreased as the D value increased. Based on this finding, the D value of 0.80 was chosen as the appropriate cut-off point for DAWL as this list contained the fewest error items and seemingly technical words. However, setting the cut-off point to 0.70 could also have been a justifiable choice since the small-scale comparison of the 0.70-0.79 list with the AWL and the AVL showed that when translated into English, most of the examples could be found in the English academic word lists. A further analysis of the lemmas with D values between 0.60 and 0.79 will be discussed Chapter 8 in relation to the presentation of Study 4, which outlines procedures for supplementing a DAWL based primarily on a 0.80 cut-off value.

#### 6.2.3.4. *Removal of error items*

This section describes the manual checking of the list of 903 lemmas for possible error items. First, the different groups of error items and the cause of them are accounted for. Then the procedure of checking the 903 lemmas for error items is described with special attention to the error items resulting from errors in the lemmatisation, tokenisation and POS-tagging of the AcaDan Corpus.

Error items refer to different items in the corpus texts that I did not manage to remove in the corpus compilation phase. As described above, error items comprise English words, proper nouns, formatting items, and items lemmatised or tokenised incorrectly. English words occur in the AcaDan Corpus because of English abstracts, citations, and references in the corpus texts. Formatting items are headers, footers, pagination, tables, figures and abbreviations used in references. The motivation behind removing proper nouns is twofold. The use of them is often closely related to the content of the text and the meanings of them are in general transparent to the reader. The reason why they occur among the selected lemmas for the DAWL is that they either refer to publishing houses, cities, or frequently occurring surnames in Danish (see Table 6.11). Finally, the annotation of the AcaDan Corpus in relation to lemmatisation, tokenisation and POS-tagging had some faults which resulted in some lemmas occurring in an incorrect form.

The removal of error items was carried out by having three colleagues and myself go through the 903 lemmas individually and marking any items that could be perceived to be an error. In average, 15 percent of the 903 words were marked as error items by my three colleagues and me. Only in 3.5 percent of the 903 words resulted in disagreement between our markings. The marking resulted in the total marking of 151 items out of which 32 were not marked by all four of us. Of these 32 items, I removed five items without further analyses because they were either proper names or formatting items leaving 27 items for further analysis. Concordance data from the AcaDan corpus supported the removal of these 27 items. Table 6.11 contains a categorisation of the 151 error items into five types, the causes of them, and the course of action taken in relation to them.

Table 6.11. Overview of error items

Error item	English words	Proper nouns <sup>24</sup>	Formatting items	Items lemmatised, tokenised or POS-tagged incorrectly	Misc.
<b>Number of removed items</b>	82	9	26	33	1
<b>Examples</b>	<i>describe</i> <i>requirements</i> <i>exploration</i> <i>only</i>	<i>københavn</i> <i>routledge</i> <i>mørch</i> <i>larsen</i>	<i>o.</i> <i>vol.</i> <i>ed.</i> <i>p</i>	<i>gisk</i> <i>ger</i> <i>developmenten</i> <i>doe</i>	<i>tilstede</i> (present)

The miscellaneous category only contains the item *tilstede*. It appears to be a verb, meaning *permit*, but a search in the corpus showed that all occurrences *tilstede* were a misspelling of the adverbial phrase *til stede* meaning *present*. The lemma *stede* occurred among the 903 lemmas and a corpus search showed that it was only used in the phrase *til stede*. Therefore, the lemma *tilstede* was deleted as any instances of it were used in the same meaning as the lemma *stede* was used.

The majority of the 33 items tokenised or lemmatised incorrectly were English words that had been lemmatised as if they were Danish words, e.g. *doe*, *wher* and *developmenten*. A few of the items were cut-off suffixes, e.g. *gisk*, *ger*, *ek*. All these were removed. Seven error items in this category were replaced with other lemmas after having been checked in the corpus. These are described below:

1. The word *forvejen* was incorrectly lemmatised as *forvej* by the lemmatiser. In the DSL list of the 10,000 most frequently used lemmas in Danish, the word is listed as *forvejen*. This is also the

<sup>24</sup> Proper nouns are not capitalised here because the lemma lowercase option was chosen for all steps in the vocabulary selection in order to include both initial small and capital letters, as described in Section 6.2.3.

entry in the Danish Dictionary. Moreover, all occurrences in the AcaDan Corpus are in the definitive form *forvejen* as part of the phrase *i forvejen* (in advance). Therefore, *forvej* was replaced with *forvejen* in the DAWL.

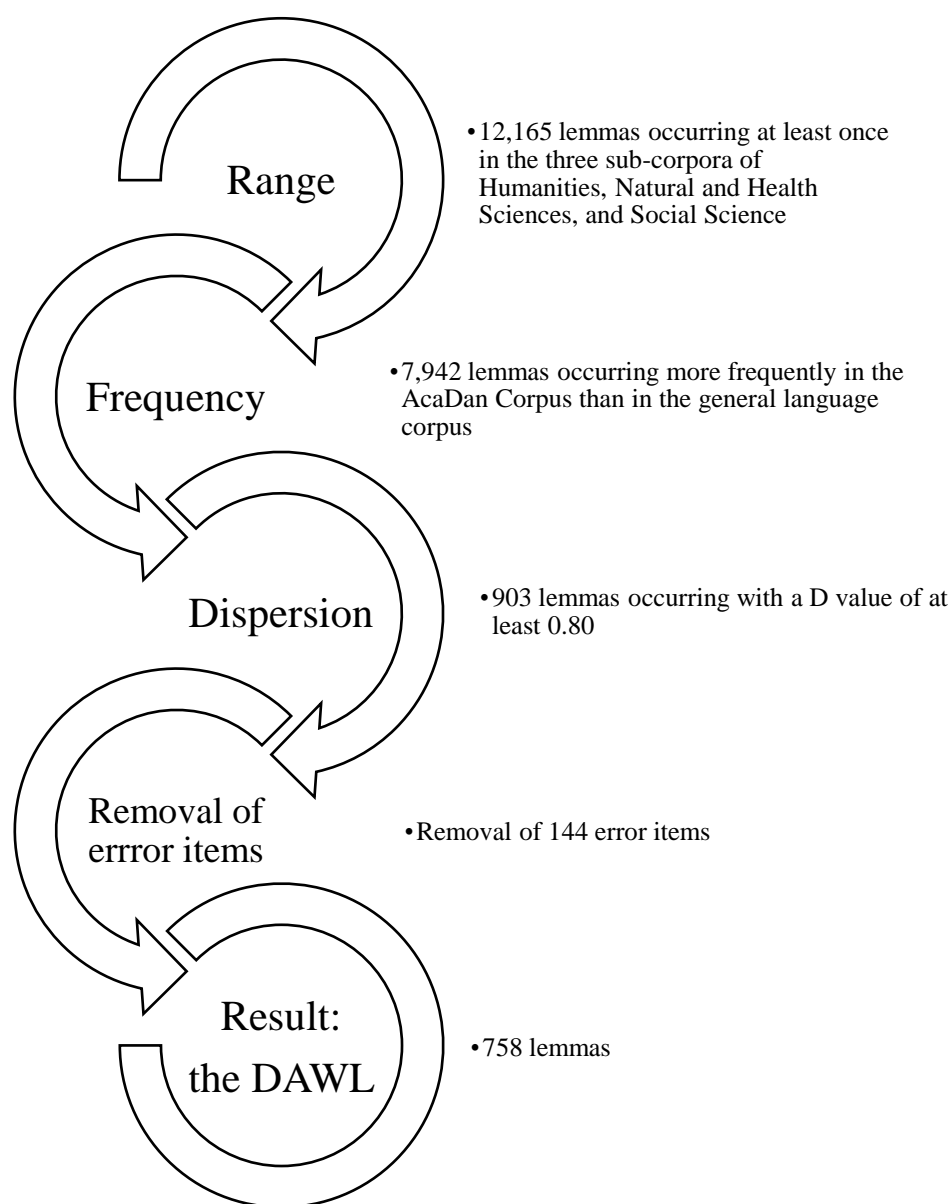
2. The item *indledningsvis* POS-tagged as a verb was replaced by the adverb *indledningvis* since all occurrences were as adverbial and the POS tagging as verb was thus wrong. Moreover, *indledningsvis* is not an entry in the Danish Dictionary or in the DSL list of the 10,000 most frequently used lemmas in Danish.
3. The item *me* was replaced with *men* (but) because the concordance data showed that all occurrences of *me* were in fact the conjunction *men* occurring after a full stop. The lemmatiser had lemmatised this particular occurrence as *me*. It could be argued that this item should be removed because it only refers to occurrences after full stop. I decided to keep it in the DAWL because it did satisfy the three criteria applied for the word selection.
4. The item *med.* was replaced with *med* (with) because the concordance data showed that the majority of instances were in fact the preposition *med* occurring just before a full stop. The lemmatiser had lemmatised this occurrence as *med.* As with the lemma *men* described above, I decided to keep this item as it did satisfy the three criteria applied for the word selection.
5. The item *ringer* was in fact the comparative form, *ringere*, of the adjective *ringe* (poor, bad), but the lemmatiser had lemmatised it incorrectly as *ringer*. Moreover, it was POS-tagged as a noun. This item was replaced with the lemma *ringe* as an adjective. Including *ringe* as lemma with both inflections (*ringere*, *ringest*) in the DAWL may be a misrepresentation in that it was only the comparative form, *ringere*, that was extracted. On the other hand, *ringer* is not a lemma in Danish and in order to represent the comparative use of *ringe* in the DAWL, I decided to include the correct form.
6. The item *svinger* was replaced with the lemma *svinge* as all occurrences were the present tense of the verb *svinge*, and these occurrences were incorrectly lemmatised as *svinger*. A search on the lemma *svinge* shows that this lemma does not occur in the present tense in the corpus. To extract instances of the present tense, one needs to search the lemma *svinger*.
7. The item *varetager* (noun) is replaced with the lemma *varetage* (verb) as *varetager* only occurred in six out of 36 instances as a noun. The rest of the occurrences were in the verbal form.

In total, 145 error items were removed from the list of the 903 lemmas, resulting in a list of 758 lemmas.



#### *6.2.3.5. Summary*

In the preceding sections, I have described the development of the DAWL as depicted in Figure 6.3. The criteria of range, frequency, and dispersion used in the vocabulary selection have been detailed, discussed, and justified. The range criterion extracted all lemmas occurring at least once in the three sub-corpora of Humanities, Social Science, and Natural and Health Science. The frequency criterion was carried out by using a log-likelihood ratio test that extracted all lemmas that occurred more frequently in the academic corpus than in the general language corpus. The dispersion criterion extracted 903 lemmas occurring with a D value of 0.80 or more. Thus, via criteria of range, frequency, and dispersion 903 lemmas occurring more frequently in academic than in general language and with an even distribution were selected for the DAWL. These 903 lemmas were manually inspected for error items and this reduced the DAWL to comprise 758 lemmas.



*Figure 6.3. Development of the DAWL*

Before I describe the procedures for evaluating the DAWL in Section 3.2.3.7., I give an account of how the 758 lemmas were analysed in relation to part of speech.

#### *6.2.3.6. Part of speech analysis of the DAWL*

The part of speech analysis was carried out by manually assigning each lemma a part of speech. The part of speech assignment for each lemma was checked by searching the lemma in the AcaDan Corpus to see its POS-tag(s). As mentioned above, the DAWL lemmas were extracted in such a way that lemmas with more than one part of speech was extracted as one single item. This means that a small number of DAWL lemmas have more than one part of speech (see Sections 6.2.3.7 and 6.4.2). A final

presentation of the DAWL must clearly show which lemmas occur in the AcaDan Corpus with more than one part of speech.

#### *6.2.3.7. Evaluation of the DAWL*

In the development of word lists, evaluation in the form of lexical coverage analyses is used to demonstrate the validity of a given word list. The DAWL is evaluated by measuring its coverage in different corpora. Given that no prior corpus of Danish academic language use existed when this study was carried out, the evaluation of the DAWL was in fact carried out by using the AcaDan corpus, but also a smaller corpus of academic language. A general language corpus was also used. Thus, the DAWL was tested against a general language corpus, the AcaDan corpus, and a second, smaller academic corpus. Information on these corpora is shown in Table 6.1 in Section 6.2.1 and described in more detail in Chapter 4. The vocabulary load analysis programme AntWordProfiler (Anthony, 2014), used in Study 1 to measure the lexical coverage of the 2,000 most frequent lemmas in Danish, was also used in this study. In order to be used in this lexical coverage programme, the DAWL had to be transformed into a so-called base word list. Moreover, to measure the contribution of general high frequency vocabulary to the DAWL's coverage (research question 5), six additional base word lists were developed. The next section describes the development of all seven base word lists.

#### *Developing base word lists for lexical coverage*

In total, seven base word lists were developed on the basis of the DAWL lemmas in order to answer research questions 3 and 5. As described previously, a base word list is organised into groups of items, either in the form of word families or lemmas. In the case of lemma-based base word lists, such as the ones described here, the groups consist of a baseform and inflections. Table 6.12 gives an overview of the seven lists. As can be gathered from Table 6.12, the seven lists can be used to measure the coverage of 1) the entire DAWL, 2) the DAWL minus general high frequency items, and 3) the general high frequency lemmas in the DAWL. Thus, the motivation for separating general high frequency lemmas from the DAWL lemmas was to be able to measure the contribution of the general high frequency lemmas to the coverage of the DAWL. The general high frequency lemmas in the DAWL are identified in Section 6.4.5 to which I refer for the details of the overlap between general high frequency vocabulary and the DAWL.

Table 6.12. Overview of the base word lists developed on the basis of the DAWL lemmas

Name of list	Content	Number of types/tokens
The DAWL	All 758 DAWL lemmas *	758/4,386
The DAWL minus the nine most frequent lemmas	749 DAWL lemmas	749/4,362
The nine most frequent lemmas	9 DAWL lemmas: <i>den, af, som, med, om, denne, eller, sig, anden</i>	9/24
The DAWL minus the 212 academic lemmas from Study 1	546 DAWL lemmas	546/3,125
The 212 academic lemmas from Study 1	212 DAWL lemmas	212/1,270
The DAWL minus the 363 general high frequency lemmas	395 DAWL lemmas	395/2,305
The 363 general high frequency lemmas	363 DAWL lemmas	363/2,091

\* The number of groups, i.e. lemmas in the base word list is 758 because the 12 lemmas with two parts of speech are each listed under one baseform. For example, the lemma *styrke* is listed with both its verbal and nominal inflections.

A number of aspects related to inflections, variance in spelling, and the marking of repeated items required particular attention in the preparation of these seven lists. In the following four sections, I report on these aspects.

## Inflections

Inflections were added to the DAWL lemmas using the DSL full-form lexicon (Det Danske Sprog- og Litteraturselskab, n.d.-b). This full-form lexicon contains 80,000 words and their inflections and was also used in Study 1. Zeros were added after both the baseform of the lemma and the inflections to make them readable for a vocabulary load analysis programme. Figure 6.3 illustrates how the lemma *accept* is organised in the base word list with a baseform (*accept*) and inflections.

ACCEPT 0
ACCEPTS 0
ACCEPTEN 0
ACCEPTENS 0
ACCEPTER 0
ACCEPTERS 0
ACCEPTERNE 0
ACCEPTERNES 0

Figure 6.3. The lemma **accept** in the base word list

In contrast to Study 1, I decided to remove inflections that by the DSL are termed “highly unlikely or unattested but paradigmatically correct.” (Det Danske Sprog- og Litteraturselskab, n.d.-b). An example is the lemma *detaljeret* which in the comparative and superlative is preceded by *mere* (most) and *mest* (most) respectively, and not the comparative and superlative forms suggested by the full-form lexicon: *detaljeredere* and *detaljerederest*. The reason why these forms are listed in the full-form lexicon is that it is an automatically derived list. In addition, I removed unlikely occurrences of the genitive from the list. In Danish, the genitive is formed by adding an ‘s’ to the lexical item, and in the 80,000, most items have a genitive form whether it is idiomatic or not. Thus, removal of the genitive and the comparative and superlative forms was only implemented in the case of highly unlikely forms. If in doubt about the likeliness of a given form, I checked it in KorpusDK (Det Danske Sprog- og Litteraturselskab, n.d.-c), a public general language corpus, to see if the form occurred at all. If it did occur in KorpusDK, the form was kept. Genitive forms of adjectives and verbs were kept if the item in question is also commonly used as a noun. For instance, the lemma *studere* contains the present participle *studerende* which is commonly used as a noun (as in “De **studerende** ser sig selv og deres fag som én disciplin<sup>25</sup>” which translates as “The **students** see themselves and their subjects as one discipline”). A few passives were removed because they are highly unlikely to occur in language use, and KorpusDK had no occurrences of them. An example is the passive of the verb *eksistere* (exist).

## Spelling

The orthography recommended by the Danish Language Council was followed throughout the development of the list. As an example, the plural form of the noun *praksis* (practice) has two spellings, *praksisser* and *praksiser*. Only the first form was included as this is the recommended form. In some cases, the Danish Language Council allows more than one spelling form and, in these cases both forms were included in the list. An example is *nærliggende* which can also be spelled *nærtliggende*. Moreover, the lemma *ressource* can also be spelled *resurse*, and, therefore, the lemma *ressource* contains both spelling forms. Similarly, the lemma *scenario* includes the spelling *scenarie*. The stressed numeral *én* was as a baseform replaced by *een* which is the informal way of writing this numeral because the programme cannot read acute accents.

---

<sup>25</sup> Juul, A., & Vallgård, S. (2014). Tværfaglighed som ideal og praksis – folkesundhedsvidenskabelige erfaringer i at uddanne på tværs af faggrænser. *Dansk Universitetspædagogisk Tidsskrift*, 9(16), 113–123.

## Repeated items

The vocabulary load analysis programme will list all repeated items as errors when performing the analysis. Therefore, it was necessary to mark repeated items with a hashtag in the base word list to make them unreadable for the programme. This was done by running a draft version of the base word list through the vocabulary load analysis programme. The programme detected a number of error items that occurred more than once in the list, either within a group (baseform + inflections) or across groups. For example, the singular and plural indefinite form of a noun could have the same form as in the case of *tiltag* (initiative), or the infinitive and imperative form of a verb could be the same. Also, across lemmas there are examples of repeated items, e.g. the imperative form of the lemma *tiltage* (increase) is the same as the singular and plural indefinite forms of the lemma *tiltag*. Furthermore, 18 lemmas were found to have more than one part of speech. For example, the lemma *evne* is both a noun and verb. These were organised as one lemma and any repeated occurrence across the inflections was marked with a hashtag to avoid an error message when running the vocabulary load analysis programme. I will return to the issue of lemmas with more than one part of speech in Section 6.3. Table 6.13 shows which steps were taken to ensure that the 18 lemmas with more than one part of speech were reflected properly in the lemma list.

Table 6.13. Changes made to the 18 lemmas in the lemma list

Course of action taken	Lemma
No changes. The lemma has no inflections irrespective of part of speech.	<i>om</i> <i>end</i>
Inflections for the adjective were added and any repetitions were marked.	<i>nær</i> <i>endelig</i> <i>samtidig</i>
Inflections for noun were added and any repetitions were marked.	<i>trods</i>
Inflections for both verb and noun were added and any repetitions were marked.	<i>evne</i> <i>følge</i> <i>hvile</i> <i>nytte</i> <i>omtale</i> <i>parallel</i> <i>række</i> <i>sigte</i> <i>styrke</i> <i>stamme</i>
Inflections for both adjective and noun were added and any repetitions were marked.	<i>formel</i>
Inflections for the adjective were added and any repetitions were marked. The noun has no plural form.	<i>vis</i>

### 6.3. Summary

In summary, using the measures of range, frequency, and dispersion, 903 lemmas were selected for a Danish Academic Word List. These 903 lemmas occurred significantly more frequently in the AcaDan corpus than in a comparison corpus of general language (the Journalisten.dk corpus), and they occurred with an even distribution in the four disciplines of Humanities, Natural Science, Social Science, and Health Science as represented in the AcaDan corpus. 144 error items were removed from the list of the 903 lemmas, resulting in a final list of 758 lemmas which will be described in the following chapter. This list of 758 lemmas was analysed in relation to part of speech and expanded with the lemmas' inflections to make it usable for lexical coverage analysis.

### 6.4. Results: A Danish Academic Word List

In this section, the five research questions of Study 2 are answered. The outline of the section follows the five research questions outlined below.

- 1) Which lemmas occur with a higher frequency in academic written language than in non-academic written language and with an even distribution across academic written language and may therefore be identified as academic words and selected for a DAWL?
- 2) Which part of speech categories do the selected DAWL lemmas belong to and in what proportions?
- 3) What is the overlap of the DAWL with Danish general high frequency vocabulary and with the 402 words identified as academic in Study 1?
- 4) What is the coverage of the DAWL in academic language, in different academic disciplines and in general language?
- 5) What is the contribution of the general high frequency DAWL lemmas to the coverage provided by the DAWL in academic language, in different academic disciplines and in general language?

#### 6.4.1. Research question 1: The Danish Academic Word List

A total of 758 lemmas occur with a higher frequency in academic written language than in non-academic written language and with an even distribution across academic written language. These 758 lemmas constitute the Danish Academic Word List (DAWL). However, it may be a misrepresentation to call these 758 lexical items lemmas as some of them have more than one part of speech, as will be shown in the answering of research question 2. As mentioned earlier, the reason for this is that tools used in the vocabulary selection, the Word list function and the Whitelist function in Sketch Engine, do not distinguish between parts of speech. This means that items with more than

one part of speech are extracted as one item, and as such the lemmas extracted for the DAWL are in fact flemmas (Pinchbeck, 2014 in Nation, 2016, p. 26), i.e. sets of lexical items sharing the same stem but with different parts of speech. In the two following sections on the frequency and dispersion of the DAWL lemmas, they are, however, analysed the way they were identified as academic words. This means that the frequencies and dispersion values are given for the 758 lemmas, and, as such, no distinction is made in relation to parts of speech.

#### 6.4.1.1. Frequency of the DAWL lemmas in the AcaDan corpus

Table 6.14 provides an overview of the frequency of occurrence of the DAWL lemmas in frequency belts of 1,000 together with examples. The majority of the lemmas, 83.2%, occur with a frequency below 1,000. Table 6.15 shows how these 632 lemmas are distributed in belts of 100, and Table 6.16 gives examples of lemmas with frequencies below 100 in the AcaDan Corpus. The frequency distribution of the 758 lemmas in the AcaDan Corpus suggests that the methods applied for extracting DAWL lemmas have to a high degree managed to extract items with a frequency above 100 in the corpus: 565 of the 758 lemmas equal to 74.5% have a frequency above 100. 193 or 25 percent out of the 758 lemmas have a frequency below 100 occurrences (see Table 6.14). This should be seen in comparison with Coxhead's Academic Word List (2000) which excluded items with frequencies below 100.

Table 6.14. Distribution of the 758 lemmas in frequency belts of 1,000 across the AcaDan Corpus

Frequency of occurrence	Number of items	%	DAWL lemmas (examples <sup>26</sup> )
>10,000	9	1.19	<i>af, anden, den, denne, eller, med, om, sig, som</i>
9,001-10,000	1	0.13	<i>men</i>
8,001-9,000	0	0	
7,001-8,000	1	0.13	<i>mellem</i>
6,001-7,000	0	0	
5,001-6,000	7	0.92	<i>end, dansk, deres, forhold, forskellig, mere, to</i>
4,001-5,000	2	0.26	<i>lille, vise</i>

<sup>26</sup> Translations of the DAWL lemmas are given in Appendix F. I do, however, provide translations when referring to single words as in Table 6.15.



<b>3,001-4,000</b>	10	1.32	<i>både, del, derfor, form, først, hvilken, hvordan, høj, måde, således</i>
<b>2,001-3,000</b>	32	4.22	<i>analyse, betydning, dermed, dog, eksempel, forbindelse, følge, grad, resultat, sammenhæng, samtidig, skabe, udvikling</i>
<b>1,001-2,000</b>	66	8.71	<i>basere, beskrive, central, erfaring, imidlertid, metode, perspektiv, primær, proces, påvirke, undersøge</i>
<b>6-1,000</b>	630	83.11	<i>afgørende, derimod, diskussion, eksempelvis, forekomme, forklare, funktion, grundlag, indeholde, interview, relevant</i>
<b>Total</b>	758	100%	

Table 6.15. Frequency distribution of the 630 lemmas with a frequency below 1,000 across the AcaDan Corpus

Frequency of occurrence	Number of items	% of the DAWL
<b>901-1,000</b>	19	2.51
<b>801-900</b>	29	3.83
<b>701-800</b>	28	3.69
<b>601-700</b>	27	3.56
<b>501-600</b>	28	3.69
<b>401-500</b>	42	5.54
<b>301-400</b>	67	8.84
<b>201-300</b>	72	9.5
<b>101-200</b>	125	16.49
<b>&lt;100</b>	193	25.46
<b>Total</b>	630	83.11

Table 6.16. Examples of DAWL lemmas <100

Examples of DAWL lemmas <100	Frequency in the AcaDan Corpus	English translation
<i>brugbarhed</i>	6	usefulness
<i>føromtalt</i>	9	before mentioned
<i>redefinere</i>	16	redefine
<i>formodning</i>	37	presumption
<i>betragtelig</i>	65	considerable
<i>Tydeliggøre</i>	93	elucidate
<i>differentiere</i>	96	differentiate
<i>vigtighed</i>	96	importance
<i>opsummere</i>	97	summarise
<i>nytte</i>	99	use
<i>ydermere</i>	99	furthermore

If a frequency cut-off of 100 had been used in the selection of DAWL lemmas, as Coxhead (2000) did, instead of a ratio measure, these 193 low-frequency items would not have been a part of the identified lexical inventory of academic Danish. From 1,001 to 10,000, the number of lemmas decline

and then it increases in the >10,000 frequency belt. This pattern has to do with occurrences of general high-frequency words such as *den* (it), *af* (of), *som* (which), *med* (with), *om* (about), *denne* (this), *eller* (or), *sig* (oneself), and *anden* (another) in the DAWL. These are the nine most frequent items in the corpus. These items are also in the top of the 2,000 most frequent lemmas in Danish (see Chapter 5, Study 1). Among these nine DAWL lemmas, the most frequent item is *den* with a frequency of about 85,00 while the least frequent item within these nine items is *anden*, which occurs roughly 11,000 times. They are identified as academic because they are more frequent in academic language than in non-academic language. It should be noted, however, that some of the high-frequent function words in Danish (e.g. *af*, *anden*, *eller*) were not more frequent in academic than in non-academic language when using the relative frequency ratio measure (see Section 6.2.3.2) which selected all items 1.5 times more frequent in the AcaDan corpus than in the non-academic comparison corpus (Journalisten.dk). Since they were selected by the log-likelihood ratio test, these items should be further analysed in relation to the functions they perform to find the reason why they occur more frequently in academic language, at least according to the statistical test used, and how important they are in academic language. For instance, Durrant (2013, p. 8, 2009) points at ‘this’ as being an important English high frequency word to master as an academic word due to its anaphoric discourse function in academic language.

#### 6.4.1.2. Dispersion of the DAWL lemmas

The 758 lemmas are dispersed in the corpus with D values between 0.80 and 0.98. As can be gathered from Table 6.17, the majority of the DAWL lemmas, almost 82 percent, occur with a D value between 0.80 and 0.89. In contrast, roughly 18 percent have D values between 0.90 and 1.00. This even distribution of the DAWL lemmas signifies that these words are used by academics irrespective of discipline.

Table 6.17. Dispersion of the DAWL lemmas in the AcaDan Corpus

Dispersion	Number of lemmas (percent)	Examples
<b>0.80-0.85</b>	411 (54.22%)	<i>modificere, tilstræbe, yderlig</i>
<b>0.86-0.89</b>	207 (27.31%)	<i>beskeden, ligeledes, interesse</i>
<b>0.90-0.95</b>	133 (17.55%)	<i>betragtelig, mål, beskrive</i>
<b>0.96-0.98</b>	7 (0.92%)	<i>afgøre, bestå, kompliceret</i>
<b>Total</b>	(100%)	

Since the range criterion in this study was a rather simple one (a lemma had to occur in all three sub-corpora at least once), it suffices here to say that the 758 DAWL lemmas occurred in the three

disciplinary sub-corpora used in the range analysis. However, a frequency count in the two sub-corpora of Natural Science and Health Science used in the dispersion analysis showed that the 758 lemmas occur in both of these two disciplines. This indicates that setting a rather simple range criterion (a lemma had to occur in three sub-corpora at least once) and merging two academic disciplines into one sub-corpora (texts from Natural Science and Health Science comprised one sub-corpus in the range analysis in order to have equal-sized sub-corpora) did not hinder the extraction of DAWL lemmas with a wide range. Choosing a high dispersion value of 0.80 no doubt helped ensuring an even distribution of the lemmas.

#### 6.4.2. Research question 2: Parts of speech of the DAWL lemmas

Table 6.18 provides the part of speech distribution of the DAWL lemmas. In Appendix F, the parts of speech are given for each DAWL lemma. In total, eight different parts of speech were found present in the DAWL, with verbs being the most frequent occurring part of speech followed closely by nouns and then adjectives. Abbreviations were not assigned any part of speech but are included in Table 6.18. Because of the way the lemmas were identified as academic, a lemma may have more than one parts of speech. This issue is detailed below in Section 6.4.2.1.

In Table 6.18, the lemmas are categorised according to the part of speech in which they occur most frequently in the AcaDan Corpus. An example is the lemma *evne* (ability, be able to) which occurs as both a noun (ability) and a verb (be able to). Since the nominal use is more frequent in the AcaDan Corpus than the verbal use, it is categorised as a noun in Table 6.18.

*Table 6.18. Distribution of part of speech in the DAWL*

<b>Part of speech</b>	<b># of lemmas – percentages in ()</b>
Verbs	240 (31.66%)
Nouns	238 (31.4%)
Adjectives	182 (24.01%)
Adverbs	57 (7.52%)
Pronouns	11 (1.45%)
Conjunctions	9 (1.19%)
Prepositions	6 (0.79%)
Numerals	6 (0.79%)
Abbreviations	9 (1.19%)
Total	758 (100%)

#### 6.4.2.1. DAWL lemmas with multiple parts of speech

In this section, I describe the DAWL lemmas occurring with more than one part of speech. Twenty-one items were categorised as having multiple parts of speech based on the part of speech analysis as outlined above. Only lemmas occurring as more than one part of speech in the AcaDan corpus were included in this analysis. This means that a lemma such as *central* which can be a noun meaning ‘centre’ and an adjective meaning ‘central’, was not categorised as a lemma with multiple parts of speech because it only occurs as an adjective in the AcaDan Corpus. These 21 items, of which three are homofoms, are displayed in Table 6.19 with frequency figures in brackets after each part of speech. The frequencies of each part of speech are given in brackets. For the homofoms, the different meanings are given in brackets after the candidate.

Table 6.19. The 21 DAWL lemmas with multiple parts of speech

<b>Homoforms</b>	
<b><i>Lemma</i></b>	<b><i>Part of speech</i></b>
<i>formel</i> (formula or formal)	adjective (239) noun (116)
<i>vis</i> (certain or way)	adjective (1,125) noun (318)
<i>stamme</i> (originate or tribe)	verb (202) noun (72)
<b>Multiple parts of speech</b>	
<b><i>Lemma</i></b>	<b><i>Part of speech</i></b>
<i>end</i>	conjunction (5,023) adverb (37)
<i>endelig</i>	adverb (487) adjective (226)
<i>evne</i>	noun (516) verb (12)
<i>følge</i>	verb (2,470) noun (459)
<i>hvile</i>	verb (110) noun (25)
<i>ligesom</i>	conjunction (1,006) adverb (43)
<i>nytte</i>	noun (73) verb (26)
<i>nær</i>	adjective (1,103) preposition (57) adverb (46)
<i>om</i>	preposition (18,908) conjunction (1,834)
<i>omtale</i>	verb (445) noun (105)
<i>overfor</i>	preposition (346) adverb (31)
<i>parallel</i>	adjective (123) noun (87)
<i>række</i>	noun (2,123) verb (76)
<i>samtidig</i>	adverb (1,754) adjective (323)
<i>sigte</i>	verb (211) noun (23)
<i>styrke</i>	noun (352) verb (253)
<i>sådan</i>	adjective (1,510) adverb (1,058)
<i>trods</i>	noun (426) preposition (265)

For most of the lemmas listed in Table 6.19, the different parts of speech are only possible to discern when inflected. As an example, *evne* as a noun is inflected *evnes*, *evner*, *evners*, *evnerne*, *evnernes*, and as a verb *evner*, *evnede*, *evnet*, *evnende*, *evn*. In contrast, *om*, *end*, *trods*, *endelig*, and *samtidig* do not change morphology according to part of speech. As can be gathered from the frequency information in Table 6.19, in some cases, the difference in frequencies between the parts of speech is notable with one part of speech being much more frequent than the other. For example, in the case of *samtidig*, the adverbial use is much more frequent (1,754 hits) than the adjective use (323). A pedagogical argument for listing all 18 lemmas with multiple parts of speech as separate items can be made. In that way, it would be transparent for the users of the word list that there is more than one syntactic use of the lemma. In the case of adjectives, however, it could also be argued that if a learner knows how adjectives can be used as adverbs (commonly by adding the suffix ‘t’), there is no need for listing the adjectival and adverbial forms separately. Ultimately, I decided that those multiple part of speech lemmas that included verbal and nominal uses (e.g. *nytte*, *sigte*) as well as adjectival and nominal uses (e.g. *parallel*, *formel*) should be listed as separate items in the DAWL. This meant that in total, 12 items were added to the DAWL. These are listed together with their parts of speech in Table 6.20.

Table 6.20. The 12 lemmas with are listed with separate parts of speech

DAWL Lemmas	Part of speech	English translation
<i>evne</i>	noun	ability
	verb	to be able to
<i>formel</i>	adjective	formel
	noun	formula
<i>følge</i>	verb	follow
	noun	consequence
<i>hvile</i>	verb	rest
	noun	rest
<i>nytte</i>	noun	usefulness
	verb	be of use
<i>omtale</i>	verb	refer to
	noun	mention
<i>parallel</i>	adjective	parallel
	noun	parallel
<i>række</i>	noun	sequence
	verb	reach
<i>sigte</i>	verb	aim
	noun	aim
<i>stamme</i>	verb	originate
	noun	tribe
<i>styrke</i>	noun	strength
	verb	strengthen
<i>vis</i>	adjective	certain
	noun	way

It should be noted that these 12 items are listed with their different inflections according to part of speech in the DAWL base word list as described in Section 6.2.3.7.

Only two percent of the DAWL lemmas have more than one part of speech, and far from all of these are homoforms in that the meaning is the same irrespective of part of speech. In comparison, Wang and Nation (2004) found that 60 items out of the 570 word families of the AWL were homoforms, and Nation and Parent (2016, p. 42) caution that it is important to mark homoforms, which they use as a cover term for homonyms, homographs, and homophones, in a pedagogical list. In Table 6.21, Nation and Parent's (2016, p. 41) definitions of homonyms, homographs, and homophones are given.

Table 6.21. Nation and Parent's (2016) definition of homoforms

<b>Homoforms</b>		
<i>Homonyms</i>	<i>Homographs</i>	<i>Homophones</i>
words that are spelled and pronounced the same, but have different meanings and are regarded as different lexical units.	words spelled the same but pronounced differently, and which have different meanings and are regarded as different lexical units.	words pronounced in the same way but spelled differently, with different meanings, and which are regarded as different lexical units.

Out of the three homoforms in Table 6.20, two are homographs (*vis*, *formel*) and one is a homonym (*stamme*). Within the 15 items with more than one part of speech, homoforms also occur. As an example, the noun *følge* is a homonym because in the meaning ‘company’, it is inflected differently in the singular definite (*følget*) from *følge* in the meaning ‘sequence’ or ‘consequence’ (*følgen*). In the case of *følge*, AcaDan concordance data showed that the noun *følge* is most often used in phrases like *som følge af* and *i følge* and that the ‘company’ meaning is very infrequent in the AcaDan corpus. Another example is *sigte* (aim) which can also mean ‘sieve’. Based on concordance data, the meaning ‘sieve’ was excluded.

Nation and Parent (2016) recommend that homographs and homonyms be listed as individual items in a word list because they differ in meaning. Therefore, the three homoforms among the DAWL lemmas are listed as separate items. Together with the listing of the nine lemmas described above, this adds 12 items to the DAWL resulting in a final list of 770 lemmas. The DAWL can be seen in its entirety in Appendix F together with translations and parts of speech. The list is frequency-ranked with the most frequent lemma at the top. In addition, DAWL items that overlap with general high frequency vocabulary are marked accordingly. This issue is addressed in Section 6.4.4. Having described the DAWL in relation to frequency, dispersion, and parts of speech, I will now move on to describing the results of the evaluation of the DAWL.

### 6.4.3. Research question 3: Evaluation of the DAWL

To answer research question 3, lexical coverage analyses were carried out using the seven base word lists described in Section 6.2.3.7. For the sake of clarity, information on all seven lists is given in Table 6.22.



Table 6.22. Overview of the base word lists developed on the basis of the DAWL lemmas

Name of list	Content	Number of types/tokens
The DAWL	All 758 DAWL lemmas *	758/4,386
The DAWL minus the nine most frequent lemmas	749 DAWL lemmas	749/4,362
The nine most frequent lemmas	9 DAWL lemmas: <i>den, af, som, med, om, denne, eller, sig, anden</i>	9/24
The DAWL minus the 212 academic lemmas from Study 1	546 DAWL lemmas	546/3,125
The 212 academic lemmas from Study 1	212 DAWL lemmas	212/1,270
The DAWL minus the 363 general high frequency lemmas	395 DAWL lemmas	395/2,305
The 363 general high frequency lemmas	363 DAWL lemmas	363/2,091

\* The 12 lemmas with two parts of speech are each listed as one group in the base word list. For example, the lemma *styrke* is listed with both its verbal and nominal inflections. Consequently, the number of types or groups is 758.

First, the coverage of the DAWL itself in academic and general language is given. Then, the coverages of the other six lists are given. As mentioned in Section 6.2, the motivation for separating general high frequency lemmas from the DAWL lemmas and developing in total seven base word lists was to be able to measure the contribution of the general high frequency items to the coverage of the DAWL.

#### 6.4.3.1. Lexical coverage of the DAWL

Table 6.23 provides the coverage of the DAWL in academic language (the two academic language corpora) and in general language (the general language corpus).

Table 6.23. The coverage of the DAWL over academic and general language

	The AcaDan Corpus	The second academic language corpus	The General Language Corpus
<b>The DAWL</b>	26.09%	27.83%	18.96%

As can be seen in Table 6.23, the DAWL has a higher coverage in academic language than in general language. This implies that the DAWL does in fact reflect a language use different from general language use, a result which is desirable for a word list of academic vocabulary. Similarly, the DAWL provided coverage similar to each other in the four academic disciplines represented in the AcaDan Corpus (see Table 6.24). This further confirms that the DAWL lemmas are in fact academic words.

However, the coverage in the Social Science sub-corpus was approximately three percent higher than those in the other disciplines.

The difference in coverage between the two academic corpora can be explained by the composition of the second academic language corpus. Almost 90 percent of the texts in the corpus stem from the Humanities and Social Science disciplines. The DAWL also has the highest coverage over these two disciplines in the AcaDan corpus (see Table 6.24).

*Table 6.24. The coverage of the DAWL over the four disciplines in the AcaDan corpus*

	Natural Science	Health Science	Humanities	Social Science
<b>The DAWL</b>	24.27%	24.75%	25.65%	28.19%

#### 6.4.4. Research question 4: General high frequency words among the DAWL lemmas

In this section, research question 4 is answered through three separate analyses. These analyses are carried out using the final DAWL of 770 lemmas as described above. First, I report on the results of the overlap analysis between the 402 general high frequency lemmas identified as academic in Study 1 and the 770 DAWL lemmas. Then, I give the results of the overlap analysis between the DAWL lemmas and Danish general high frequency vocabulary in the form of the 2,000 most frequently used lemmas in Danish. Finally, I report on the overlap between the 10,000 most frequently used lemmas in Danish and the DAWL lemmas.

##### 6.4.4.1. Overlap with the 402 general high frequency lemmas of academic nature from Study 1

Almost a third of the DAWL lemmas overlap with these 402 lemmas meaning that a substantial portion of the DAWL lemmas are also general high frequency words. To be precise, more than fifty percent (212) of the 402 lemmas occur in the DAWL. Examples of these are listed in Table 6.25.

*Table 6.25. Examples of the 212 DAWL lemmas that overlap with the 402 lemmas from Study 1*

Overlapping items						
<i>afgørende</i>	<i>central</i>	<i>forekomme</i>	<i>imidlertid</i>	<i>model</i>	<i>oprindelig</i>	<i>resultat</i>
<i>afhængig</i>	<i>danne</i>	<i>foretage</i>	<i>inddrage</i>	<i>modsatning</i>	<i>opstå</i>	<i>retning</i>
<i>afsnit</i>	<i>dels</i>	<i>forhold</i>	<i>indeholde</i>	<i>mulig</i>	<i>organisere</i>	<i>studere</i>
<i>aktiv</i>	<i>dominere</i>	<i>form</i>	<i>indflydelse</i>	<i>mål</i>	<i>pege</i>	<i>styrke</i>
<i>bestå</i>	<i>eksempelvis</i>	<i>grad</i>	<i>konklusion</i>	<i>niveau</i>	<i>proces</i>	<i>supplere</i>
<i>betegne</i>	<i>eksistere</i>	<i>grundlag</i>	<i>konkret</i>	<i>nævne</i>	<i>påvirke</i>	<i>svag</i>
<i>betragte</i>	<i>erfaring</i>	<i>henholdsvis</i>	<i>levende</i>	<i>nødvendig</i>	<i>regel</i>	<i>typisk</i>
<i>betydning</i>	<i>etablere</i>	<i>hensyn</i>	<i>ligeledes</i>	<i>nødvendigvis</i>	<i>relativ</i>	<i>udbrede</i>

#### 6.4.4.2. Overlap with the 2,000 general high frequency lemmas of Danish

Moreover, measuring the overlap between the 2,000 most frequent lemmas in Danish and the DAWL shows that almost half of the items of DAWL are general high frequency items. In total, 363 of the most common 2,000 lemmas in Danish occur in the DAWL. Of these 363 lemmas, 151 do not occur among the 212 lemmas identified above. Examples of these 151 general high frequency items are listed in Table 6.26. In Appendix F, the DAWL lemmas also found among the 2,000 most frequent lemmas in Danish are in bold and italics.

Table 6.26. Examples of general high frequency lemmas among the DAWL

Overlapping lemmas				
<i>acceptere</i>	<i>endelig</i>	<i>konstatere</i>	<i>pågældende</i>	<i>tilsyneladende</i>
<i>afgøre</i>	<i>fastholde</i>	<i>kritisk</i>	<i>reel</i>	<i>udelukkende</i>
<i>afløse</i>	<i>forklaring</i>	<i>medvirke</i>	<i>samtidig</i>	<i>udfordre</i>
<i>bekræfte</i>	<i>gentage</i>	<i>måde</i>	<i>sigte</i>	<i>udpege</i>
<i>beskæftige</i>	<i>gælde</i>	<i>nogenlunde</i>	<i>spørgsmål</i>	<i>understrege</i>
<i>daværende</i>	<i>holdning</i>	<i>optræde</i>	<i>svare</i>	<i>vigtig</i>
<i>diskutere</i>	<i>ifølge</i>	<i>overveje</i>	<i>såkaldt</i>	<i>øvrig</i>

#### 6.4.4.3. Overlap with the most frequent 10,000 lemmas in Danish

The DAWL was also analysed in relation to which frequency bands the 770 DAWL lemmas fall into in the DSL list of the 10,000<sup>27</sup> most frequent lemmas in Danish (Det Danske Sprog- og Litteraturselskab, n.d.-d) (henceforth the Top 10,000 list). As reported above, around 50 percent of the DAWL lemmas are found among the first and second 1,000 bands. Table 6.27 shows to which bands the DAWL lemmas belong.

<sup>27</sup> This is a later version of the list of the 5,000 most frequently used lemmas in Danish used in Study 1, but based on the same corpus. See also Chapter 3.

Table 6.27. The DAWL divided into frequency bands of the DAWL

Frequency band	Number of lemmas	Examples
<b>1-1,000 (K1)</b>	208	<i>årsag, udvikling, sådan, pege, konsekvens, indgå, forsøg, etablere</i>
<b>1,001-2,000 (K2)</b>	157	<i>analyse, betegne, fastholde, hen vise, koncentrere, modsætning, påvirke</i>
<b>2,001-3,000 (K3)</b>	93	<i>anføre, betegnelse, illustrere, kompliceret, sigte, undtagelse</i>
<b>3,001-4,000 (K4)</b>	72	<i>understøtte, samtidig, kriterium, formulering, endog, afsæt</i>
<b>4,001-5,000 (K5)</b>	47	<i>balancere, færdighed, igangværende, redegøre, tilskrive, udforske</i>
<b>5,001-6,000 (K6)</b>	41	<i>afklaring, disposition, evne, observation, parallel, substans</i>
<b>6,001-7,000 (K7)</b>	27	<i>vægte, stadighed, nævneværdig, hvormed, gennemgribende, bevirke</i>
<b>7,001-8,000 (K8)</b>	22	<i>bibeholde, foreskrive, handlekraft, sammenfatte, udsnit, videreudvikle</i>
<b>8,001-9,000 (K9)</b>	16	<i>anvendelig, given, håndgribelig, relevans, skitsere, tvungen</i>
<b>9,001-10,000 (K10)</b>	12	<i>forventelig, indledningsvis, medvirke, uegnet, ønskelig</i>

Note that the 21 lemmas identified in research question 2 as occurring with more than one part of speech occur in more than one frequency band, or more than once within a single frequency band, which is why the total number of items in Table 6.26 is 781. As can be seen in Table 6.27, five of these 21 lemmas (*end, følge, om, sådan* and *trods*) occur twice or three times in the K1 band. In Table 6.19, Section 6.4.2.1, the frequencies of these 21 lemmas' parts of speech were given. Comparing these frequencies with how these lemmas are distributed in the Top 10,000 list, as shown in Table 6.28, shows that in most cases they occur most frequently in the same part of speech whether it is in the AcaDan corpus or in the Top 10,000 list.

Table 6.28. Comparison of frequencies for DAWL lemmas with more than one part of speech

DAWL lemma	Frequency bands	Parts of speech (ranked after frequency in Top 10,000)	Parts of speech (ranked after frequency in the AcaDan corpus)
<i>end</i>	Twice in K1	Conjunction, adverb	Conjunction, adverb
<i>endelig</i>	K1 + K2	Adverb, adjective	Adverb, adjective
<i>evne</i>	K1 + K6	Noun, verb	Noun, verb
<i>formel</i>	K3 + K4	Adjective, noun	Adjective, noun
<i>følge</i>	K1 + K2	Verb, noun	Verb, noun
<i>hvile</i>	K3 + K7	Verb, noun	Verb, noun
<i>ligesom</i>	K1 + K2	Conjunction, adverb	Conjunction, adverb
<i>nytte</i>	K4 + K6	Verb, noun	Noun, verb
<i>nær</i>	K1 + K2 + K4	Adjective, preposition, adverb	Adjective, preposition, adverb
<i>om</i>	K1	Preposition, conjunction, adverb	Preposition, conjunction
<i>omtale</i>	K2 + K4	Verb, noun	Verb, noun
<i>overfor</i>	K2 + K6	Adverb, preposition	Preposition, adverb
<i>parallel</i>	K6 + K7	Adjective, noun	Adjective, noun
<i>række</i>	K1 + K2	Noun, verb	Noun, verb
<i>samtidig</i>	K1 + K4	Adverb, adjective	Adverb, adjective
<i>sigte</i>	K2 + K3	Verb, noun	Verb, noun
<i>stamme</i>	K2 + K4	Verb, noun	Verb, noun
<i>styrke</i>	K1 + K2	Noun, verb	Noun, verb
<i>sådan</i>	K1	Adverb, adjective	Adjective, adverb
<i>trods</i>	K1	Preposition, noun	Noun, preposition
<i>vis</i>	K1 + K2	Adjective, noun	Adjective, noun

#### 6.4.5. Research question 5: Lexical coverage of the general high frequency words in the DAWL

The relatively high coverage of the DAWL irrespective of language type (academic or general) can be explained by the fact that the DAWL comprises roughly 50 percent general high frequency vocabulary as shown in the overlap analyses with general high frequency vocabulary in Section 6.4.4. The six lexical coverage analyses carried out in order to answer research question 5 confirm this explanation. Table 6.29 provides the coverages of 1) the DAWL, 2) the DAWL without the nine most frequent general high frequency lemmas in the AcaDan Corpus, 3) the nine most frequent general high frequency lemmas in the AcaDan Corpus, 4) the DAWL without the 212 overlapping general high frequency lemmas from Study 1, 5) the 212 overlapping general high frequency lemmas from

Study 1, 6) the DAWL minus the 363 general high frequency lemmas, and finally the coverage of 7) the 363 general high frequency lemmas.

*Table 6.29. The coverage of the DAWL with and without general high frequency items*

	<b>List</b>	<b>The AcaDan Corpus</b>	<b>The second academic language corpus</b>	<b>The General Language Corpus</b>
1	The DAWL	26.09%	27.83%	18.96%
2	The DAWL minus the nine most frequent lemmas	14.80%	16.25%	8.80%
3	The nine most frequent lemmas	11.29%	11.58%	10.17%
4	The DAWL minus the 212 academic lemmas from Study 1	15.60%	17.73%	14.39%
5	The 212 academic lemmas from Study 1	10.54%	10.14%	4.59%
6	The DAWL minus the 363 general high frequency lemmas	2.96%	3.06%	1.56%
7	The 363 general high frequency lemmas	23.22%	24.88%	17.53%

The coverage of the DAWL minus the nine most frequent lemmas is about ten percent lower than the coverage of the DAWL regardless of language type. Note that the list containing the removed nine items (*den, af, som, med, om, denne, eller, sig, anden*) provides somewhat similar coverage over the two language types. In contrast, the DAWL minus the nine most frequent lemmas differs in coverage dependent on language type, which indicates that these nine high frequent lemmas are important for academic language. The coverage of the DAWL minus the 212 academic lemmas from Study 1 is similar to the coverage of the DAWL minus the nine most frequent lemmas in the academic language corpora as the coverages in these two corpora are reduced by about ten percent. The coverage result of the 212 academic lemmas from Study 1 demonstrates that it is the remaining DAWL lemmas, and not the 212 general high frequency items, that provide most of the coverage of the DAWL. The coverage of the DAWL minus the 363 general high frequency lemmas provided us with even more information on the coverage provided by the DAWL. Separating the 212 overlapping Study 1 lemmas from the 363 general high frequency items results in 151 lemmas. Isolated, these 151 lemmas provide a coverage of 12.68% in the AcaDan Corpus which is higher than that of the 212 academic lemmas from Study 1 (10.54%). It is six out of the nine most frequent DAWL lemmas (three of them, *af, denne, and eller*, are among the 212 academic lemmas from Study 1) that contribute to the 12.68% coverage of the 151 lemmas.

As can be seen in Table 6.30, a similar pattern can be detected when looking at the coverage of the DAWL in the four academic disciplines with the high frequency lemmas influencing the coverage numbers considerably. While the coverages of the different lists in the four disciplines are higher than that of general language, the disciplines also differ from each other. The DAWL, whether or not it includes general high frequency lemmas, offers a higher coverage of the Humanities and the Social Science disciplines than of the Natural and Health Sciences. As in Study 1, it seems that the general high frequency vocabulary provides the highest coverage within the Humanities and Social Sciences indicating that the language use of these disciplines is closer to general professional writing, such as journalist language, than that of the two other disciplines which make use of more specialised language.

*Table 6.30. Coverage of the DAWL with and without general high frequency lemmas in the four academic disciplines of the AcaDan Corpus*

	<b>Social Science</b>	<b>Humanities</b>	<b>Health Science</b>	<b>Natural Science</b>
The DAWL	28.19%	25.65%	24.75%	24.27%
The DAWL minus the nine most frequent lemmas	16.21%	13.96%	14.73%	13.83%
The nine most frequent lemmas	11.99%	11.69%	10.02%	10.45%
The DAWL minus the 212 academic lemmas from Study 1	16.95%	16.56%	13.43%	13.32%
The 212 academic lemmas from Study 1	11.32%	9.14%	11.36%	10.99%
The DAWL minus the 363 general high frequency lemmas	3.17%	2.86%	2.91%	2.8%
<b>The 363 general high frequency lemmas</b>	<b>25.15%</b>	<b>22.88%</b>	<b>21.92%</b>	<b>21.52%</b>

The findings from the lexical coverage analyses carried out to answer research questions 4 and 5 highlight the fact that general high frequency vocabulary comprises a significant portion of the DAWL, and that the most frequently used lemmas in Danish contribute notably to the coverage of the DAWL. On the other hand, since these general high frequency lemmas occur significantly more frequently in the AcaDan Corpus than in the general language corpus of Journalisten.dk, they arguably perform important academic functions in academic language use. In the next section, the specifications of the DAWL are given based on Nation's (2016, pp. 131–143) framework for critiquing a word list with some adjustment to the methodology behind the DAWL.

#### 6.4.6. Specifications of the Danish Academic Word List

Table 6.31 gives the specifications of the DAWL based on Nation (2016). The last specification, “Possible criticism”, will be discussed in Section 6.5.1. As can be seen, the size of the list is set to 770 lemmas because 12 lemmas with more than one part of speech are separated into independent items as reported in Section 6.4.2.1.

*Table 6.31. Specifications of the DAWL*

<b>Purpose</b>	Description of Danish academic vocabulary Teaching of academic vocabulary
<b>Size of list</b>	770 lemmas
<b>Organisation of the list</b>	The DAWL can be organised according to dispersion, frequency, part of speech, general high frequency items, and alphabetically.
<b>Unit of counting</b>	Lemma
<b>Capitalisation</b>	No distinction between capitalised and non-capitalised words
<b>Numbers</b>	No numbers, but numerals are included
<b>Criteria for list</b>	Range, frequency, and dispersion: The range criterion extracted all items occurring at least once in the three sub-corpora of Humanities, Social Science, and Natural and Health Science. The frequency criterion was carried out by using a log-likelihood ratio test that extracted all items more frequent in the academic corpus than in the general language corpus. The dispersion criterion extracted all items occurring with a D value of 0.80 or more.
<b>Removed items</b>	143 items: English words, proper nouns, words lemmatised or tokenised incorrectly, formatting items, single letters, and abbreviations
<b>General high frequency items</b>	The DAWL contains 363 lemmas from the 2,000 most frequent lemmas in Danish. Moreover, almost 90% of the words on the list are found within the 10,000 most frequent lemmas of Danish.
<b>Part of speech categorisation</b>	Yes
<b>Corpus</b>	The AcaDan Corpus of 3.3 million words was used for vocabulary selection.
<b>Possible criticism</b>	Corpus size and composition: representation of disciplines and subject areas; no textbooks, only research articles and reports. Not validated in a corpus of the same size as the AcaDan Corpus. Extraction methods.

In the preceding sections, I have given the results of the identification and vocabulary selection analyses described in Section 6.2, and described the identified academic vocabulary in Danish in



relation to frequency of occurrence in the AcaDan Corpus, dispersion, part of speech, and overlap with general high frequency vocabulary. In doing so, I have shown how the DAWL represents Danish academic vocabulary. In the next and penultimate section of Chapter 6, I will discuss the findings of Study 2 in comparison with academic word lists in other languages as described in Chapter 3, and highlight how Study 2 has contributed to our knowledge of Danish academic language and vocabulary. I will also report on the limitations of Study 2. I conclude Chapter 6 with a summary and by giving the rationale for Study 3.

## 6.5. Discussion

In this discussion, I will focus on two themes that are closely related and therefore they will be discussed together. The first theme is related to the issue of general high frequency vocabulary and its overlap with academic vocabulary, which has been a continuous focus area of this thesis. The second theme concerns methodological choices made in the identification and selection of words for the DAWL and the limitations related to these choices. Before discussing these two themes, I will discuss the other main findings of Study 2.

Study 2 has expanded our understanding of Danish academic vocabulary in a number of ways. Firstly, Study 2 identified 770 lemmas as academic and these were selected for the DAWL. These were identified in the AcaDan corpus using measures of range, frequency and dispersion. The initial result of these measures was a list of 758 lemmas after the removal of 143 error items (English words, proper nouns, annotation and tagging errors). The list of 758 lemmas was analysed in relation to their frequencies in the corpus and in relation to part of speech, and overlap with general high frequency vocabulary. With regard to the frequency distribution of the DAWL, 83 percent of them occurred with frequencies between 6 and 1,000, and around 200 lemmas had a frequency below 100. The frequency analysis revealed that nine items occurred with frequencies above 10,000 up to 85,000. These were general high frequency function words. Moreover, the DAWL consists of 47 percent general high frequency vocabulary (363 lemmas). In addition, 28 percent of the academic words (212 items) identified in Study 1 occurred in the DAWL. The fact that almost half of the items in the DAWL are general high frequency lemmas supports the notion that general high frequency vocabulary is an important component of academic vocabulary. The issue of general high frequency vocabulary and its overlap with the DAWL words will be discussed further below.

Secondly, Study 2 has expanded our understanding of Danish academic vocabulary in relation to the syntactical properties. The part of speech analysis carried out in relation to research question 2

uncovered that the DAWL, in fact, comprises 770 lexical items. The reason for this was that a small number of the lemmas occurred in two parts of speech with similar frequencies in the AcaDan Corpus which justified separating these items according to parts of speech in the DAWL. Further, it demonstrated that the majority of the DAWL words were nouns and verbs, 31 percent each. This finding stands in contrast to the part of speech distribution of the Swedish Academic Word List (Ribeck et al., 2014) which is similar to that of the Academic Key Word List (Paquot, 2010). In these lists, the most frequent word class was nouns comprising 42 percent in the SAWL. Verbs constituted 26 percent of the SAWL. Also, in contrast to especially the SAWL is the proportion of adjectives in the DAWL. They comprise 24 percent in the DAWL and only 14 percent of the SAWL. In the AKL, 19 percent of the items are adjectives. Paquot (2010, p. 59) relates this relatively high occurrence of adjectives to a substantial number of nouns in the AKL as adjectives modify nouns. The same explanation can be given in relation to the DAWL. The analyses of Study 3 will shed more light on how the issue of part of speech is related to the functions of academic vocabulary, e.g. in relation to referring to academic activities and stance setting.

The evaluation of the DAWL showed a lexical coverage of more than 25 percent in academic language and 18 percent in general language. Excluding the nine most frequent items from general high frequency vocabulary reduced the coverage with about ten percent. Unsurprisingly, the 363 general high frequency items in the DAWL contributed to the list having a high coverage, but the differences in coverage between the two language types analysed support that the DAWL is representative of academic vocabulary. The somewhat similar coverage figures of the DAWL in the four academic disciplines represented in the AcaDan Corpus supports the claim that the measures applied in the development of the DAWL have succeeded in identifying words occurring across academic disciplines. Furthermore, the disciplinary coverage differences of the DAWL echo the coverages provided by the Academic Word List (Coxhead, 2000) in that this list had a notable higher coverage (12 percent) in the disciplinary sub-corpus of Commerce compared to the other three sub-corpora of Arts, Law, and Science. Here, the AWL covered about nine percent in each.

Study 2, in line with Study 1, highlights the important relationship between academic vocabulary and general high frequency vocabulary, a relationship that has been of central interest to academic word list developers as shown in Chapter 3. The coverage of the DAWL in general language (19 percent) compared to that of especially the first 1,000 most frequent words (76 percent, see Study 1) in Danish supports the claim made that the DAWL is representative of a specialised vocabulary. On the other hand, comparing the DAWL coverage to that of the second 1,000 most frequent words in Danish

emphasises the role that the 363 general high frequency words play in relation to coverage of the DAWL. Also, when comparing with the coverages of other academic word lists, the issue of general high frequency vocabulary is important to consider. Compared to the Swedish and Norwegian academic word lists which had coverages of approximately eight percent in academic language corpora (Johansson et al., 2017, p. 154), the coverage of the DAWL is relatively high. The same is the case when comparing the DAWL's coverage to that of Coxhead's (2000) Academic Word List (8-10 percent), and also Gardner and Davies' (2014) Academic Vocabulary List even though the latter list has a notably higher coverage (almost 14 percent) than the AWL. The issue of general high frequency vocabulary, and in particular the proportion of it in the DAWL and in other academic word lists, are discussed further in the next paragraph.

As shown in Section 6.4.4, 363 out of the 770 lemmas can also be found among the general high frequency vocabulary of Danish as represented by the 2,000 most frequently used lemmas in Danish from the DSL list of the 5,000 most used lemmas in Danish. Only 86 lemmas are outside the 10,000 most used lemmas in Danish as demonstrated in Section 6.4.4. The DAWL is comparable to Paquot's (2010) Academic Key Word List in its percentage of general high frequency vocabulary. The AKL comprises approximately 47 percent words also occurring in the General Service List (West, 1953). Gardner and Davies (2014) include an interesting discussion about the issue of high frequency vocabulary in the Academic Word List (Coxhead, 2000) initiated by a study they carried out in 2010 in which they demonstrated that 79 percent of the Academic Word List word families occur within the first four frequency levels of the COCA (Davies, 2008). As they argue, and which has been demonstrated in relation to the DAWL above in Section 6.4.5, it is exactly because of the relatively high proportion of general high frequency words in the Academic Word List that it provides good coverage of academic language. This argument resonates with Cobb and Horst (2004) findings for French which showed that the first 2,000 words of French provided a substantial higher coverage of academic language (81.27%) than did the first 2,000 words of English academic language (70.42%). Based on these findings, Cobb and Horst argue that French academic language makes high use of the most frequently occurring words in French. The question that Cobb and Horst's research leaves us with is whether these general high frequency words have separate academic meanings and functions which can explain their use in academic language. Or if they occur in academic language because they are used in their general meanings. After all, general high frequency vocabulary is general vocabulary because of its broad range in different text types. However, as has been argued by Durrant (2009, 2013), the general high frequency vocabulary may serve different functions in academic

language than in general language, an argument that again highlights the importance of focusing on the issue of polysemy in the development and use of word lists.

From what has been discussed here so far, two issues are important to consider in connection to general high frequency vocabulary as academic vocabulary. The first issue has to do with polysemy. The fact that general high frequency words can be identified as academic because of their higher frequency in academic language than in non-academic language has implications both for how we categorise the words of academic texts as well as for how we teach these words. As discussed in Study 1 and also argued by Durrant (2013), among others, learning general frequency vocabulary is by no means an easy task precisely because of the wide range and the polysemous nature of general vocabulary. The academic nature of these words echoes the issue of pre-technical words which especially Nordic studies of vocabulary have focused on primarily from a learning perspective (referencer). The objectives of these studies have been to identify which general words in their technical senses are not understood by learners. There is reason to argue that the same perspective could be applied to the general high frequency words among the academic vocabulary of Danish (or of any language). However, in order to so, we need to explore more thoroughly the academic nature of these words not only in relation to coverage, but also to their functions in the academic text. This is partly done in Study 3 which investigates the functions of the DAWL words, and, thus, also the use of academic general high frequency words.

The second issue related to the occurrence of general high frequency vocabulary in the DAWL is primarily a methodological one. The investigation of the coverages of general high frequency vocabulary in Study 1 laid the foundation for choosing the appropriate identification and selection criteria for the DAWL. In this respect, I followed what other academic word lists developers have experimented with, i.e. comparative frequency measures such as ratio and keyness. As described in Chapter 3, the Swedish academic word list (Ribeck et al., 2014), and to some degree the Norwegian Academic Word List (Hagen et al., 2016), used stop lists of general high frequency vocabulary in order to operationalise the first principle of creating academic word lists: “Exclusion of high frequency non-academic words that can be found in every corpus.” (Johansson et al., 2017, p. 151). This principle was not operationalised in the vocabulary selection for the DAWL. This is why the DAWL includes very high frequency items such as *af*, *anden*, and *den*. An argument can be made that such words do not have specific academic senses and therefore should be removed via a stop list. Using a list of the 1,000 most frequent words in the language as Ribeck et al. would risk eliminating important academic words, however. Therefore, special concern should be given when using a stop

list that only eliminates very general high frequency items, an issue that Hagen et al. (2016) discuss in some detail. As can be seen from Section 6.4.4, words like *årsag*, *pege*, and *forsøg* (see also Table 6.26) would not have been included in the DAWL if a stop list of the first 1,000 words in Danish had been used.

### 6.5.1. Limitations

Study 2 has a number of limitations related to the decisions made in the design and compilation of the corpora used for identifying and selecting words for the DAWL. These will be discussed now. First of all, I have argued that the DAWL is representative of Danish academic vocabulary due to its coverage in academic language compared to general language. However, it can be argued that by only including professional academic writing in the AcaDan Corpus, the DAWL only represents this form of academic language use. This is an argument that I would agree with in full if the sole purpose of the DAWL was to function as a pedagogical word list aiding e.g. university L2 students in academic reading. If that was the case, the AcaDan corpus should have contained the kind of texts that students meet in higher education, i.e. textbooks. However, the primary aim of developing the DAWL was to identify which words can be said to be core academic words in Danish and for that reason, as described in Chapter 4 on corpus design, professional academic writing was chosen as the prototypical kind of academic language. Admittedly, there is also a practical side to this as also explained in Chapter 4. Compiling a large, representative corpus of academic language takes time and resources, and I needed to focus on what could be done within the scope of this PhD project. Additionally, the size of the AcaDan Corpus can be questioned. In comparison with e.g. the corpora behind the Swedish and Norwegian academic word lists (Hagen et al., 2016; Ribeck et al., 2014), the AcaDan Corpus is rather small. However, as proven by Paquot (2010), also small corpora can be used for identifying academic vocabulary. Finally, as laid out in Chapter 4, it was not possible to compile a corpus of academic Danish language use resembling in size the corpora used in other academic word list studies due to the fact that no Danish corpora of academic language existed beforehand.

Another criticism that can be raised against the DAWL concerns the applied measures used in the identification and word selection analyses. Here I will address the issues of range and dispersion and the way the AcaDan corpus was divided according to disciplines. I will also touch upon the issue of using a comparative frequency measure and the importance of choosing a suitable comparison corpus. Firstly, the range criterion applied in the development of the DAWL is a rather simple one. A DAWL lemma has to occur at least three times in the AcaDan corpus and at least once in each of the three

disciplinary sub-corpora of Natural and Health Science, Social Science, and Humanities. The argument for measuring range in this way was that the dispersion measure should ensure that the words selected would be evenly dispersed in the texts of the corpus. Moreover, the way the different disciplines and especially sub-disciplines are represented in non-equal portions in the AcaDan Corpus made it difficult to employ a range criterion such as the ones used by Coxhead (2000) and Dang et al. (2017) which required equal-sized sub-corpora of disciplines and subject areas for measuring range. Moreover, in relation to range and dispersion, the decision to use different sub-corpora divisions for each measure could be said to affect the validity of the resulting list. Secondly, it can be argued that I have relied too much on the dispersion criterion to make up for the simplicity of the range criterion, especially considering that the Juilland's D measure has received some criticism in recent years as it has been found to decrease in sensitivity when used on corpora with many parts (Biber et al., 2016). It could be that some of the DAWL words are not as evenly distributed in the AcaDan Corpus as assumed. In addition, the dispersion criterion was set at 0.80 which, as described in Section 6.2.3.3, resulted in a number of potential academic words being excluded. These words are further investigated in Study 4. The motivation for Study 4 thus lies in the fact that the decisions made in the dispersion analysis are based on experiments and intuition and not as in Dang et al. (2017; see also Dang, 2017) on the lexical coverage of different pilot lists with different dispersion values.

Despite the limitations discussed here, Study 2 has provided us with new insight into the nature of Danish academic vocabulary, primarily through the identification of 770 lemmas which represent a core list of Danish academic vocabulary. An important implication of the findings in Study 2 that resonates with the findings of Study 1 is that the categories of general, academic, and technical vocabulary into which we divided the words of academic discourse are indeed overlapping.

## 6.6. Summary and rationale for Study 3

This chapter has focused on the study of vocabulary selection for a Danish Academic Word List using a quantitative, corpus-based approach. Study 2, through its corpus-based identification and selection analyses, has provided us with a description of a specific lexical inventory of Danish vocabulary. This inventory is represented by the Danish Academic Word List, the DAWL, which comprises 770 lemmas. Its status as a representation of Danish academic vocabulary is proven by its relatively high coverage of 26 to 28 percent in the two academic language corpora as shown in Section 6.4.3.

Study 2 thus presented a list of Danish academic vocabulary and described this list in relation to frequency, dispersion, part of speech, and general high frequency vocabulary. The overall question

that guides the next study of this thesis, Study 3, is what it is that makes these DAWL words academic in a functional perspective, providing two functional classifications of the DAWL items.

## Chapter 7. Study 3 – Exploring the functions of the DAWL words

### 7.1. Introduction and research question

This chapter presents Study 3 in which I analysed the 770 DAWL words according to the functions they perform in the academic text. The purpose of this functional analysis was to move beyond the simple frequency occurrences of academic words in academic texts in order to reach an understanding of the actual use of the words, and thus a function-based qualification of why the DAWL words constitute an academic vocabulary in Danish. Such an understanding is not only useful for our general understanding of Danish academic vocabulary but also for the teaching of the DAWL words, especially for productive purposes as it will enable teachers to show how these words are integral for creating the scientific discourse used in academic texts. In Danish, Study 3 presents the first functional categorisation of words extracted from an academic language corpus since earlier categorisations (e.g. Stray Jørgensen, 2004) have been based on more intuitive suggestions as posited in Chapter 2.

The functional analysis of the DAWL words is guided by one research question:

- 1) What functions do Danish academic words perform?

Two functional analyses were carried out. First, a macro-categorisation of the words primarily based on the pragmatic-functional classifications employed in academic multi-word unit studies (e.g. Ädel & Erman, 2012; Chen & Baker, 2010; Simpson-Vlach & Ellis, 2010; Biber et al., 2004) and described in Chapter 2. This analysis served as foundation for the second functional analysis which took its point of departure in the framework developed by Hirsh (2004) for functional categorisation of the lexical items in the Academic Word List (Coxhead, 2000). An important assumption behind Hirsh's framework is that academic words are used by academics irrespective of discipline because academic words have certain functions that are to a high degree subject-independent, and are used to perform different functions related to the **textual**, **ideational**, and **interpersonal** levels of the academic texts (Halliday, 1976). Hirsh (2004), based on comprehensive analyses of the occurrences of Academic Word List (Coxhead, 2000) words in different academic texts, established seven functional categories related to these three levels of the text: **metatextual**, **extratextual**, and **intratextual** (textual level); **scholarly processes, states of affairs, and relations between entities** (ideational level) and **authoritative** (interpersonal level). Analysing the DAWL words according to academic functions resulted in an expansion of Hirsh's analytical framework, which will be elaborated on in Section 7.2.2 below. Most notably, three categories of **technical words**, **phrasal elements** and **stance** were added.



The category of technical words relates to overlap zone II in the vocabulary circle shown in Figure 2.1, and it highlights the notion that academic words can take on technical senses in specific contexts as also shown by Hyland and Tse (2007) and Chung and Nation (2004). The category of phrasal elements points to the fact that some words in the DAWL also occur in formulaic sequences. The stance category was added as a means to elaborate on how academic words are used for carrying out the interpersonal metafunctions of the academic text.

It should be noted already here that the analyses carried out in Study 3 showed that most DAWL words have multiple functions in the academic text. Moreover, the analyses of Study 3 were carried out by looking at the DAWL words outside of context and then finding examples for this somewhat intuitive analysis in the AcaDan Corpus. I will return to this issue in the discussion section of this chapter. Before reporting further on the study itself, I will outline the structure of this chapter.

The majority of this chapter is dedicated to the analyses of the DAWL words according to their functions. The analysis section, Section 7.2, first reports on the findings from the first macro-analysis, and then moves on to the more fine-grained analysis carried out. The chapter concludes with a discussion of some of the issues related to analysing the functions of academic vocabulary, and a rationale for the fourth and last study of this thesis.

## 7.2. A functional categorisation of the DAWL words

In this section, I describe how the analyses of Study 3 were carried out.

### 7.2.1. The first analysis: Macro categories

A macro-categorisation of the DAWL words was carried out before applying Hirsh's more fine-grained framework. Three macro categories were established on the basis of the different pragmatic-functional classifications described in Chapter 2, and the 770 DAWL words were assigned to these categories. The categorisation was checked by another researcher and a second analysis was carried out based on the feedback I received from her. The three macro categories were:

1) **Words related to the academic workflow.** This category was derived from Martin's (1976) 'vocabulary of the research process', Stray-Jørgensen's (2004) 'investigation word category', and the concept of academic language functions (Bailey et al., 2007). This category contained primarily verbs and nouns denoting academic activities and processes. Examples of these are *visualisere* (visualise), *udvikle* (develop), *tilstræbe* (aim at), *medvirke* (contribute), *aktivitet* (activity), *fastholdelse* (insistence), *krav* (demand, n), *overensstemmelse* (accordance), and *årsag* (cause).

2) **Descriptive words.** This category contained two sub-categories: evaluation and neutral. The first sub-category was similar to the Evaluation category in Simpson-Vlach and Ellis's (2010) functional classification of academic formulas (see Chapter 2). This first sub-category contained adjectives and adverbs that show the author's stance to the content. Examples of evaluative descriptive words are *afgørende* (crucial), *bemærkelsesværdig* (notable), *entydig* (unambiguous), *forventelig* (probable), *hensigtsmæssig* (appropriate), and *relevant* (relevant). The second sub-category included all neutral adjectives and adverbs. Examples of neutral descriptive words are *aktuel* (current), *dansk* (Danish), *enkelt* (single), *forskellig* (different), *igangværende* (ongoing), and *potentiel* (potential).

3) **Discourse-organising words.** Similar to the category of Descriptive words, this category was based on the pragmatic-functional taxonomy of Biber, Conrad, and Cortes (2004). This category contained primarily conjunctions and adverbs, but also some nouns. Examples of these are *afsnit* (section), *derimod* (on the contrary), *endog* (even), *føromtalt* (before-mentioned), *indledningsvis* (by way of introduction), and *således* (thus).

From the categorisation of the DAWL words into these three macro categories, three other categories emerged for around 70 of the DAWL words, as mentioned above. These words exhibited properties that fall outside the categories listed above. A portion of them was categorised as having a technical sense and are analysed further in Section 7.2.2.8. Others are elements in phrases, and these are described in Section 7.2.2.9. Finally, a group of words seemed to describe the institutional settings of research and education. As will be shown below, these words fall into one of the categories set forth by Hirsh (2004, 2010) and are thus described in Section 7.2.2 below.

### 7.2.2. The second analysis: A functional framework of Danish academic vocabulary

To reach a deeper understanding of what it is that makes the DAWL words academic besides their occurrences in academic texts and their broad academic functions as described above, a second analysis focusing even more closely on the functions they perform was necessary. For this purpose, Hirsh's framework for functional categorisation as described in Chapter 2 was applied as a point of departure. Table 7.1 provides the functional framework developed by and printed in Hirsh (2004, p. 52).

Table 7.1. Framework for the functional categorisation of academic vocabulary (adapted from Hirsh, 2004)

Metafunction	Category	Sub-categories	Examples AWL item(s) in bold and source text in italics
textual	metatextual	headings	<b>Methods</b> We studied two pairs of villages ( <i>Med1</i> )
		in-text cues	In this <b>section</b> , we will consider some more complicated cases. ( <i>Econtext1</i> )
	extratextual	previous research	Unfortunately, most <b>previous</b> work on integration has focused on the distribution of race within schools, implicitly assuming that...( <i>Soc1</i> )
		other bodies	After discussion with the Côte d'Ivoire <b>Ministry</b> of health, enrolment was stopped... ( <i>Med2</i> )
		borrowed methods/materials <sup>28</sup>	Our statistical analyses used Epi Info ( <b>version</b> 6) and SPSS ( <b>version</b> 6.1). ( <i>Med3</i> )
		ethical consent	After obtaining informed <b>consent</b> , we recorded the yellow fever vaccination status...( <i>Med3</i> )
		further research	We leave this puzzle to further <b>research</b> ( <i>Econ1</i> )
	intratextual	conjunctions	<b>Nonetheless</b> , the prospect of rapid global improvement remains bleak. ( <i>Med1</i> )
		carrier words	We test this <b>assumption</b> using a Hausman test, adding the residuals from a Probit model of regime choice to the regressions. ( <i>Econ1</i> )
ideational	scholarly processes		We <b>assessed</b> the effect of fly control on public health in a pilot study in Gambian villages. ( <i>Med1</i> )
	state of affairs	context	Because the exact date of death was not <b>available</b> for some patients, we used statistical methods for interval censored survival data. ( <i>Med2</i> )
		setting (location/period)	(The regressions include the dummies SSA for Sub-Saharan Africa, ASP for the Asia-Pacific <b>region</b> and MDE for the Mid-East...( <i>Econ1</i> )
		participants	Loss to follow up, mainly owing to inclusion of <b>temporary immigrants</b> in the baseline data, was similar for intervention and control groups... ( <i>Med1</i> )
		characteristics	While there is a clear structural division between two groups, this division is based on <b>grade</b> not race. ( <i>Soc1</i> )
	relations between entities		Disulfide bonding between cysteines in different polypeptide chains of oligomeric proteins plays a <b>crucial</b> role in ordering the structure of complex proteins... ( <i>Chemtext1</i> )
interpersonal	authoritative		We <b>undertook</b> a pilot study to investigate the role of domestic flies in the transmission of trachoma and diarrhoea. ( <i>Med1</i> )

<sup>28</sup> I have added "materials" to Table 7.1 since Hirsh (2004) calls the category "borrowed materials" on p. 57, but "borrowed methods" in the table on p. 51.

First, the macro-categories outlined in the previous section were compared to Hirsh's framework. The first and second categories, **words related to the academic workflow and descriptive words**, correspond broadly to Hirsh's three ideational categories: **states of affairs**, **relation between entities**, and **scholarly processes**. The third category, **discourse-organising words**, relates to the **textual** metafunction in Hirsh's framework, which focuses on how the writer organises the text and guides the reader through it.

Secondly, the analysis of the DAWL words according to Hirsh's framework revealed the need for adjusting it. In the following, I will describe the changes made and the motivation behind them. The final analytical framework applied in this analysis is provided in Table 7.2 below. Within the **textual** metafunction, an extra sub-category, **metacommunication**, was added to the **metatextual** category as a number of DAWL words are used for communicating to the reader what the next chapter, section, or paragraph centre on. The five subcategories of the **extratextual** category were reduced to four in that **previous research** and **further research** were merged into one category. This was done due to a notable overlap between the DAWL words falling into these two sub-categories. The sub-category of **other bodies** was expanded and renamed as **institutional** because the macro-analysis carried out prior to this analysis showed that the DAWL contains a small number of words that fall outside the functional-pragmatic categories used in that analysis. They are academic words that are used across education and research settings to describe the activities that are central to research and education such as 'lecture', 'university', 'curriculum', and 'research project'. Thus, the **other bodies**' sub-category was renamed 'institutional' to signal that it not only contains words related to research but also to education. Also within the **textual** metafunction, the **intratextual** sub-category **conjunctions** was renamed so that not only conjunctions could fall into this sub-category since many words classified as adverbs can also have the function of linking ideas together. Hirsh (2004, p. 59) gives an example with the word *nonetheless* which is in fact classified as an adverb and not a conjunction in different dictionaries such as the Oxford Advanced Learner's Dictionary (Hornby, 2010) and the Gyldendal's English-Danish Dictionary (Kjærulff Nielsen, 1988). Therefore, the sub-category is here renamed as **linking words** to be able to include both conjunctions and adverbs.

The last category in Hirsh's framework, **authoritative** differs from the others in that the words of this category are related more to the stylistics of academic prose than to describing the subject matter of the research being carried out. This category is described as containing academic words such as 'undertook', 'assistance', and 'predominantly' that "could arguably be replaced with more general

language words such as ‘did’, ‘help’, and ‘mainly’ without significantly changing the meaning of the text.” (Hirsh, 2004, p. 65). The use of the more formal words given here suggests that the writer attempts to adhere to formality expectations expressed sometimes implicitly by the academic community. Certainly, academic writers in Danish use words that fulfil the authoritative, stylistic functions described here, but categorising DAWL words into this category demands a particular attention to the context in which the words occur, and thus cannot be done without extensive analysis of the texts. In addition, DAWL words that would fall into this category, e.g. the use of *introduktion* (introduction) instead of *indledning*, have other academic functions as will be described below. Hence, the category of **authoritative** belonging to the interpersonal metafunction was omitted. Instead, an extra category of **stance** was added to the framework within the **interpersonal** metafunction. Stance is an essential feature of academic writing which, however, does not seem to be fully included in Hirsh’s framework. Adding a category of stance to the **interpersonal** metafunction in the analytical framework is motivated by the occurrence of a large number of what I termed **evaluation words** in the category of **descriptive words** in the macro-analysis carried out initially. The concept of stance covers two opposite ways that academic writers can either “stamp their personal authority on to their arguments or step back and disguise their involvement.” (Hyland, 2005, p. 176). In other words, stance concerns the degree of confidence the writer has in the arguments and claims of the text. It is here we find the use of hedges and emphatics. Hedges are used for showing that the writer does not fully commit to the propositions of the text. Emphatics, conversely, are markers of certainty and are used to highlight the importance and strength of e.g. a result or an argument. Besides this component of evidentiality (Hyland, 1999, p. 3), stance also involves an affective component referring to the writer’s more varied attitudes towards arguments, persons, findings, etc. expressed through attitude markers which can be used for conveying e.g. surprise, agreement, or frustration. In addition, stance includes a relation dimension that concerns how academic writers refer to themselves and the audience in order to build a form of relationship with the readers. In doing so, writers can use explicit markers to acknowledge the presence of the reader and to encourage them to be active participants in the text (Hyland, 1999, pp. 3–5). Adding a category of stance to the analytical framework is a first step in showing how stance is expressed in Danish academic writing.

This second analysis was checked by a second researcher, and Table 7.2 provides the number of DAWL words in each of the categories and sub-categories described above including the two categories of **technical** and **phrasal elements**. As can be seen, the number of words in total is larger than the number of items in the DAWL (770) due to the multifunctionality of the words as mentioned

above. This is also the reason why no numbers of words are given for the sub-categories as they tend to overlap (cf. Hirsh, 2004, p. 95)

*Table 7.2. The functional framework for the final categorisation of the DAWL words*

Metafunction	Category	Sub-categories
Textual (403)	metatextual (79)	headings
		in-text cues
		metacommunication
	extratextual (124)	previous and future research
		institutional
		borrowed methods/materials
		ethical consent
	intratextual (200)	linking words
carrier words		
Ideational (632)	scholarly processes (309)	
	state of affairs (211)	context
		setting (location/period)
		participants
		characteristics
relations between entities (102)		
Interpersonal (196)	Stance (196)	
Technical (9)		
Phrasal element (19)		

As we can see, almost half of the functions fulfilled by the DAWL words are ideational, a third of the functions are textual, and only roughly every sixth function is interpersonal. This distributional pattern may not be surprising when we look more closely at the functions the various categories and sub-categories describe in an academic text. In the following sections, examples of the DAWL words together with concordance data from the AcaDan corpus are given in relation to the categories and sub-categories outlined in Table 7.2. Each category and its sub-categories are introduced with a description of the functions the words that fall into these categories perform. Since the analytical framework applied closely resembles that of Hirsh (2004, 2010), the descriptions of those categories and sub-categories corresponding to Hirsh's framework are based on his description of them.

#### *7.2.2.1. Metatextual*

Words in the metatextual category help readers to orient themselves in the text as a whole and make connections between different parts of the texts. The first sub-category, **headings**, signals to the reader what the section in question contains. Hirsh (2004, p. 54) points out that this sub-category with its predictive cues in form of headings is vital for both the bottom-up processing of the text as well

as the top-down processing as it affords the reader with cues for the processing. The DAWL lemma *afrunding* (rounding off, final comments) can be used as a heading in the end of an article or a thesis as a final comment much in line with another DAWL lemma *konklusion*, as the example from the AcaDan corpus shows:

***Afrunding*** Sammenligningen mellem de forskellige kategoriseringer af patienter i henholdsvis en amerikansk og en dansk kontekst i denne artikel er naturligvis ganske kort og tentativ. (Humanities)<sup>29</sup>

English translation: **Final comments** The comparison between the different categorisations of patients in an American and in a Danish context, respectively, in this article is naturally rather short and tentative.

In fact, in eight out of 13 occurrences, this is how the word is used in the AcaDan Corpus, which underlines how *afrunding* serves an academic function. Other DAWL words that can be used as headings and thus can fall into this sub-category are *afsnit* (section), *analyse* (analysis), *baggrund* (background), *diskussion* (discussion), *disposition* (outline), *introduktion* (introduction), and *metode* (method). All the DAWL words in this sub-category are nouns. Some of them also have other functions outside the textual metafunction such as *analyse* and *diskussion* which will be shown below. The words *afsnit* and *baggrund* occur in headings with specifiers in the AcaDan corpus in order to specify what the section in question is about in particular. Below is an example of how *baggrund* is used as a part of a multiple word heading marked in bold:

**1 indledning og baggrund Rygsmerter og ryg sygdomme** En samfundsøkonomisk analyse fra 2011 udarbejdet af Statens institut for Folkesundhed (3) beregnede omkostningerne forbundet med ryg sygdomme og rygsmerter i Danmark. (Health Science)<sup>30</sup>

English translation: **1 introduction and background Back pain and back disorders** A socio-economic analysis from 2011 from the National Institute of Public Health (3) calculated the costs related to back disorders and back pain in Denmark.

The second sub-category, **in-text cues**, comprises the DAWL words mentioned above as these can also be used for referring back or forth in the text. Other DAWL words in this sub-category are *andetsteds* (elsewhere), *begyndelse* (beginning), *forudgående* (preceding), *føromtalt* (before-mentioned), *indledningsvis* (by way of introduction), *nedenfor* (below), *ovenfor* (above), *ovenstående*

---

<sup>29</sup> Johansen, B. S., & Johansen, K. S. (2008). Islam i dansk psykiatri. Tidsskrift for Islamforskning, 3(1), 30–43.

<sup>30</sup> Roos, E., Bliddal, H., Christensen, R., Hartvigsen, J., Mølgaard, C., Søgaard, K., & Zebis, M. K. (2013). Forebyggelse af Skader og Sygdomme i Muskler og Led. Vidensråd for Forebyggelse.

(the above), and *sidstnævnte* (the latter). These are words that link the different parts of a text together, and enable the reader to connect these parts as shown in the example from the AcaDan corpus below. As such, there is an overlap of the DAWL words in the heading sub-category and in this sub-category. Below, context is provided for the DAWL word *indledningsvis*, and it shows how the author refers back in the text (also the phrase *som nævnt* could fall into this category if the focus was on academic phrases).

Som nævnt **indledningsvis** er en af de fundamentale vanskeligheder ved kausal inferens, at vi ikke kan observere, hvad der ville have været virkningen, hvis årsagen (kontrafaktisk) var udeblevet. (Social Science)<sup>31</sup>

English translation: As mentioned, **by way of introduction**, one of the fundamental difficulties in relation to causal inference is that we cannot observe what the effect would have been if the cause (counterfactual) had not occurred.

As mentioned above, I expanded the metatextual category with the subcategory **metacommunication** to include academic words used for this purpose, that is, for telling the reader what is going to happen in the form of academic functions in the following paragraph, section, or chapter, or more generally, what the text is about. Granted, both headings and in-text cues are also used for metacommunicative purposes, but introducing a third sub-category in the metatextual category allows for a more fine-grained analysis. Examples of DAWL words in this metacommunication sub-category are primarily verbs such as *beskæftige* (as in *beskæftige sig med*: deal with), *fremstille* (describe), *introducere* (introduce), *konkludere* (conclude), *nævne* (mention), *omhandle* (concern) *omtale* (comment), *opsummere* (summarise), *redegøre* (give an account of), and *svare* (answer). Below is an example from the AcaDan Corpus in which *beskæftige* is used for metacommunicating the content of the text itself.

Denne artikel **beskæftiger** sig med den miljømæssige faktor, som hedder: Tidligt tab af primære molarer som følge af caries. (Health Science)<sup>32</sup>

English translation: This article **deals** with the environmental factor called: Early loss of primary molar tooth caused by caries.

---

<sup>31</sup> Hariri, J. G. (2012). Kausal inferens i statskundskaben. *Politica*, 44(2), 184–201.

<sup>32</sup> Hess, P., & Kreiborg, S. (2014). Forebyggelse af uønskede tandvandring efter tidligt mælketandstab. *Tandlægebladet*, 118(11), 902–9.



### 7.2.2.2. Extratextual

As discussed in Chapter 4, academic writing is to a very high degree about showing your membership of an academic community, i.e. a shared “community of practice” (Wenger, 1998; Hyland, 2004). The words that fall into the **extratextual** category are used for referring to other academics’ research in different ways, both to identify gaps and to support the researcher’s own claims, arguments and findings. Examples of DAWL words that fall into the **extratextual** category are categorised according to the four sub-categories outlined in Table 7.2.

The first sub-category, **previous and future research**, is used for: 1) referring to previous research in order to either point out weaknesses and gaps in the previous research literature to justify the research being carried out, and 2) pointing out that the findings of the research in question need to be further explored and substantiated. The DAWL words belonging to this sub-category are adjectives such as *tidlig* (previous), often in its comparative form, *tidligere*, specifying a certain piece of research, and *fremtidig* (future), and nouns like *litteratur* (literature), *forskningsprojekt* (research project), and *udforskning* (exploration). In the concordance data below, we see how the authors of a Natural Science article point out a gap in previous research:

Da der i *tidligere* undersøgelser er set indikationer på aerob nedbrydning af 68 diestrene, men at dette ikke er tilbunds gående belyst, kan der udføres aerobe nedbrydningsforsøg af diestrene i laboratoriet repræsenterende et let påvirket sediment. (Natural Science)<sup>33</sup>

English translation: Since **previous** research have seen indications of aerobic disintegration of diesters 68, and this has not been thoroughly explored, aerobic disintegration experiments of diesters in the laboratory can be conducted as a representation of a lightly influenced sediment.

The words that fall into the second sub-category of **institutional** are especially proper nouns and titles of institutions in line with the examples given in Hirsh (2004, p. 57). Institutional DAWL words are for example:

- *forelæsning* (lecture)
- *forskningsprojekt* (research project)
- *ministieriel* (ministerial)
- *obligatorisk* (mandatory)

---

<sup>33</sup> Fjordbøge, A. S., Kjeldsen, P., Petersen, P. A., & Durant, N. D. (2007). Oprensning af forureningen på depotet ved Høfde 42 ved hjælp af nul-valent jern. Miljøministeriet.

- *official* (official)
- *offentlig* (public)
- *reference* (reference)
- *strategi* (strategy)
- *studere* (study, verb)
- *system* (system)
- *universitet* (university)
- *videnskabelig* (scientific)
- *værdigrundlag* (value statement)

The purpose of this sub-category is to be able to refer to the specific context in which the investigation in question is carried out as can be seen in this concordance example from a Humanities research article:

Men alligevel, når det drejer sig om evidensbegrebets betydning, kommer vi nok tættere på ved at referere til de forskningsmiljøer, politiske og *ministerielle* sammenhænge, der tydeliggør interessen for evidensbaseret. (Humanities)<sup>34</sup>

English translation: However, when looking at the meaning of the concept of evidence, it helps to refer to the research environments and the political and **ministerial** contexts that make the interest for evidence-based research explicit.

As with many of the other DAWL words, also these words are multifunctional as can be seen with *forskningsprojekt*, which occurs in two sub-categories within the extratextual category. As mentioned above, this sub-category does not only pertain to research but also to education and education administration. For example, international staff members need to be able to navigate the vocabulary of university administration when taking on teaching responsibilities and administrative duties. Similarly, students in higher education are expected to be able to understand what is meant by *færdigheder* (skills) in contrast to *kompetencer* (competencies) in the curriculum. Knowledge of such terms is also highly relevant for L2 Danish users in higher education.

The third sub-category, **borrowed methods and materials**, contains words that are used for referring to external entities such as programmes, frameworks, and systems that the researchers have used in order to carry out their investigation. Examples of DAWL words in this sub-category are *kategori* (category), *ressource/resurse* (resource), *spekter/spektrum* (spectrum), and *model* (model). This is

---

<sup>34</sup> Borgnakke, K. (2015). Evidensbevægelsen i spændingsfeltet mellem sikker viden og ikke-viden. *Ceptra-Striben*, (17), 22–31.

illustrated in the concordance example below in which the authors of a Natural Science research article refer to the Raman spectroscopy, a technique used in chemistry:

De ”bølger”, der ses i *spektret*, skyldes, at selve Ramanspektret er af dårlig kvalitet og svarer til støj i *spektret*. (Natural Science)<sup>35</sup>

English translation: Those “waves” that are seen in the **spectrum** are there because the Raman **spectrum** in itself is of poor quality, and it corresponds to noise in the **spectrum**.

DAWL words in the fourth sub-category, **ethical consent**, are e.g. *accept* (accept, noun), *acceptere* (accept, verb), and *tilkendegivelse* (declaration). In relation to this sub-category, Hirsh (2004, p. 58) points out how especially the Health Science discipline makes use of words that illustrate that the researchers have gained consent from participants and research committees in order to carry out their research. The example from the AcaDan corpus below exemplifies this. However, the issue of ethical consents is also relevant in research in other academic disciplines.

(...) i enkelte tilfælde kontaktede jeg, med informantens **accept**, informantens behandler og delte mine observationer for at sikre en optimal lydhørhed hos behandlere. (Health Science)<sup>36</sup>

English translation: (...) in a few cases, I contacted, after **accept** from the informant, the informant’s therapist, and shared my observations to ensure optimal responsiveness from the therapists.

#### 7.2.2.3. Intratextual

The category of intratextual is used for linking ideas together within a text. Hirsh defines two sub-categories: **conjunctions** and **carrier words**. As mentioned above, the conjunction sub-category was renamed **linking words** since many words classified as adverbs can also be said to carry the function of linking ideas together. This sub-category is used for establishing different kinds of relationship between ideas: *causative*, *contrastive*, *temporal*, and *summative*. The latter category, **carrier words**, comprises words used anaphorically or cataphorically to refer to and classify ideas outlined in other parts of the text. As Hirsh (2004, p. 60) points out, carrier words are often non-specific nouns made specific only when linked to ideas referred to other places in the text.

---

<sup>35</sup> Reeler, N. E. A., Nielsen, O. F., Sauer, S. P. A., Kjærgaard, H. G., Borring, N., Filtenborg, T., ... Wadum, J. (2013). Raman af hvide pigmenter. *Dansk Kemi*, 94(12), 30–34.

<sup>36</sup> Buus, N. (2011). Forbedring af adherence til medicinsk depressionsbehandling med udgangspunkt i brugerperspektivet. Syddansk Universitet. IST - Institut for Sundhedstjenesteforskning. Helbred, Menneske og Samfund.

In Table 7.3, examples of **linking words** in the DAWL are given divided into the different uses of these words. Note that many of these words have multiple functions, and therefore occur in more than one type of intratextual linking relation. In the following, I will give examples taken from the AcaDan Corpus on how these linking words perform the function of linking ideas together in different ways.

Table 7.3. Types of linking words in the DAWL

DAWL linking words			
<i>Causative</i>	<i>Contrastive</i>	<i>Temporal</i>	<i>Summative</i>
<i>derfor</i> (therefore)	<i>derimod</i> (on the other	<i>dernæst</i>	<i>altså</i> (consequently)
<i>dermed</i> (thus)	hand	(subsequently)	<i>dermed</i> (thus)
<i>derved</i> (thereby)	<i>dog</i> (however)	<i>endvidere</i>	<i>hermed</i> (by this)
<i>hermed</i> (by this)	<i>hvorimod</i> (whereas)	(moreover)	<i>således</i> (thus)
<i>hvormed</i> (by which)	<i>imidlertid</i>	<i>først</i> (first)	
<i>hvorved</i> (by means of	(nevertheless)	<i>idet</i> (as)	
which)	<i>men</i> (but)	<i>mens</i> (while)	
<i>idet</i> (because, as)	<i>mens</i> (whereas)	<i>samtidig</i>	
<i>nødvendigvis</i>		(simultaneously)	
(necessarily)		<i>ydermere</i>	
<i>således</i> (thus)		(furthermore)	

The following fragment from a Social Science article shows how the DAWL word *idet* shows a **causative** relationship between the proposition set forth in the first part of the sentence, and the proposition in the last part of the sentence:

Akademiske tekster arbejder meget åbenlyst ud fra en kumulativ vidensopbygning, **idet** forfatteren eksplicit søger at redegøre for tekstens intertekstualitet (...). (Social Science)<sup>37</sup>

English translation: Academic texts, clearly, are constructed based on an accumulative knowledge **as** the writer explicitly seeks to account for the intertextuality of the text (...).

The **contrastive** link between different assertions in a sentence is evident in this example from an article from humanities with the DAWL word *mens*:

<sup>37</sup> Nicolaisen, M. S. (2014). Analog tekst i digital kontekst. Nordisk Tidsskrift for Informationsvidenskab Og Kulturformidling, 2(3), 41–53.

Ifølge princippet epoché er det første udsagn en ren beskrivelse af det oplevede fænomen, **mens** det andet er udtryk for vurdering der siger mere om betragteren end om billedets fænomenologiske fremtrædelsesform. (Humanities)<sup>38</sup>

English translation: According to the principle epoché, the first statement is a simple description of the experienced phenomenon **whereas** the second statement is a manifestation of assessment that says more about the viewer than about the picture's phenomenological form of appearance.

According to Hirsh (2004, p. 59), **temporal** linking words indicate a time sequence of the propositions they join together. This can be seen twice in the following example from Natural Science:

Jeg vil illustrere dette **først** med et lille eksempel og **der næst** betragte paralleller mellem Euklids geometri og Kandinskys teori (...). (Natural Science)<sup>39</sup>

English translation: I will illustrate this **first** with a small example, and **subsequently** consider the parallels between the geometry of Euclid and Kandinsky's theory (...).

The last kind of linking relationship, the **summative**, functions, as the title indicates, as a way of summarising the content of the preceding proposition. The DAWL words *således* and *dermed* can be used in this function:

I en besvarelse er det angivet at klubben/banen har brugt DGU's konsulent, og to har angivet, at de bruger en ekstern konsulent. Kun få respondenter har **således** angivet, at de har fået professionel hjælp til at identificere skadevolderne. (Natural Science)<sup>40</sup>

English translation: One survey reply states that the club/field has used the DGU consultant, and two state that an external consultant has been used. Only few respondents have **thus** stated that they have received professional help to identify what has caused the damages.

The second sub-category of the intratextual category, **carrier words**, shares with the linking words the function of linking propositions together. Given that carrier words such as the English words "issue" are lexical words used to refer and classify ideas and concepts set forth in other parts of the text (Hirsh, 2004, p. 60), they are as they occur in the academic texts almost without meaning if the reader cannot recognise the ideas or concept to which they refer. In that way, they have undergone a process of grammaticalisation as suggested by Meyer (1990 in Nation, 2016, p. 300). The example below from the AcaDan corpus with DAWL words implies that the same process of delexicalisation

---

<sup>38</sup> Funch, B. S. (2003). Den fænomenologiske metode i museologisk forskning. *Nordisk Museologi*, (1), 17.

<sup>39</sup> Jensen, A. H. J. (2009). Matematik er kunst uden pensel, kunst er matematik uden kridt. *Kvant. Tidsskrift for Fysik og Astronomi* (4).

<sup>40</sup> Jensen, A. M. D., Mortensen, B., & Paaske, K. (2012). Pesticidforbrug og pesticidbelastning på golfbaner. Miljøstyrelsen.

may be happening to a set of lexical words in Danish. In the example below, the DAWL word *proces* is used anaphorically to refer to the concepts of “fordybelse og tolkning” in the first sentence of the example (see translation below).

De udstillede installationer og værker er alle konceptuelle og kræver fordybelse og tolkning. Disse **processer** søges afhjulpnet gennem muligheden for at vælge guidede tours med forskellige indgangsvinkler (...). (Humanities)<sup>41</sup>

English translation: The exhibited installations and works are all conceptual and demand concentration and interpretation. These **processes** can be remedied by making it possible to choose guided tours of different perspectives (...).

Another example of a DAWL words used as a carrier word is given below. Here the carrier word is used cataphorically, or what Hirsh terms “as an advanced label” for the proposition that follows:

**Baggrunden** for den unge Casper Roses udtalelse er, at Liberal Alliance i 2010 præsenterede ni spidskandidater til næste folketingsvalg. (Social Science)<sup>42</sup>

English translation: The **background** for young Casper Rose’s statement is that Liberal Alliance in 2010 presented nine top candidates for the next general election.

#### 7.2.2.4. Scholarly processes

The category ‘scholarly processes’ belongs to the ideational metafunction as described in Table 7.1 and Table 7.2. As can be seen from Table 7.4, the DAWL words in the scholarly processes category are used for describing some central academic language functions (Bailey et al., 2007) of the research process. Their function in academic writing is to enable the writer to describe the activities, tools, and mental processes related to the research carried out (Hirsh, 2004, p. 61). This category corresponds largely to Martin’s (1976) category of vocabulary related to the research process in hard science disciplines, but as Hirsh (2004, p. 61) argues on the basis of his study, words belonging to this category are also used in the soft sciences. In particular, the verbs *analysere* (analyse), *kategorisere* (categorise), and *undersøge* (investigate) are frequent academic words that are used across disciplines as the examples in Table 7.4 show. Given the nature of this category, it is not surprising that this category contains the largest number of DAWL words, 309 to be precise.

Table 7.4 lists verb examples of these paired with concordance data from the AcaDan Corpus. The DAWL nouns in this category include *analyse* (analysis), *granskning* (study), *kategori* (category),

---

<sup>41</sup> Skot-Hansen, D. (2013). Fransk kultur i undtagelsestilstand - franske kunst- og kulturarenaer i en digital æra. Nordisk Kulturpolitisk Tidskrift, 16(02), 201–216.

<sup>42</sup> Knudsen, T. (2011). Den politiske djøficering. Samfundsoekonomen, (3), 37–41.

*konklusion* (conclusion), *metode* (method), *model* (model), *målestok* (measure), *overvejelse* (consideration), and *udredning* (explanation) are included in this category. Some of these occur both as nouns and verbs, i.e. both the nominal and verbal forms are in the DAWL, and in total there are 39 of these noun-verb pairs in the **ideational** sub-category of **scholarly processes**. Examples of these are:

- *afdekke-afdækning* (uncover-uncovering)
- *beskrive-beskrivelse* (describe-description)
- *definere-definition* (define-definition)
- *konkludere-konklusion* (conclude-conclusion)
- *modificere-modification* (modify-modification)
- *placere-placering* (place-placement)
- *tolke-tolkning* (interpret-interpretation)
- *udforske-udforskning* (explore-exploration)
- *udvikle-udvikling* (develop-development).

It is not surprising that the DAWL contains so many nominalisations considering that this is a defining trait of academic language as described in Chapter 2.

Table 7.4. Examples of DAWL words in the scholarly processes sub-category

DAWL word	AcaDan example	English translation
<i>analysere</i> (analyse)	Forskel i antal af hudproblemer og graden af hudproblemer relateret til GA og fødselsvægt blev <b>analyseret</b> med Chi2 test og med odds ratio (OR) og tilhørende 95% CI. (Health Science) <sup>43</sup>	Difference in number of skin conditions and the degree of skin conditions related to GA and birth weight was <b>analysed</b> using Chi2 test and odds ratio (OR) and matching 95% CI.
	Udgangssedimentet blev <b>analyseret</b> i triplikat for 17 stoffer (se tabel 3.4.1). (Natural Science) <sup>44</sup>	The output sediment was <b>analysed</b> in triplicate for 17 substances (see table 3.4.1).
	I dette afsnit <b>analyseres</b> først den danske og siden den grønlandske ende af den relation, som udgøres af Rigsfællesskabet. (Social Science) <sup>45</sup>	In this section, first the Danish and then the Greenlandic side of the relation constituted by the Danish National Community are <b>analysed</b>
<i>kategorisere</i> (categorise)	Undersøgelsens population er <b>kategoriseret</b> i fem etniske grupper alene ud fra data om eget og forældres fødeland. (Health Science) <sup>46</sup>	The population of the study is <b>categorised</b> into five ethnic groups solely based on data on their own and their parents' birth country.
	De refleksivt anvendte pronominer der ikke korefererer med det grammatiske subjekt, blev <b>kategoriseret</b> i de nedenstående seks underkategorier. (Humanities) <sup>47</sup>	The pronouns used reflexively that do not co-refer with the grammatical subject were <b>categorised</b> into the six sub-categories below.
<i>Systematisere</i> (systematise)	(...) for så vidt som det foranlediger mig til mere udførligt at forklare og <b>systematisere</b> resultaterne eller indsigterne fra mine tidligere forskningsarbejder (...) (Social Science) <sup>48</sup>	(...) in so far as it caused me, in detail, to explain and <b>systematise</b> the results and knowledge from my previous research (...)

<sup>43</sup> Ramsgaard-Jensen, T. F., Horskjær, M., Jensen, L. B., Due, K. M., & Grønkjær, M. (2014). n-CPAP behandling forårsager alvorlige nasale hudproblemer hos de mindste præmature børn. *Klinisk Sygepleje*, 28(1), 35–45.

<sup>44</sup> Fjordbøge, A. S., Kjeldsen, P., Petersen, P. A., & Durant, N. D. (2007). Oprensning af forureningen på depotet ved Høfde 42 ved hjælp af nul-valent jern. *Miljøministeriet*.

<sup>45</sup> Gad, U. P. (2008). Når mor/barn-relationen bliver teenager. *Politica - Tidsskrift for Politisk Videnskab*, 40(2), 111–133.

<sup>46</sup> Helweg-Larsen, K., Kastrup, M., Baez, A., & Flachs, E. M. (2007). Etniske forskelle i kontaktmønsteret til psykiatrisk behandling. Syddansk Universitet. Statens Institut for Folkesundhed.

<sup>47</sup> Jensen, T. J. (2009). Refleksivt anvendte pronominer i moderne dansk. *Ny Forskning I Grammatik*, 16, 131–151.

<sup>48</sup> Hansen, J. A., & Hammerslev, O. (2010). Bourdieu og staten. *Praktiske Grunde. Tidsskrift for Kultur- Og Samfundsvidenskab*, (1-2), 11–33.



---

<i>undersøge</i> (investigate)	Vi har <b>undersøgt</b> eksponeringsvariablens betydning for udfaldet (storforbrug) i det samme år, hvilket vil sige, at vi gennemfører en tværsnitsundersøgelse, idet der ikke indgår et tidselement fra eksponering til udfald. (Health Science) <sup>49</sup>	We have <b>investigated</b> the influence of the exposure variable on the outcome (massive expenditure) in the same year, which means that we conduct a cross section study, as there is no time element from exposure to outcome.
	Da et af formålene med denne artikel er at <b>undersøge</b> spændvidden i legitimeringen af brugerinddragelsen(...) (Social Science) <sup>50</sup>	Since a purpose of this article is to <b>investigate</b> the span in the legitimatising of user involvement (...)

---

#### 7.2.2.5. State of affairs

Words in the ideational category of state of affairs are used for describing and especially for specifying the subject matter of the text, i.e. the **context, setting, participants, and characteristics** of the research reported on. These sub-categories tend to overlap (Hirsh, 2004, p. 53), but an attempt to exemplify each of these has been made below. These unclear distinctions of these sub-categories emphasise the multifunctionality of academic words mentioned in the introduction to this chapter.

The **context** sub-category of this category comprises words that describe external factors and variables influencing the research, e.g. societal matters that in some way influence the investigation carried out. Put differently, by Hirsh (2004, p. 62), words in this sub-category are used for explaining the true nature of what is observed in the investigation. The examples below show how DAWL words (*fastlå, underliggende*) can be used for describing contextual factors influencing the research.

Desuden **fastslår** loven, at man i statsskovene skal fremme udviklings- og forsøgsvirksomhed. (Natural Science)<sup>51</sup>

English translation: Moreover, the law **requires** that developmental and experimental activities are prioritised by the state forest.

---

<sup>49</sup> Hvidtfeldt, U. A., Vinther-Larsen, M., Petersen, C. B., Thygesen, L. C., & Grønbæk, M. (2006). Ældre og alkohol. Statens Institut for Folkesundhed, Socialministeriet - Styrelsen for Specialrådgivning og Social Service.

<sup>50</sup> Rasmussen, C. H. (2015). Brugerinddragelse og kulturpolitisk kvalitet. Nordisk Kulturpolitisk Tidsskrift, 18(1), 76–95.

<sup>51</sup> Graudal, L., Nielsen, U. B., Schou, E., Thorsen, B. J., Hansen, J. K., Bentsen, N. S., & Johannsen, V. K. (2013). Muligheder for bæredygtig udvidelse af dansk produceret vedmasse 2010-2100. Institut for Geovidenskab og Naturforvaltning, Københavns Universitet.

Det må dog erkendes, at vi savner en dybere forståelse af faktorerne bag den stærkt bekymrende langtidstendens til fald i den **underliggende** danske produktivitetsvækst. (Social Science)<sup>52</sup>

English translation: It needs to be acknowledged, however, that we lack a deeper understanding of the factors behind the seriously worrying long-time tendency for decrease in the **underlying** Danish productivity growth.

The **setting** sub-category tells the reader about the location and time period of the activities and processes of which the research is focused on. As Hirsh argues (2004, p. 62), **setting** words are describing general concepts, but are nonetheless academic words as they, like other **state of affairs** words, specify the subject matter of the academic text. The DAWL word *årti* (decade) is in the AcaDan Corpus example below used to describe a development that is important for the research reported on in a Social Science text.

Der er sket en stor udvikling i brugerundersøgelser de seneste **årtier**. (Social Science)<sup>53</sup>

English translation: Major developments in user inquiries have taken place in recent **decades**.

As the **state of affairs** category overall is preoccupied with specifying the subject matter of the research being written about, the **participant** sub-category represents “the core subject matter of the text” (2004, p. 63) and as such, it does not only specify animate entities but also abstract ideas and concepts participating in the research process. Similarly, the words that fall into the **participant** sub-category are highly topic-specific while still academic as they are used across academic disciplines (Hirsh, 2004, p. 63). The DAWL contains some words that could be used to specify and describe the participants of the research carried out. Below some examples of these words in context are given.

Dette har vi tilføjet efterfølgende på baggrund af data fra Tilbudsportalen, beriget med svar fra spørgeskemaet, hvor **lederne** selv har svaret på, hvilken form der kendetegner deres plejecenter. (Health Science)<sup>54</sup>

English translation: We have added this afterwards using data from the Tilbudsportalen including responses from the survey in which the **managers** have replied which type describes their care facility.

---

<sup>52</sup> Sørensen, P. B. (2010). Nedturen i dansk økonomi. Samfundsoekonomen, 2010(1), 29–34.

<sup>53</sup> Pors, N. O. (2011). Evidens om bibliotekernes brugere. Dansk Biblioteksforskning, 6(2/3), 65–81.

<sup>54</sup> Hjelmar, U., Bhatti, Y., Rostgaard, T., Petersen, O. H., Vrangbæk, K., Larsen, P. T., & Jacobsen, L. (2016). Kvalitet på offentlige og private plejecentre i Danmark.

En meget forenklet tegning, figuren nederst til højre, viser princippet for et instrument, der bygger på anvendelsen af ovenstående *formel*. (Natural Science)<sup>55</sup>

English translation: A very simplified drawing, the figure bottom right, shows the principle for an instrument that builds on the use of the **formula** stated above.

In contrast to the first sub-category, **context**, which specifically aims at describing contextual factors impacting the research, **characteristics**, the last sub-category of **state of affairs**, is concerned with characterising the participants of the research. As such, many of the DAWL words categorised into the second macro category of **descriptive words** in the first functional analysis of the DAWL words fall into this sub-category. Below are given some AcaDan Corpus examples of DAWL words used for characterising elements of the research reported on.

(...) ud fra et kønsperspektiv kan det eksempelvis antages, at køn betyder noget for, hvordan ‘other emotion management’ aktiviteter fordeler sig blandt medarbejdere. Køns- og emotionssociolog Shields anfører eksempelvis, at kvinder i en *nordamerikansk* sammenhæng generelt mener, at de forventes at udtrykke positive følelser over for andre (...). (Social Science)<sup>56</sup>

English translation: (...) from a gender perspective, it can be assumed that gender means something for how ‘other emotion management’ activities are distributed among the employees. For example, gender- and emotion sociologist Shields argues that women in a **North-American** context generally think that they are expected to express positive feeling towards other people (...).

En sådan prioritering bør give mere plads til mere simple projekter vedrørende modernisering af *traditionelle* dambrug. (Natural Science)<sup>57</sup>

English translation: Such a prioritisation should leave more room for more simple projects concerning modernisation of **traditional** fish farming.

#### 7.2.2.6. Relations between entities

Hirsh argues that words in this category carry out an essential function in academic discourse by showing how ideas, concepts, and entities are related and affect each other (Hirsh, 2004, p. 64). By describing often causative relations between entities, the words of this category contribute to the particular research’s creation of new knowledge. In the example below from the AcaDan corpus, two

---

<sup>55</sup> Rasmussen, S. E. (2009). Analyse ved røntgendiffraktion-Rietveld-metoden. *GeologiskNyt*, 19(6).

<sup>56</sup> Poder, P. (2010). Når medarbejdere håndterer hinandens følelser. *Tidsskrift for Arbejdsliv*, 12(3), 72–86.

<sup>57</sup> Nielsen, R., Thøgersen, T. T., Andersen, J. L., Dalskov, J., & Kusier, R. (2015). Situationsbeskrivelse af den danske fiskeri-, akvakultur- og fiskeindustri sektor. Institut for Fødevarer- og Ressourceøkonomi, Københavns Universitet.

DAWL words occur which both show how including one element of the research process (*betydningen af sociale faktorer*) contributes to an important finding signalled by the word *værdifuldt*.

Det er derfor håbet, at yderligere analyser kan gennemføres, som kan belyse de hypoteser, som vi indledningsvist har beskrevet. Specielt vil det være *værdifuldt* at medinddrage *betydningen* af sociale faktorer for de påviste etniske forskelle. (Health Science)<sup>58</sup>

English translation: Thus, it is the hope that further analyses can be conducted that can shed light on the hypotheses described in the introduction. In particular, it would be **of value** to include the **impact** of social factors on the shown ethnical differences.

Also, DAWL words like *rolle*, *påvirke* and *bevirke* have the function of specifying relations between entities as shown below:

Tabellen viser klart at tilhørsforholdet er størst – og stigende – i forhold til lokalområdet, og at Europa, eller for den sags skyld Verden, ikke spiller nogen *rolle* forstået på den måde at disse svarmuligheder næsten ikke angives af nogen af svarpersonerne. (Social Science)<sup>59</sup>

English translation: The table clearly shows that the membership is biggest – and increasing – in relation to the local area, and that Europe, or the World for that matter, does not play any **role** as indicated by the fact that these replies are not given by any of the respondents.

To forhold *påvirker* direkte og med øjeblikkelig virkning det udbytte man får fra skoven. (Natural Science)<sup>60</sup>

English translation: Two factors **influence** directly and instantaneously the outcome produced by the forest.

#### 7.2.2.7. Stance

The DAWL words that fall into the stance category can further be classified according to the three components or functions of stance as proposed by Hyland (1999): **evidentiality** which covers both hedges and emphatics, **affect** pertaining to the attitudes of the writer, and **relation** used for addressing the reader of the text. Examples of DAWL words used for expressing stance are given in Table 7.5,

---

<sup>58</sup> Helweg-Larsen, K., Kastrup, M., Baez, A., & Flachs, E. M. (2007). Etniske forskelle i kontaktmønsteret til psykiatrisk behandling. Syddansk Universitet. Statens Institut for Folkesundhed.

<sup>59</sup> Gundelach, P. (2001). National identitet i en globaliseringstid. Dansk Sociologi, 12(1), 63–80.

<sup>60</sup> Graudal, L., Nielsen, U. B., Schou, E., Thorsen, B. J., Hansen, J. K., Bentsen, N. S., & Johannsen, V. K. (2013). Muligheder for bæredygtig udvidelse af dansk produceret vedmasse 2010-2100. Institut for Geovidenskab og Naturforvaltning, Københavns Universitet.

which is divided into the three functions outlined in Section 7.2.2. Concordance examples from the AcaDan Corpus are given for one word in each function.

As can be seen from Table 7.5, the example given in the **evidentiality** sub-function of **hedges** shows how the authors of a Social Science article make a small reservation when explaining a result using the DAWL word *sandsynligvis*. Another evidentiality marker is the use of the **emphatic** *entydigt* (unambiguous) in the extract from a Natural Science article, which signals how certain the authors are in the truth of their results. In the example of the **affect** function in Table 7.5, the authors show their surprise in the results by using *bemærkelsesværdig* (remarkable). The example given for the function of **relation** shows how the author of a Social Science research article engages in relating to the reader and the surrounding academic community not only by the use of the pronoun *os* (us), which granted is not in the DAWL, but also by using the DAWL word *vise* to direct the reader to the context of his proposition.

Table 7.5. DAWL words with stance functions

Function	Sub-function	DAWL words	AcaDan examples
Evidentiality	hedges	<i>antage</i> (assume) <i>antydde</i> (suggest) <i>ofte</i> (often) <i>delvis</i> (partly) <i>imidlertid</i> (however) <i>mulig</i> (possible) <i>potentiel</i> (potential) <i>relativ</i> (relative) <i>sandsynligvis</i> (probably) <i>såkaldt</i> (so-called) <i>ubetydelig</i> (insignificant)	Den primære forklaring på dette er <b>sandsynligvis</b> , at det i de fleste tilfælde er nærmest umuligt at dokumentere en effekt. (Social Science) <sup>61</sup>  English translation: The primary explanation for this is <b>probably</b> that in most cases it is almost impossible to document an effect.
	emphatics	<i>afgørende</i> (crucial, crucially) <i>bestemt</i> (certain, certainly) <i>central</i> (central) <i>entydig</i> (unambiguous) <i>fundamental</i> (fundamental) <i>gyldighed</i> (validity) <i>indgående</i> (exhaustively) <i>indlysende</i> (obvious) <i>klarhed</i> (clarity) <i>styrke</i> (strength) <i>veletableret</i> (well-established) <i>vigtig</i> (important) <i>væsentlig</i> (essential, essentially)	Dermed er det moderne biologiske overdrev <b>entydigt</b> et spørgsmål om vegetation og jordbund, ikke historie. (Natural Science) <sup>62</sup>  English translation: Thus, modern, biological commons is <b>unambiguously</b> a question about vegetation and soil, not history.
Affect	Attitude markers	<i>bemærkelsesværdig</i> (remarkable, remarkably) <i>beskeden</i> (small, modest) <i>betragtelig</i> (considerable) <i>betydningsfuld</i> (meaningful) <i>eksakt</i> (exact) <i>forståelig</i> (understandable) <i>indlysende</i> (obvious)	Ikke desto mindre er der tale om <b>bemærkelsesværdige</b> resultater, som kalder på større indsigt i , hvad der foregår, når forsøgspersoner træffer beslutning om forsøgsdeltagelse. (Health Science) <sup>63</sup>  English translation: Nonetheless, these are <b>remarkable</b>

<sup>61</sup> Rasmussen, C. H. (2015). Brugerinddragelse og kulturpolitisk kvalitet. Nordisk Kulturpolitisk Tidsskrift, 18(1), 76–95.

<sup>62</sup> Dam, P. (2013). Overdrev - fra ekstensivt landbrug til intensiv natur?. Landbohistorisk Tidsskrift, 10(2), 9-35.

<sup>63</sup> Wadmann, S. (2013). Informeret samtykke i kliniske forsøg: teknikaliteter, tillid og tætte relationer. Etik I Praksis, 7(2), 31–46.

			results that call for more insight into what happens when test subjects decide on test participation.
Relation	Relation markers	<i>vor</i> (our) <i>vise</i> (show)	<p>Lad os begynde med noget af den ”negative viden” – dét, som samfundsvidenskaberne har <i>vist</i> os, at vi ikke kan. (Social Science)<sup>64</sup></p> <p>English translation: Let us begin with some of the “negative knowledge”, that which the social sciences have <b>shown</b> us that we cannot do.</p>

#### 7.2.2.8. Technical words in the DAWL

As described in Chapter 2, the distinctions between general, academic, and technical vocabulary are not always clear, and both general and academic words can be used in technical senses. In the development of the DAWL (Study 2), attention was directed to ensure that the words identified as academic and selected for the DAWL were as core academic as possible, and the dispersion analysis revealed that there were words identified via the chosen measures and criteria that could also have technical senses. As such even the strictest objective measures cannot prevent academic words taking on technical senses in certain contexts, a phenomenon I termed discipline-dependent polysemy. The nine words listed below are DAWL words that at first sight could be deemed more technical than academic:

- *arv* (inheritance, heritage)
- *cirkulation* (circulation)
- *cirkel* (circle)
- *formel* (formula)
- *færdighed* (skill)
- *heterogenitet* (heterogeneity)
- *indvandring* (immigration)
- *koordinat* (co-ordinate)
- *spekter/spektrum* (specter)

<sup>64</sup> Kurrild-Klitgaard, P. (2012). Frihed mellem fornuft og skepsis. Kritik, (206), 31–39.

Using concordance data from the AcaDan Corpus, I will demonstrate how a DAWL word like *cirkulation* is used in ways that are general to academics, and contrast it with other examples from the same corpus that highlights their discipline-dependent polysemy.

In the Natural Science and Health Science sub-corpora, *cirkulation* is used as a technical word as in the example below. However, in the Humanities and Social Science disciplines the word is used more frequently as well as in a more figurative sense as shown by the example from subject area of literary studies.

Brug af den biogene pulje vil ikke øge den mængde kulstof, som er i 'fri' ***cirkulation*** mellem de ikke fossile kulstofpuljer (...). (Natural Science)<sup>65</sup>

English translation: Use of the biogene pool will not increase the amount of carbon that is in 'free' ***circulation*** between the non-fossil carbon pools (...).

Opgaven var derfor at sprænge kunstinstitutionens produktion og ***cirkulation*** af anæstetiserende billeder gennem en revolutionær æstetisk praksis (ÆPM: 23). (Humanities)<sup>66</sup>

English translation: The task was therefore to disrupt production and ***circulation*** of anaesthetising pictures of the culture institution through a revolutionary aesthetic practice (ÆPM: 23).

The word *cirkel* is not used as a technical word in the AcaDan corpus. Rather, it is used in a more figurative sense to refer to a development that leads to a point similar to the starting point as shown in the example below.

Denne nærmest tautologiske ***cirkel*** af gensidig anerkendelse er med til at skabe stor ulighed i koncentrationen af symbolsk kapital. (Social Science)<sup>67</sup>

English translation: This almost tautological ***circle*** of mutual recognition contributes in part to the creation of great inequality in the concentration of symbolic capital.

On the other hand, as the examples below from the Health Science and Natural Science disciplines illustrate, *cirkel* is also used to refer to concrete figures and specifications:

---

<sup>65</sup> Graudal, L., Nielsen, U. B., Schou, E., Thorsen, B. J., Hansen, J. K., Bentsen, N. S., & Johannsen, V. K. (2013). Muligheder for bæredygtig udvidelse af dansk produceret vedmasse 2010-2100. Institut for Geovidenskab og Naturforvaltning, Københavns Universitet.

<sup>66</sup> Holm, I. W. (2010). Æstetik og politik. K And K: Kultur Og Klasse: Kritik Og Kulturanalyse, (110), 147–155.

<sup>67</sup> Larsen, A. G., & Ellersgaard, C. H. (2012). Status og integration på magtens felt for danske topdirektører. Praktiske Grunde. Tidsskrift for Kultur- Og Samfundsvidenskab, (2-3), 9–30.



Aktuelt er den distale del af den palatinale kanal udrenset (principielt angivet med hvide **cirkler** b-C), men endnu ikke den mesiale del af kanalen (b). (Health Science)<sup>68</sup>

English translation: Currently, the distal part of the palatal canal is cleansed (in principle shown in white **circles** b-C), but not yet the mesial part of the canal.

Området med diatomit er omkranset af prikker. Røde **cirkler** er borer med diatomit. (Natural Science)<sup>69</sup>

English translation: The area with diatomite is surrounded by dots. Red **circles** are drillings with diatomite.

In most of the occurrences, however, it is used for referring to a group of people sharing the same interests or having the same goal:

I de litterære og kunstneriske **cirkler** taler man om postironi, uden at lægge mærke til hvor absurd et sådan udtryk egentlig er. (Social Science)<sup>70</sup>

English translation: In the literary and artistic **circles**, they talk about post irony without noticing how absurd such a term actually is.

The last seemingly technical DAWL word that I will look at here is *heterogenitet* (heterogeneity). This word is used across the academic disciplines in the AcaDan Corpus to denote entities and concepts that are composed of different elements, and is in that way a word that takes its meaning after the context in which it is used. The following example is taken from the Health Science discipline, and is about differences in educational backgrounds among specialists, which is referred to as “educational heterogeneity”.

Den uddannelsesmæssige **heterogenitet** understreges af, at mange specialister – f.eks. læger, sygeplejersker og socialrådgivere (...) gennem årene f.eks. har taget masteruddannelser i forvaltning, offentlig ledelse og public governance (...). (Health Science)<sup>71</sup>

English translation: The educational **heterogeneity** is emphasised by the fact that many specialists – e.g. doctors, nurses, and social workers (...) during the year have

---

<sup>68</sup> Bjørndal, L., Bruun, G., Pedersen, S. D. N., & Langemark, C. (2014). Den biomekaniske udrensning -rationale, effekt og kliniske principper. *Tandlaegebladet*, 118(7), 528–38.

<sup>69</sup> Pedersen, G. K., Pedersen, S. A. S., Bonde, N. C., Heilmann-Clausen, C., Larsen, L. M., Lindow, B. E. K., ... Willumsen, P. S. (2011). Molerområdet geologi – sedimenter, fossiler, askelag og glacialtektonik. *Geologisk Tidsskrift*, 2011(12), 41–135.

<sup>70</sup> Andersen, B. S. (2008). Ironiens kultur og kulturens ironi. *Dansk Sociologi*, 19(1), 9–30.

<sup>71</sup> Beck Jørgensen, T., & Vrangbæk, K. (2013). Den gode forvaltning - på basis af hvilke værdier? *Nordisk Administrativ Tidsskrift*, 90(3), 115–134.

taken further education in administration, public management and public governance (...).

In the example below from the subject area of sociology within the Social Science discipline, the word is used to represent an abstract idea.

Eftersom de samfundsmæssige betingelser gør, at den postmoderne bevidsthed lever af intensiveringen, afspejler selvfortællingerne en sørgende opstemthed – et underskud af krop og et underskud af den vildskab, som ***heterogeniteten*** repræsenterer. (Social Science)<sup>72</sup>

English translation: Since the societal conditions cause the modern consciousness to live off the intensification, the self-narrative reflects a mourning excitement – a deficit of body and a deficit of the wildness that the **heterogeneity** represents.

The examples given here of DAWL words with both academic and technical senses emphasise the point made by Malmström, Pecorari, and Shaw (2018, p. 37) that it is the circumstances that decide whether a word is academic. This is what I termed discipline-dependent polysemy in Chapter 2.

#### 7.2.2.9. Phrasal elements in the DAWL

In total, 19 items in the DAWL can be categorised as phrasal elements in that they most frequently occur in a phrase, and not as an independent word. Consequently, a phrasal element is here defined as a lexical item that most frequently occurs in a phrase, and not as an independent word both in the AcaDan Corpus and in Danish in general. The Word Sketch Tool in Sketch Engine (Kilgarriff et al., 2014, 2004) was used to ascertain the collocational behaviour of these 19 items. This tool shows which words a given word co-occur most frequently with.

Table 7.6 lists the items that only occur in phrases in the AcaDan Corpus. These items are marked as phrasal elements in the DAWL (see Appendix F) as these words should be taught as part of phrases. For example, the word *stede*, which is in fact an old inflected form of *sted* (place), is only used in the phrase *til stede*, and it would not make sense to teach this item as a single word. Likewise, the verb *munde* rarely occurs without the prepositions *ud* and *i* (and also exists as *udmunde i*). According to the Danish Dictionary (Det Danske Sprog- og Litteraturselskab, n.d.-a), it is often used in professional texts in a figurative sense, and in the AcaDan corpus all instances of the word are in the figurative sense:

---

<sup>72</sup> Heinskou, M. B. (2007). Seksualitet mellem risiko og chance. *Dansk Sociologi*, 1(18), 55–75.

Artiklen **mun­der ud i** en skitsering af, hvordan de kan anvendes i en analyse af et værk af Chohreh Feyzdjou.<sup>73</sup>

English translation: The article **concludes** with an outline of how they can be applied in an analysis of a work by Chohreh Feyzdjou.

Table 7.6. DAWL words only occurring as phrasal elements

DAWL word	Phrases
<i>almindelighed</i> , noun (generality)	<i>i almindelighed</i> (in general) <i>i al almindelighed</i> (generally)
<i>bekostning</i> , noun (expense)	<i>på bekostning af</i> (at the expense of)
<i>forvejen</i> , noun (ahead)	<i>i forvejen</i> (ahead)
<i>fremmest</i> , adjective (foremost)	<i>først og fremmest</i> (first and foremost)
<i>henblik</i> , noun (concerning)	<i>med henblik på at</i> (in anticipation of)
<i>hensyn</i> , noun (consideration)	<i>med hensyn til</i> (concerning) <i>tage hensyn til</i> (take into consideration) <i>under hensyn til</i> (in view of) <i>uden hensyn til</i> (irrespective of)
<i>hobe</i> , verb (heap)	<i>hobe sig op</i> (accumulate)
<i>mente</i> , noun (number carried)	<i>in mente</i> (in mind, number carried)
<i>munde</i> , verb (result)	<i>munde ud i</i> (result in)
<i>overens</i> , adverb (correspond)	<i>stemme overens med</i> (agree with)
<i>stede</i> , noun (place)	<i>til stede</i> (present)
<i>sigt</i> , noun (sight)	<i>på længere sigt</i> (in the long term) <i>på kort sigt</i> (in the short term) <i>på sigt</i> (at sight)
<i>stadighed</i> , noun (steadiness)	<i>til stadighed</i> (constantly)

<sup>73</sup> Petersen, A. R. (2012). Kulturel erindring i migrationens splintrede spejl. *Passepartout*. *Skrifter for Kunsthistorie*, 18. årgang (33), 16–30.

Table 7.7 lists the items that both occur in phrases and as single-word units. The table provides the phrases in which they occur in the AcaDan Corpus most frequently with the phrase in which the word occurs most frequently. It also gives examples of these items occurring as single-word units from the AcaDan Corpus.

Table 7.7. DAWL words occurring as phrasal elements and as single-word units

DAWL word	Phrases	Occurrence as independent word	English translation
<i>grad</i> , noun (degree, extent)	<i>i høj grad</i> (to a high degree) <i>i stigende grad</i> (increasingly) <i>til en vis grad</i> (to a certain degree)	hvis temperaturen falder til under 5 <b>grader</b> C (Natural Science) <sup>74</sup>  (...) hvilket forklares med en stigende <b>grad</b> af metabolisk regulering af foderoptagelsen. (Natural Science) <sup>75</sup>	If the temperature drops to below 5 <b>degrees</b> C  (...) which is explained by an increasing <b>degree</b> of metabolic regulation of feed intake.
<i>fald</i> , noun (fall)	<i>i hvert fald</i> (definitely) <i>i værste fald</i> (at worst) <i>i så fald</i> (in that case) <i>i givet fald</i> (if occasion should arise)	Der er sket en stigning siden 2000 for Kristendemokraterne, Dansk Folkeparti og Venstre, og et <b>fald</b> for SF. (Social Science) <sup>76</sup>	English translation: An increase has taken place since 2000 for the Christian Democrats, the Danish Folk Party and Venstre, and a <b>decrease</b> for SF.
<i>stand</i> , noun (condition; state; profession; social class)	<i>i stand til</i> (able to) <i>ude af stand til</i> (unable to)	Det var en mægtig styrke at have rod i en klassebevidst <b>stand</b> , styrket af økonomisk opgang og socialt fællesskab. (Natural Science) <sup>77</sup>	There was a powerful strength in being rooted in a class-conscious <b>profession</b> , strengthened by financial growth and a social community.  It has great or some significance for their

<sup>74</sup> Jensen, A. M. D., Mortensen, B., & Paaske, K. (2012). Pesticidforbrug og pesticidbelastning på golfbaner. Miljøstyrelsen

<sup>75</sup> Nørgaard, P., Nielsen, M. V., Helander, C., Eknæs, M., & Nadeau, E. (2015). Drægtige fårs foderoptagelse aftager med stigende tyggetidsindeks. *Faar*, 80(1), 8–11.

<sup>76</sup> Kosiara-Pedersen, K. (2014). Partimedlemmernes deltagelse og syn på partidemokrati 2000-2012. *Politica - Tidsskrift for Politisk Videnskab*, 46(3), 274–295.

<sup>77</sup> Kærgård, N., & Dalgaard, T. (2014). Dansk landbrugs strukturudvikling siden 2. verdenskrig. *Landbohistorisk Tidsskrift*, 2014(1-2), 9–33.

		Det har stor eller nogen betydning for deres golfspil, at banen er i god <b>stand</b> . (Natural Science) <sup>78</sup>	golf that the course is in good <b>condition</b> .
<i>trods</i> , noun, adverb, preposition (despite)	<i>på trods af</i> (contrary to) <i>til trods for</i> (in spite of)	(...) at det er medierne, der (overvejende) har magten og sætter dagsordenen <b>trods</b> skærpede redaktionelle vilkår. (Humanities) <sup>79</sup>	(...) that it is the media that (predominantly) has the power and sets the agenda <b>despite</b> severe editing conditions.
<i>tværs</i> , adverb (across)	<i>på tværs af</i> (across)	Den anden mulighed er, at indskriftbåndet er begyndt på stenens ene side og er fortsat <b>tværs</b> over toppen af stenen. (Humanities) <sup>80</sup>	The other possibility is that the inscription band is started on one side of the stone and then continued <b>across</b> the top of the stone.
<i>vidt</i> , adverb, adjective (wide)	<i>for så vidt</i> (in so far as)	Det er arbejdsgruppens opfattelse, at behandling med MDT er <b>vidt</b> udbredt blandt både fysioterapeuter og kiropraktorer. (Health Science) <sup>81</sup>	It is the opinion of the task force that treatment with MDT is <b>widely</b> used among physiotherapists and chiropractors.

Another important aspect to focus on when teaching these phrasal elements is their collocational behaviour in terms of adjectives. A word like *grad* can be modified by different adjectives, and the Word sketch analysis carried out in Sketch Engine shows that *høj* (high) (a DAWL word) and *stigende* (increasing) are two most used adjectives with *grad*. Table 7.8 shows which other adjectives are used to modify *grad* as part of the phrase *i MODIFIER grad* in the AcaDan corpus. DAWL words besides *grad* are marked in bold.

<sup>78</sup> Jensen, A. M. D., Mortensen, B., & Paaske, K. (2012). Pesticidforbrug og pesticidbelastning på golfbaner. Miljøstyrelsen

<sup>79</sup> Kristensen, N. N. (2009). Det er et spørgsmål om at gøre det nemt, men ikke at føre pennen. N O R D I C O M - Information, 31(1-2), 81–112.

<sup>80</sup> Nielsen, M. L. (2012). "Og efter sin elskede"?! Danske Studier, 2012, 5–23.

<sup>81</sup> Kjær, P., Junker, K., Kongsted, A., Schjøttz-Christensen, B., Møller, C., Ris Hansen, I., ... Melbye, M. (2015). National klinisk retningslinje for ikke-kirurgisk behandling af nyopstået rodpåvirkning i nakken.

Table 7.8. Modifiers in the phrase ‘i x grad’ in the AcaDan corpus

<i>i mindre grad</i>	less so
<i>i særlig grad</i>	particularly
<i>i nogen grad</i>	to some extent
<i>i overvejende grad</i>	considerably
<i>i varierende grad</i>	to a varying degree
<i>i samme grad</i>	to the same extent
<i>i hvilken grad</i>	to which degree
<i>i ringe grad</i>	to a very small extent
<i>i større grad</i>	more so
<i>i udpræget grad</i>	decidedly
<i>i tilstrækkelig grad</i>	sufficiently
<i>i væsentlig grad</i>	essentially
<i>i en sådan grad</i>	to an extent that
<i>i en eller anden grad</i>	to some extent
<i>i større eller mindre grad</i>	more or less so

This analysis of the phrasal elements concludes the functional categorisation of the DAWL words. In the next section, I will summarise the findings and briefly discuss some of the issues that have arisen in this categorisation process. I will in particular address the applicability of Hirsh’s functional framework, and I will discuss how academic words are multifunctional and sometimes polysemous. In doing so, I also include some methodological considerations and limitations in relation to carrying out a functional categorisation based on a lists of words.

#### 7.4. Discussion and limitations

In this chapter, I have presented Study 3, which provided a functional categorisation of the DAWL words. In particular, the purpose of Study 3 was to elucidate the functional properties of the DAWL words, and thus give a more usage-based description of the DAWL essential to our understanding of the functional nature of Danish academic vocabulary, and for future pedagogical applications of the DAWL. The functional categorisation was carried out through two analyses. The first analysis, which operated with three macro-categories, the academic workflow, descriptive words, and discourse-organising words, showed that most of the DAWL words fell into these categories. However, a small portion of words were categorised as technical, institutional, and as phrasal elements. The second analysis took its point of departure in Hirsh’s (2004, 2010) analytical framework developed from the Hallidayan metafunctions of textual, ideational, and interpersonal. The above-mentioned three

additional categories were kept in the second analysis, and thus Hirsh's framework was expanded upon. Specifically, the technical and phrasal elements categories were added as separate components to the framework and the institutional category replaced the other bodies' sub-category of the textual metafunction. Moreover, a category of stance replaced the authoritative category in the interpersonal metafunction. Hirsh's framework with the mentioned additions and changes proved to be highly applicable for analysing the DAWL words. This in turn supports the claim made in Study 2 that the DAWL is representative of a Danish academic vocabulary, not just in relation to frequency and lexical coverage, but also in relation to functional properties of the words.

I have also described a small number of DAWL words with general as well as technical meanings, and while the technical meaning is important for subject matter understanding, the general academic nature of these words is underlined by the fact that these words are used mostly in somewhat figurative and non-concrete senses. The functional analysis thus highlights the overlap between academic and discipline-specific vocabulary described in the vocabulary circle in Chapter 2, and the need to focus on academic words that fall into this category.

In fine-tuning the framework as described above, one runs the risk of presenting a simplified depiction of the functional properties of academic vocabulary. For example, it is clear from the categorisation of the DAWL words into the three sub-categories of the metatextual category that more analyses using concordance data and also statistical representation as in Hirsh (2004, 2010) would provide a clearer depiction of the functions of these words. In particular, such analyses would prove or disprove if an academic word does indeed fall into the suggested category. As an example, I intuitively categorised *fremstille* (give an account of) as a metacommunication word, but when looking at its occurrences in the AcaDan Corpus, it is not used in this function. However, the intention behind adding a sub-category of metacommunication to the metatextual category was to illustrate which DAWL words can be used for this important feature of academic writing. As such, attempting a functional categorisation of a word list at all is a way of clearly suggesting which words are used for which purposes. In doing so, the context in which the words occur should be taken into account as Hirsh (2004) did in his study.

Further, the categorisation of the DAWL words emphasised the multifunctionality of academic vocabulary as many of the DAWL words fell into more than one functional category. Arguably, this is not only due to academic words having multiple functions dependent on context, but it can also be seen as a result of how the categorisation analysis was carried out. The analysis of the DAWL word

was carried out by assigning the words to the different categories and sub-categories without considering all the contexts, i.e. by going through all concordance data for each word in the AcaDan Corpus. This can be considered a weakness of the study. Undoubtedly, a more comprehensive categorisation analysis that either follows the methodology of many lexical bundle studies concerned with functional categorisation of these (e.g. Ädel & Erman, 2012; Biber et al., 2004; Chen & Baker, 2010; Simpson-Vlach & Ellis, 2010), or that assigns the occurrence of academic words in a certain number of academic texts, as Hirsh (2004, 2010) did, would contribute to a more valid representation of the functions of Danish academic vocabulary. Hence, the analyses of Study 3 should be seen as a basis for further research into the functions of academic vocabulary in Danish. Nonetheless, Study 3 does advance our understanding of which words can be used for performing different academic functions, and thus verifies the academic nature of the inventory of the DAWL.

In the following, I will discuss some pedagogical implications of the functional categorisation of the DAWL words. Firstly, in the teaching of Danish to academics, the institutional words should be given attention, as they are important for their integration in the workplace. This set of words should be supplemented by the terminology lists developed by the University of Copenhagen (Københavns Universitet, 2008) which contain Danish and English terminology related to administration. Moreover, Study 3 lends empirical basis to teaching materials on academic writing. The word categories established by Stray Jørgensen (2004) and described in Chapter 2 of this thesis could be further developed, using the findings of Study 3, and including demonstration of which words are used for which purposes in academic writing. Thus, from a pedagogical standpoint, Study 3 also adds a productive aspect to the DAWL.

## 7.5. Rationale for Study 4

While the functional analyses carried out in Study 3 have contributed to a deeper understanding of the nature of Danish academic vocabulary, these analyses only pertain to the words selected for the DAWL. The DAWL words met certain criteria of range, frequency, and dispersion detailed in Study 2. As such, there are words in the AcaDan Corpus that would have been included, if different cut-off points (e.g. dispersion values of 0.60 or 0.70 instead of 0.80) had been employed in the selection criteria. Study 4 will focus on these words by supplementing the DAWL with items morphologically and semantically related to the DAWL words, and which have met some of the cut-off points applied in Study 2. It is expected that the added words will, to a high degree, fall into the same functional categories as the DAWL words they are related to. Study 4 will thus expand the number of words that



can be used for fulfilling the functions set forth in the functional framework developed in Study 3 on the basis of Hirsh (2004, 2010). In the next chapter, Chapter 8, Study 4 is presented.

## Chapter 8. Study 4 – Supplementing the DAWL

### 8.1. Introduction and research questions

The focus of this chapter is on the methodology, analyses and findings of Study 4. Study 4 partly departs from the quantitative approach taken in Study 2 and in other academic word list studies by considering the morphological and semantical nature of some of the lemmas that fell outside the cut-off chosen for the dispersion criterion in relation to the DAWL. In Study 2, a dispersion cut-off value at 0.80 was chosen after trialling extractions based on 0.70 and 0.60 values. Thus, in the development of the DAWL, a large group of words that satisfied the first two criteria of range and frequency were excluded because they did not satisfy the dispersion criterion. The motivation for Study 4 was that the dispersion analyses carried out in Study 2 showed that among the lemmas with dispersion values between 0.60 and 0.79 in the AcaDan Corpus there were lemmas that could be considered academic lemmas. Therefore, in Study 4, lemmas that in Study 2 satisfied the first two criteria of range and frequency and had dispersion values between 0.60 and 0.79 were analysed according to their morphological and semantical relatedness to the DAWL words. The purpose of Study 4 was thus to investigate if these lemmas could be added to the DAWL. By adding lemmas to the DAWL according to semantical and morphological relatedness, Study 4 moreover explored how a resulting list of the DAWL lemmas and related lemmas could be organised into groups of related items, thus moving from a strictly lemma-based list to a list of word groups containing both inflections as well as closely related derivations comparable to the concept of word families (Bauer & Nation, 1993). The research question guiding this supplementation analysis was:

- 1) How many lemmas fulfilling the DAWL criteria of range and frequency and with dispersion values between 0.60 and 0.79 are related morphologically and semantically to lemmas in the DAWL?

To measure the contribution of the added lemmas, Study 4 also measured the lexical coverage of the supplemented DAWL, the S-DAWL and compared it to that of the DAWL.

This chapter is divided into eight sections. Section 8.2 focuses on the methodology of Study 4 and Section 8.3 reports on the findings of Study 4 followed by a summary in Section 8.4. In Sections 8.5 to 8.7 I discuss different issues pertaining to the findings of Study 4 including limitations. The chapter is concluded with rationale for the final chapter of this thesis, Chapter 9.

## 8.2. Supplementing the DAWL

As described above, the basis for this supplementation analysis comprised lemmas which satisfied the two criteria of range and frequency used for developing the DAWL in Study 2, but which did not meet the third criterion of occurring in the AcaDan Corpus with a minimum dispersion value of 0.80. Thus, the lemmas explored in Study 4 are those that occurred in all three sub-corpora representing the academic disciplines of Humanities, Social Science, and Natural and Health Sciences. Further, they occurred with significant higher frequency in the AcaDan corpus than in the general language corpus, Journalisten.dk. Only lemmas with a dispersion value above 0.60 but below 0.80 were used. Using these lemmas as the basis for the supplementation analysis ensured that the adding of words to the DAWL was primarily objective in that these lemmas fulfil two out of the three criteria used in the identification of academic words.

The procedure itself for adding lemmas to the DAWL comprised comparing the DAWL with these lemmas, which had already been extracted in the dispersion analysis for word selection in Study 2. All error items had been removed as part of the dispersion analysis in Study 2. The lemmas were organised into two lists as detailed in Table 8.1.

*Table 8.1. The basis for the supplementation analysis*

<b>Basis for the supplementation analysis</b>	
0.60 list	0.70 list
1,181 lemmas with dispersion values between 0.60 and 0.69	956 lemmas with dispersion values between 0.70 and 0.79

These two lists of in total 2,137 lemmas thus comprised the basis for supplementing the DAWL. In comparison, the DAWL contains 770<sup>82</sup> lemmas. For a lemma in either of the two lists to be added to the DAWL, it had to satisfy two criteria:

- 1) A lemma must be morphologically related to a DAWL lemma either by derivation or by compounding.
- 2) The meaning of a lemma must be closely related to that of a DAWL lemma.

---

<sup>82</sup> Note that 758 lemmas were selected from the AcaDan Corpus for the DAWL and 12 items were added as there were 12 lemmas that occurred as two parts of speech. Therefore, the final number of words in the DAWL is 770.

This means that even if a lemma is morphologically related to a DAWL lemma, the meaning of it must be close to the meaning of the DAWL lemma. For example, even though *færdiggørelse* (completion) contains the same stem as the DAWL lemma *færdighed* (skill), the meaning of *færdiggørelse* is so far from that of *færdighed* that it cannot be included in the S-DAWL. Before going into detail about how the supplementation was carried out, I will in Section 8.2.1 outline some basic principles for word-formation in Danish using the DAWL lemma *anvendelig* (usable, applicable) as an example.

### 8.2.1. Morphological relatedness: derivations and compounds

In Danish as in English, derivations are formed through affixation. Usually, by the adding of a prefix or a suffix to the stem of a word, a new word is created often with a new part of speech (Hansen & Heltoft, 2011, p. 241). For example, if the suffix *-lse* is added to the verb *anvende* (use), it becomes a noun, *anvendelse* (use, application). Likewise, by adding the suffix *-lig*, *anvende* turns into an adjective, *anvendelig* (usable). Not all derivations include a change of part of speech. For example, by adding the prefix *u-* to the adjective *anvendelig* (usable), the part of speech stays the same, but the meaning of *uanvendelig* (unusable) is the opposite of *anvendelig*. Table 8.2 groups together the different derivations (in bold) of *anvende*, which could all form part of a hypothetical level 6-word family in Danish (Bauer & Nation, 1993). Inflections, which also occur via affixation, are provided to show how extensive a word family for this item would be.

Table 8.2. Derivations and inflections of *anvende*

<b>VERB</b>	<i>anvende</i>	<b>infinitive</b>
	<i>anvender</i>	present tense
	<i>anvendes</i>	present tense passive
	<i>anvendte</i>	past tense
	<i>anvendtes</i>	past tense passive
	<i>anvendt</i>	past participle
	<i>anvendende</i>	present participle
	<i>anvend</i>	imperative
<b>NOUN</b>	<i>anvendelse</i>	indefinite singularis
	<i>anvendelses</i>	indefinite singularis possessive
	<i>anvendelsen</i>	definite singularis
	<i>anvendelsens</i>	definite singularis possessive
	<i>anvendelser</i>	indefinite pluralis
	<i>anvendelsers</i>	indefinite pluralis possessive
	<i>anvendelserne</i>	definite pluralis
	<i>anvendelsernes</i>	definite pluralis possessive
<b>NOUN</b>	<i>anvendelighed</i>	indefinite singularis
	<i>anvendeligheds</i>	indefinite singularis possessive
	<i>anvendeligheden</i>	definite singularis
	<i>anvendelighedens</i>	definite singularis possessive
<b>ADJECTIVE</b>	<i>anvendelig</i>	common gender
	<i>anvendelige</i>	plural/definite
	<i>anvendeligt</i>	neuter
<b>ADJECTIVE</b>	<i>uanvendelig</i>	common gender
	<i>uanvendelige</i>	plural/definite
	<i>uanvendeligt</i>	neuter
<b>ADVERB</b>	<i>anvendeligt</i>	
<b>ADVERB</b>	<i>uanvendeligt</i>	

Compounding is another way of creating new words in Danish. A compound consists of at least two individual and independent lexical items with the final element deciding the part of speech of the compound (Lundskær-Nielsen & Holmes, 2010, p. 634). In contrast to English, most compounds are written as one string of characters in Danish, sometimes with a conjoining ‘s’ or ‘e’ (*forskningsprojekt*, *bindeled* (research project, connecting link)). In relation to the notion of word families, the question is how to decide their memberships of word families. The following compounds

with *anvendelse* as the first component could be added to the hypothetical word family in Table 8.2, (taken from the Danish Dictionary (Det Danske Sprog- og Litteraturselskab, n.d.-a)): *anvendelsesformål* (application purpose), *anvendelsesmulighed* (possible application), *anvendelsesområde* (field of application), and *anvendelsesorienteret* (application-oriented) (taken from the Danish Dictionary (Det Danske Sprog- og Litteraturselskab, n.d.-a)). However, because of the second components of these compounds which relate to rather different issues, the question is if these compounds should form their own word families.

### 8.2.2. Procedure for supplementing the DAWL

I will now describe the operationalisation of derivation and compounding as introduced above in the supplementation analysis. The first step of the analysis involved comparing the 0.60 and 0.70 lists to the DAWL. Lemmas in the two lists related to a DAWL lemma as either derivations or compounds were added to a table containing the DAWL lemmas in alphabetised order. In cases in which the morphological and/or semantical relationship could not be decided on, the Danish Dictionary (Det Danske Sprog- og Litteraturselskab, n.d.-a) was consulted for etymological information and for meanings. A lemma in the 0.60 and 0.70 lists was categorised as a **derivation** if it shared the same stem as a DAWL lemma. Thus, for the DAWL lemma *neutral* (neutral), the lemmas *neutralitet* (neutrality) and *neutralisere* (neutralise), which occur in the 0.60 list and in the 0.70 lists, respectively, were added to the DAWL since in meaning they are also closely related. Derivations also included lemmas to which prepositions are added as prefixes. As such, the verbs *fravige* (deviate from) and *afvige* (diverge) were added to the DAWL lemma *vige* (retreat). In addition, derivations also included antonyms such as *uacceptabel* (unacceptable) which was added to the DAWL lemmas *accept* (accept, noun) and *acceptere* (accept, verb), to which *acceptabel* (acceptable) was also added. Antonyms were considered to fulfil the second criterion of semantical relatedness in that their meaning is still close to that of the DAWL lemma (or the added lemma in the case of *uacceptabel-acceptabel*), albeit a negation of that meaning. In relation to the second criterion of semantical relatedness, in cases where the meaning of the derivation was markedly different from the meaning of the DAWL lemma, the derivation was not included in the S-DAWL. For example, the verb *anspore* from the 0.60 list was excluded because the meaning of *anspore* (to encourage or incite) is different from the DAWL lemma *spore*, which means to trace, track or monitor something. In contrast, the lemma *opspore* (track down, trace) was added as the meaning of this word is almost synonymous to the DAWL lemma *spore*.

For a **compound** in the 0.60 and 0.70 lists to qualify as a supplement lemma to the DAWL, it had to contain a DAWL word or a added derivation and be semantically related to the DAWL word. Again, the close relationship to an item already included in the DAWL was a condition for being considered as a potential S-DAWL item. As an example, for the DAWL lemma *analyse*, the following lemmas were found to supplement the DAWL: *analysemetode* (method of analysis), *analysemodel* (analysis model), *analyseniveau* (level of analysis), *analysestrategi* (analysis strategy), and *analysetilgang* (analytical approach). The compound *erkendelsesteori* (cognition theory) is an example of a lemma from the 0.70 list that is composed of the derivation *erkendelse* (recognition) to the DAWL word *erkende* (recognise, realise).

Alternative spellings of a word were counted as separate lemmas if the alternative spelling in fact was an added lemma. That was for example the case with the alternative spelling of *spekter* (spectre), *spektrum* which occurred in the 0.60 list. Thus, the S-DAWL contains two lemmas, *spekter* and *spektrum*, even if these are only alternative spellings of the same word. This was done to make it transparent that the alternative spelling satisfies the vocabulary selection criteria of range and frequency applied in Study 2. Conversely, the alternative spelling *resurse* to the DAWL lemma *ressource* (resource) is not listed as a separate lemma as it did not occur in the 0.60 and 0.70 lists. The inclusion of alternative spellings is motivated by the fact that these alternative spellings are part of the spelling norm in Danish as described in Chapter 6. In cases in which a lemma had more than one part of speech, the item was listed in the table as separate lemmas if they occurred as separate parts of speech in the AcaDan Corpus. For example, the item *fremme* was listed as a verb (promote), as an adverb (ahead), and as a noun (advancement) because it occurred as such in the AcaDan Corpus.

After an initial analysis of the 0.60 and 0.70 lemmas' relatedness to the DAWL words, the table containing the DAWL words and the added lemmas was checked by another researcher. This researcher was asked to assess whether the added lemmas were related morphologically and semantically to the DAWL words. Her assesment was then carefully considered and changes according to her assessment were made to the initial table. This resulted in a final table with the DAWL lemmas and the related lemmas compiled that formed the basis for the remaining analyses of Study 4. The findings of this analysis is detailed in Section 8.3. In the next section, I briefly describe how the 552 added lemmas were transformed into a base word list for lexical coverage analysis.

### 8.2.3. Measuring the lexical coverage of the S-DAWL

To measure the lexical coverage of the resulting S-DAWL, the added lemmas were added to the DAWL base word list developed in Study 2. A list of the added lemmas alone was also created. The procedure for doing so followed closely the procedures for developing base word lists described in Studies 1 and 2, and I refer to Chapters 5 and 6 for details on this issue. Here, I will summarise some considerations central for preparing the added lemmas and the S-DAWL for lexical coverage analysis. First, the base word list must include all occurring parts of speech of an item. As such, the lemma *fremme* mentioned above, was listed with inflections for all three parts of speech, and the different inflections for *spekter* and *spektrum*, respectively were listed as well. Secondly, as in Study 2, only likely occurring inflections were added to each lemma so that the base word list reflects actual language use and not what is paradigmatically possible. For example, possessive forms of past and present participles were not included in the base form list. A last consideration was the occurrence of repeated forms. As in Studies 1 and 2, all repeated forms were marked with a hashtag to make them unreadable for the lexical coverage analysis programme which cannot count repeated items. The final S-DAWL base word list contains 1,308 lemmas and 7,919 tokens.

The AntWordProfiler programme (Anthony, 2014) was used for measuring the lexical coverage of the S-DAWL over the AcaDan Corpus, the Second Academic Language Corpus, and over the General Language Corpus (see Chapter 4 for descriptions of these three corpora). The lexical coverage of the S-DAWL was also measured over the four academic disciplines of the AcaDan Corpus. The results of the lexical coverage analyses are reported on in Section 8.3.5 of this chapter.

## 8.3. The S-DAWL

This section is divided into two parts. The first part details the morphological nature of the added lemmas in the S-DAWL in relation to the DAWL lemmas. I also report on the overlap between the added lemmas and Danish general high frequency vocabulary. In the second part, I describe the S-DAWL in relation to dispersion, frequency of occurrence, and part of speech distribution. The lexical coverages of the S-DAWL and of the added lemmas alone are also provided.

In total, 552 lemmas were added to the DAWL. The 552 added lemmas constitute 41.80% of the S-DAWL, which in full contains 1,322 lemmas. Out of the 770 DAWL lemmas, 326 DAWL lemmas were related morphologically and semantically to the added lemmas with dispersion values between 0.60 and 0.79. Table 8.3 exemplifies which lemmas are added to which DAWL lemmas. The DAWL lemmas are listed so related items are placed in the same row in the table. For example, the DAWL



lemmas *fokus* (focus, noun) and *fokusere* (focus, verb) are listed after each other followed by the added lemmas *fokusering* (focusing), *fokusområde* (area of focus), and *hovedfokus* (main focus). Accordingly, the S-DAWL, provided in Appendix G, is ordered so that the added lemmas follow the DAWL lemmas. In addition, morphological and semantical related DAWL lemmas are also grouped together.

Table 8.3. Examples of DAWL lemmas and the added lemmas<sup>83</sup>

DAWL lemma		Derivations (lemma)			Compounds (lemma)			
<i>afhængig</i>		<i>uafhængig</i>	<i>afhængighed</i>	<i>uafhængighed</i>	<i>afhængighedsforhold</i>			
<i>baggrund</i>					<i>baggrundsinformation</i>	<i>baggrundsviden</i>		
<i>definere</i>	<i>definition</i>	<i>omdefinering</i>	<i>veldefineret</i>					
<i>fokus</i>	<i>fokusere</i>	<i>fokusering</i>			<i>fokusområde</i>	<i>hovedfokus</i>		
<i>hensigt</i>	<i>hensigtsmæssig</i>	<i>uhensigtsmæssig</i>			<i>hensigtserklæring</i>			
<i>indhold</i>		<i>indholdsmæssig</i>	<i>indeholde</i>					
<i>kategori</i>	<i>kategorisere</i>	<i>kategorisering</i>	<i>kategorisk</i>					
<i>mål</i>	<i>målestok</i>	<i>målbar</i>	<i>måle</i>	<i>målelig</i>	<i>målrette</i>	<i>målgruppe</i>	<i>målsætning</i>	
<i>Nuanceret</i>	<i>nuancering</i>	<i>nuancere</i>	<i>unuanceret</i>					
<i>perspektiv</i>	<i>perspektivere</i>	<i>perspektivering</i>			<i>tidsperspektiv</i>			
<i>reference</i>		<i>referere</i>			<i>referenceramme</i>	<i>referencepunkt</i>		
<i>specifik</i>		<i>specificere</i>		<i>uspecificeret</i>				
<i>tydelig</i>	<i>tydeliggøre</i>	<i>tydeligvis</i>	<i>tydeliggørelse</i>	<i>tydelighed</i>				
<i>udelukke</i>		<i>udelukkende</i>						
<i>værdi</i>	<i>værdifuld</i>	<i>værdimæssig</i>			<i>værdikonflikt</i>	<i>værdisætte</i>	<i>værdisætning</i>	<i>signalværdi</i>

<sup>83</sup> English translation are provided in Appendix G.

### 8.3.1. Description of the added lemmas: abbreviations, compounds, and derivations

In this section, I describe how the 552 added lemmas were related morphologically to the DAWL words in the form of compounds and derivations. Table 8.4 shows the distribution of the added lemmas according to their dispersion values and types of morphological relatedness. The morphological relatedness between the 326 DAWL lemmas and the added lemmas are either in the nature of derivations (70.16%) or compounds (29.11%) with the exception of a few abbreviations (0.70%). Concerning abbreviations, the DAWL lemma *eventuel* (potential) is often used in its adverbial form *eventuelt*, in the abbreviated form *evt.*, which occurs in the 0.70 list and has been included in the S-DAWL. Likewise, the abbreviated form, *hhv.* from the 0.60 list, was added to the DAWL lemma *henholdsvis* (respectively). Moreover, to the DAWL abbreviation *jf.* two different abbreviations (*jf* and *jvf*) of the imperative form of the verb *jævnføre* (compare) were included in the S-DAWL.

Table 8.4. Distribution of the added lemmas according to dispersion and morphological relatedness

	0.60-0.69	0.70-0.79	Total
<b>Derivations</b>	185	203	388
<b>Compounds</b>	115	46	161
<b>Abbreviations</b>	3	1	4
<b>Total</b>	304	251	552

As can be seen from Table 8.4, the number of derivations in either of the two lists is relatively near each other while the number of compounds in each of the two lists diverges notably with the 0.60 list offering the highest number of compounds. It is the 0.70 list that offers the fewest added lemmas, but on the other hand it offers the largest number of derivations compared to the 0.60 list. In fact, almost 81 percent of the added lemmas with dispersion values between 0.70 and 0.79 are derivations. Conversely, the 0.60 list offers the most compounds which make up almost 40 percent of the added lemmas with dispersion values between 0.60 and 0.69. In Section 8.3.1.1, I look more closely at the lemmas related via compounding to the DAWL lemmas, and then I move on to describe the derivations in Section 8.3.2.2.

#### 8.3.1.1. Compounds

In terms of the compositions of the 161 added compounds, three patterns emerge. Firstly, some of them consists of a DAWL word, and a word that does not occur in either the DAWL or the two 0.60 and 0.70 lists such as *læsefærdighed* (reading skill) in which only *færdighed* (skill) occurs in the

DAWL, or *selvforståelse* (self-knowledge) in which only *forståelse* occurs in the DAWL. Another example of this is *tidsperspektiv* (time perspective) in which *tid* is not found in either lists. Secondly, others are made up of an added lemma and a word also not included in either lists. For example, the added compound *afhængighedsforhold* (state of dependence) comprises a derivation of the DAWL lemma *afhængig* (dependent) and another DAWL lemma, *forhold* (state, condition). Thirdly, there are compounds in which both words are DAWL word. Examples are *rækkefølge* (order), *udviklingstendens* (development tendency), *udviklingsproces* (developmental process), and *naturforhold* (nature). As can be seen from the examples given here, the majority of the compounds are nouns (136 out of 161). The remaining compounds are adjectives (19), verbs (five) and adverbs (one). I return to the part of speech distribution of the full S-DAWL in Section 8.3.1.4.

Compounds are often rather specific or even technical in meaning (Moon, 1997, p. 56), and the semantic function of most compounds is to specify something with usually the first part as the specifier and the second part as the entity specified (Hansen & Heltoft, 2011, p. 240). This is also the case with the majority of the 161 compounds which are rather specific in meaning. For example, in the added lemma *baggrundsinformation* (background information), the word *baggrund* (a DAWL lemma) specifies what kind of information we are dealing with. Moreover, many of the S-DAWL compounds are more specific than the DAWL words they are related to. The DAWL lemma *proces* (process) is supplemented with four compounds, *arbejdsproces* (work process), *beslutningsproces* (decision process), *omstillingsproces* (readjustment process), and *udviklingsproces* (developmental process) in which the first parts of the compounds specify they type of process in question. Likewise, the compounds added to the DAWL lemma *videnskabelig* (scientific) refer to certain types of sciences (*religionsvidenskab* (comparative religion)) or ways of being scientific (*populærvideenskabelig* (popular), *tværvideenskabelig* (interdisciplinary)).

Considering the specification function of compounds and the fact that there is a relatively high number of compounds in the 0.60 and 0.70 lists, it could be the case that the added compounds are more discipline-specific in meaning than the DAWL lemmas they are related to. Additionally, the fact that most of the added compounds (161) occur with dispersion values below 0.70 as shown in Table 8.2, indicates that they are less evenly distributed in the AcaDan Corpus and thus used more frequently within some disciplines than in others. A closer look at the occurrences of these compounds in the AcaDan Corpus paints a more complex picture, however. For example, the lemma *produktionsforhold* (production) occurs in most instances in the Social Science and the Humanities disciplines and especially in discussions of Marxist theory. It also occurs in the sub-discipline of Food

and Resource Economics (Natural Science) referring to sea farming. There are no occurrences of it in the Health Science discipline which explains the relatively low dispersion value together with its low frequency. It could be the case that these compounds lie in a grey area between general and discipline-specific academic vocabulary. On the other hand, they may be specified versions of general academic words still used to describe processes and activities general to most academics as in the case of *baggrundsinformation* described above.

A comparison with the English Academic Collocation List (ACL) (Ackermann & Chen, 2013) in part supports the latter idea. This list contains collocations that are identified as being academic. Out of the 167 compounds in the S-DAWL (including compounds from the DAWL), 36 have equivalent entries in the ACL. These are listed in Table 8.5. The majority of the remaining 131 Danish compounds would assumedly also be translated into collocations and multiword units in English. However, a few of them actually translate into single word units in English. For example, the Danish word *målsætning* translates to ‘objective’ or ‘target’, and *videnskabsmand* translates into ‘scientist’. Many of the S-DAWL compounds do not occur as dictionary entries presumably because of their specificity. For example, *forskningslitteratur* (research literature), *hovedfokus* (main focus), and *forklaringskraft* (explanatory power) cannot be found in the Danish Dictionary (Det Danske Sprog- og Litteraturselskab, n.d.-a).

Table 8.5. S-DAWL compounds and their ACL equivalents

<b>S-DAWL compound</b>	<b>Collocation in the ACL</b>
<i>analysetilgang</i>	analytical approach
<i>arbejdserfaring</i>	professional experience
<i>baggrundsviden</i>	background knowledge
<i>enkeltindivid</i>	single individual
<i>feltstudium</i>	field research
<i>forklaringskraft</i>	explanatory power
<i>forskningsindsats</i>	research effort
<i>forskningslitteratur</i>	scholarly literature
<i>gradsforskel</i>	varying degree
<i>grundforskning</i>	basic research
<i>grundprincip</i>	basic principle
<i>hovedfokus</i>	main focus
<i>hovedformål</i>	primary purpose
<i>hovedresultat</i>	main findings, key findings
<i>hovedvægt</i>	first priority
<i>kendetegn</i>	distinctive feature, characteristic feature, key characteristic, defining characteristic, specific characteristic, main characteristics
<i>kernebegreb</i>	key concept, basic concept, central concept, defining concept, core issue
<i>kerneverdi</i>	core value
<i>kompetenceudvikling</i>	professional development, career development
<i>landbefolkning</i>	rural population
<i>ledetråd</i>	guiding principle
<i>levetilvilkår</i>	living conditions
<i>ligeværdighed</i>	equal opportunities, equal status
<i>målgruppe</i>	target audience
<i>naturforhold</i>	natural conditions
<i>naturvidenskabelig</i>	natural science
<i>nøglerolle</i>	key role
<i>organisationsstruktur</i>	organisational structure
<i>ressourcekrævende/ resursekrævende</i>	require resources
<i>slutresultat</i>	final result
<i>udviklingsproces</i>	developmental process
<i>udviklingstrin</i>	developmental stage
<i>hovedelement</i>	basic element, core element, essential element, key element, main element
<i>hovedårsag</i>	underlying cause, major cause, underlying reason, primary reason
<i>værdigrundlag</i>	shared values

Above, I have discussed the nature of the compounds added to the DAWL in terms of specificity and in relation to their status as general academic vocabulary. As discussed, some of these may be

described as belonging to the grey zone area between academic and technical vocabulary, as illustrated by the vocabulary circle in Chapter 2 (Figure 2.19). In Section 8.3.1.2, I look closely at the derivations among the added lemmas in terms of their part of speech and their composition in terms of affixation.

### 8.3.1.2. Derivations

As showed in Table 8.4, there were far more derivations (388) among the added lemmas than compounds (161). These derivations can be viewed in terms of parts of speech and in terms of affixation and this section is divided accordingly. First, I describe the parts of speech of the derivations and then I focus on how they are derived from the DAWL words in terms of prefixes.

#### *Parts of speech of the derivations*

The majority of the added derivations are nouns (186), as were the majority of the compounds, in contrast to the part of speech distribution of the DAWL in which verbs were the most frequent part of speech though closely followed by nouns. Adjectives are the second largest part of speech (111) among the added derivations. Verbs constitute 83 of the added derivations. There are six adverbs (*almindeligvis* (generally), *fremme* (ahead), *følgelig* (consequently), *givetvis* (certainly), *tydeligvis* (evidently), and *undtagelsesvis* (unusually)) and two conjunctions (*medens* (while), *samt* (plus)) among the added derivatives. The first conjunction, *medens*, is in fact an alternative spelling of the DAWL lemma *mens* but it also occurs in the 0.70 list which is why it is included among the added lemmas (see Section 8.2.1.). The second conjunction, *samt*, is added to the DAWL adjective *samtlig*. As can be seen from the six adverbs, four of them end with the suffix *-vis*, but they are related to the DAWL lemmas in different ways. For example, the two of them, *almindeligvis* and *undtagelsesvis*, are related to the DAWL nouns *almindelighed* (generality) and *undtagelse* (exception), while these three, *fremme*, *givetvis*, and *tydeligvis* are added to the DAWL adjectives *fremmest* (foremost), *given* (given), and *tydelig* (obvious). The adverb *følgelig* is related to the DAWL lemma *følge* (consequence) (which is one of the 12 items listed in the DAWL as more than one part of speech as it can both be a noun and a verb, see also Chapter 6).

If we look at the adjectives, nouns and verbs added as derivations to the DAWL, there is no clear pattern as to how these derivations and the DAWL lemmas interrelate in terms of part of speech. In other words, it cannot be said that derivations of e.g. the DAWL nouns are mostly in the form of verbs or vice versa. However, if we look at the 117 DAWL verbs for which we find adjectives among the added derivations, it is clear that only a few of the adjectives are antonyms to the DAWL verb. For

example, the DAWL verb *berøre* (touch, affect) is related to the adjective *uberørt* (untouched), and *uafhængig* (independent) occurs among the related items to the DAWL verb *afhænge* (depend) together with *uafhængig*. Some of the adjectives among the added derivations are either past participles (e.g. *veldefineret* (well-defined) which has been prefixed with *vel* meaning well) or present participles (e.g. *afgørende* (crucial) related to the DAWL verb *afgøre* (determine), and *foregående* (previous) related to the DAWL verb *foregå* (take place) of DAWL verbs which occur as separate lemmas in the AcaDan Corpus. As the DAWL is lemma-based, these forms are in fact included in the lemma, but listing *afgørende* and *foregående* as separate items in the S-DAWL may be reasonable considering that these words have their own entries in the Danish Dictionary (Det Danske Sprog- og Litteraturselskab, n.d.-a). Conversely, there is also an example of a DAWL adjective in the form of a present participle (*gennemgående* (continous, thorough)) being supplemented by the verb *gennemgå* (undergo).

If we look at the nouns among the added derivations, we can see that many of them are derived from verbs in the DAWL. The DAWL verbs *afgrænse* (define), *afhænge* (depend), *antage* (assume), *bedømme* (assess), *bekræfte* (confirm), *benævne* (designate), *betegne* (designate), *definere* (define), and *erkende* (recognise, realise), to name a few, are supplemented with the following nouns: *afgrænsning* (defination), *afhængighed* (dependence), *antagelse* (assumption), *bedømmelse* (assessment), *bekræftelse* (confirmation), *benævnelse* (designation), *betegnelse* (designation), *definition* (definition), and *erkendelse* (recognition, realisation). Also, semantically, these nouns are closely related to the DAWL verbs. As the number of verbs among the added derivations suggests, there are fewer instances of a DAWL noun being supplemented by a verb, but there are some. For example, the DAWL noun *eksempel* (example) is supplemented by the verbal derivation *eksemplificere* (exemplify) and the DAWL noun *indvandring* (immigration) is supplemented by the verb *indvandre* (immigrate). The occurrence of nominalisations among the added lemmas is not surprising considering that that this is a defining trait of academic language as described in Chapter 2.

### Prefixes

Above, I have described the derivations in terms of part of speech. In the following, I will look at the derivations in terms of the affixes that form them. Both the DAWL and the added derivations contain a large number of items to which prepositions such as *af*, *an*, *fra*, *frem*, *ind*, *på*, *sammen*, *til*, *ud*, and *under* (of, to, from, on, in, on, together, out, under) are added as prefixes. Below I only describe those



that are derived from DAWL lemmas without a prepositional prefix. Table 8.6 provides the derivations with prepositional prefixes and the DAWL lemmas they supplement in brackets. As can be seen from Table 8.6, the DAWL lemma *vise* (show) is an item that has many derivations with prepositional suffixes, but also the DAWL lemma *forstå* (understand, comprehend) has several derivations with prepositional prefixes. The prepositional prefixes most frequently occurring among the derivations are *af*, *for*, and *om*, each with five derivations followed by *op* with four derivations. A few of these derivations are phrasal verbs in that the prepositional prefix can be separate from the verb without any change in meaning. For example, the meaning of the two sentences below are the same.

(...) et af de kritiske irreversible elementer i inflammationsprocessen opstår, når inflammationscellerne **ophobes** (...).<sup>84</sup>

(...) et af de kritiske irreversible elementer i inflammationsprocessen opstår, når inflammationscellerne **hobes op** (...).

English translation: (...) one of the critical, irreversible elements in the inflammation process emerges when the inflammation cells **accumulate** (...).

Conversely, other of the derivations listed in Table 8.6 can only stand in the form of prefix plus item if the meaning is to be the same. To put it another way, the meaning of the lemma *afvise* (reject) is different from the meaning of *vise af* (signal in traffic).

---

<sup>84</sup> Bruun, G., & Bjørndal, L. (2013). Behandling af profund caries—baseret på evidens fra nyere kliniske undersøgelser. *Tandlægebladet*, 117, 322-9.

Table 8.6. Derivations with prepositional prefixes

Prefix	Added lemma	English translation	DAWL lemma
<b>af-</b>	<i>aftegne</i>	draw	<i>tegn</i>
	<i>afvige</i>	diverge, differ	<i>vige</i>
	<i>afvigelse</i>	deviation	<i>vige</i>
	<i>afvise</i>	reject	<i>vise</i>
	<i>afvisning</i>	rejection	<i>vise</i>
<b>an-</b>	<i>anvise</i>	show, assign	<i>vise</i>
<b>for-</b>	<i>fordobling</i>	doubling	<i>dobbelt</i>
	<i>forenkling</i>	simplification	<i>enkelt</i>
	<i>forforståelse</i>	preunderstanding	<i>forståelse</i>
	<i>fortolke</i>	interpret	<i>tolke</i>
	<i>fortolkning</i>	interpretation	<i>tolkning</i>
<b>fra-</b>	<i>fravige</i>	deviate	<i>vige</i>
<b>frem-</b>	<i>fremkalde</i>	induce	<i>kalde</i>
<b>før-</b>	<i>førnævnt</i>	before-mentioned	<i>nævne</i>
<b>ind-</b>	<i>indforstået</i>	informed	<i>forstå</i>
	<i>indvirkning</i>	impact	<i>bevirke</i>
			<i>medvirke</i>
<b>in-</b>	<i>indiskutabel</i>	indisputable	<i>diskutere</i> <i>diskussion</i>
<b>infra-</b>	<i>infrastruktur</i>	infrastructure	<i>struktur</i>
<b>med-</b>	<i>medføre</i>	entail	<i>føre</i>
<b>om-</b>	<i>omdefinering</i>	redefining	<i>definere</i> <i>definition</i>
	<i>omforme</i>	convert	<i>form</i>
	<i>omformulering</i>	reformulation	<i>formulering</i>
	<i>omformulere</i>	reformulate	<i>formulering</i>
	<i>omorganisering</i>	reorganisation	<i>organisering organisere</i>
<b>op-</b>	<i>ophobe</i>	accumulate	<i>hobe</i>
	<i>ophobning</i>	accumulation	<i>hobe</i>
	<i>opspore</i>	track down	<i>spore</i>
	<i>optegne</i>	record	<i>tegn</i>
<b>over-</b>	<i>overrepræsentation</i>	overrepresentation	<i>repræsentere</i> <i>repræsentativ</i>
	<i>overvurdere</i>	overestimate	<i>vurdering</i>
	<i>overvurdering</i>	overestimation	<i>vurdering</i>
<b>på-</b>	<i>påkalde</i>	invoke	<i>kalde</i>
	<i>påvise</i>	demonstrate	<i>vise</i>
	<i>påvisning</i>	proof	<i>visning</i>
<b>sammen-</b>	<i>sammenkoble</i>	link	<i>koble</i>

<b>semi-</b>	<i>semistruktureret</i>	semi-structured	<i>struktur</i>
<b>til-</b>	<i>tilgrundliggende</i>	underlying	<i>grundlag</i>
	<i>tilgrænsende</i>	adjacent	<i>grænse</i>
	<i>tilstand</i>	condition	<i>stand</i>
<b>ud-</b>	<i>udnytte</i>	utilize, exploit	<i>nytte</i>
	<i>udveksle</i>	exchange	<i>veksle</i>
	<i>udveksling</i>	exchange	<i>veksle</i>
<b>under-</b>	<i>underforstå undervurdere</i>	imply	<i>forstå</i>
		underestimate	<i>vurdering</i>

Turning to some of the other prefixes of the added derivations, it is interesting to note that while the DAWL contains no less than 41 lemmas with the prefix *be-*, there are only four examples of derivations with *be-* derived from DAWL lemmas without the *be-* prefix. The *be-* prefix can be described as a conversion prefix in that it either signals a transitivity of the word being prefixed as in *benytte* (make use of), or that something is made into something else as in *bearbejde* (prepare) (Lundskær-Nielsen & Holmes, 2010, p. 628). The four added derivations prefixed with *be-* are *begrunde* (motivate) and *begrundelse* (motivation) derived from the DAWL lemma *grunde* (base), and *besvare* (respond) and *besvarelse* (reply; solution) derived from the DAWL lemma *svare* (answer). Out of the 41 DAWL lemmas starting with the prefix *be-*, nine are supplemented with derivations, e.g. the DAWL lemma *beskæftige* (engage; employ) is supplemented by the derivation *beskæftigelse* (employment) and *bearbejdning* (preparation) is supplemented by *bearbejdelse* (processing). Most of the added lemmas to the nine DAWL lemmas starting with *be-* are nouns with the exceptions of the example just mentioned, together with *betydelig* (significant) to the DAWL lemma *betyde* (mean; signify), and *befolke* (populate) to the DAWL lemma *befolkning* (population).

The last form of prefixes that I will describe here is the kind of prefix that negatively influences the meaning of the prefixed item. I have already touched upon the use of the prefix *u-* to make an antonym of a word, and both the DAWL and the added derivations provide us with several examples of antonyms. Some of the antonyms have a positive counterpart in the DAWL or among the added lemmas such as *uvæsentlig* (inessential) (an added lemma) vs. *væsentlig* (essential) (a DAWL lemma) and *ubestemt* (undetermined) (an added lemma) and *bestemt* (determined) (a DAWL lemma). Another negative prefix is the *mis-* which is used to express the meaning of bad or wrongly (Lundskær-Nielsen & Holmes, 2010, p. 628), but there is only one occurrence of an added derivation with this prefix: *misforståelse* (misunderstanding) which supplements the DAWL lemma *forståelse* (understanding).

### *8.3.1.3. Overlap with general high frequency vocabulary*

Table 8.7 provides the overlap between the added lemmas and the 2,000 most frequently used lemmas in Danish also termed Danish general high frequency vocabulary in the studies of this thesis. The majority of the 552 added lemmas do not belong to Danish general high frequency vocabulary as only 45 lemmas occurred among the 2,000 most frequently used lemmas in Danish with 15 lemmas occurring in the first 1,000 frequency band, and 30 in the second 1,000 frequency band. This finding corresponds well with the fact that the added lemmas have lower dispersion values and are therefore not as generally used in the AcaDan Corpus as the 363 DAWL lemmas that overlapped with the 2,000 high frequency words in Danish. Of these 363 DAWL lemmas, 212 also overlapped with the 402 general high frequency lemmas that were identified as academic in Study 1. Thus, there were 189 general high frequency lemmas that did not overlap with the DAWL. The overlap between these 189 words and the added lemmas comprised only 16 lemmas (marked with an asterisk in Table 8.7). The fact that none of the 45 general high frequency lemmas added to the DAWL are compounds confirm the specific nature of the added compounds as described above.

Table 8.7. General high frequency words among the added lemmas

Frequency of occurrence		
<b>Occur among the 1,000 most frequently used lemmas in Danish</b>	<i>afvise</i>	<i>offentliggøre</i>
	<i>*forsker</i>	<i>*omfatte</i>
	<i>*forskning</i>	<i>opleve</i>
	<i>handle</i>	<i>organisation</i>
	<i>kritik</i>	<i>producere</i>
	<i>kritisere</i>	<i>rum</i>
	<i>*markere</i>	<i>samt</i>
	<i>*medføre</i>	<i>*tradition</i>
	<i>modsatte</i>	<i>*undersøgelse</i>
	<i>måle</i>	<i>virkelighed</i>
	<i>nation</i>	<i>vurdere</i>
<b>Occur among the second 1,000 most frequently used lemmas in Danish</b>	<i>*afslutning</i>	<i>indvandre</i>
	<i>*afslutte</i>	<i>karakter</i>
	<i>betingelse</i>	<i>*præge</i>
	<i>*betydelig</i>	<i>realitet</i>
	<i>dømme</i>	<i>repræsentant</i>
	<i>forholde</i>	<i>studie</i>
	<i>forvente</i>	<i>tilstand</i>
	<i>*fremme (adverb)</i>	<i>tredjedel</i>
	<i>*grundlæggende</i>	<i>uafhængig</i>
	<i>*handling</i>	<i>*udføre</i>
	<i>*henvisning</i>	<i>udnytte</i>
		<i>usædvanlig</i>

### 8.3.2. Description of the S-DAWL

In the preceding section, I have discussed the derivations added to the DAWL in relation to how they relate to the DAWL lemmas morphologically and semantically with a particular focus on their parts of speech and prefixes. In the next three sections, I move on to describe the full S-DAWL, the S-DAWL, in relation to dispersion, frequency, and part of speech distribution. The purpose of doing so is to show the differences and similarities between the two lists, the DAWL and the S-DAWL. By highlighting the properties of the S-DAWL in comparison with the DAWL, this description will further our understanding of Danish academic vocabulary.

#### 8.3.2.1. Dispersion of the S-DAWL

Table 8.8 shows the dispersion distribution of the S-DAWL. The 552 added lemmas comprise around 40 percent of the S-DAWL with the DAWL comprising the remaining 60 percent. These 60 percent

occur with a dispersion of 0.80 or more. Of the 552 added lemmas 303 or roughly 55 percent are from the 0.60 list which means they occur in the AcaDan corpus with dispersion values between 0.60 and 0.69. The other 45 percent of the added lemmas are from the 0.70 list, and, as such, they occur in the AcaDan Corpus with dispersion values between 0.70 and 0.79.

*Table 8.8. Dispersion of the S-DAWL lemmas in the AcaDan Corpus*

<b>0.60-0.69</b>	302	22.84%
<b>0.70-0.79</b>	250	18.91%
<b>0.80-0.89</b>	626	47.35%
<b>0.90-1.0</b>	144	10.89%
<b>Total</b>	1,322	100%

#### *8.3.2.2. Frequency distribution of the added lemmas*

Table 8.9 provides an overview of the frequencies of the added lemmas in the AcaDan Corpus including examples. As can be seen, the majority of the S-DAWL lemmas (423) occur with a frequency of less than 100. A substantial part (193) of the DAWL lemmas also occur with frequencies below 100, and the majority (630) occur with frequencies below 1,000. This finding corresponds with the fact that the added lemmas have lower dispersion values and are therefore not as generally used in the AcaDan Corpus. Around 17 percent of the DAWL lemmas occur with frequencies above 1,000. As shown in Table 8.9, only eight of the added lemmas occur this frequently in the AcaDan Corpus.

Table 8.9. Frequency distribution of the added lemmas in the AcaDan Corpus

Frequency of occurrence	Number of items	%	Added Lemmas (examples)
>3000	1	0.18	<i>undersøgelse</i>
2001-3000	1	0.18	<i>samt</i>
1001-2000	6	1.09	<i>forskning, handle, opleve, relation, rum, vurdere</i>
901-100	2	0.36	<i>anvendelse, studie</i>
801-900	3	0.54	<i>handling, individ, karakter</i>
701-800	7	1.27	<i>forsker, forvente, kritik, måle, omfatte, organisation, udføre</i>
601-700	3	0.54	<i>betydelig, grundlæggende, medføre</i>
501-600	8	1.45	<i>forholde, producere, præge, systematisk, tilstand, tradition, virkelighed, videnskab</i>
401-500	5	0.91	<i>formulere, indvandrere, konstruktion, nærværende, potentiale</i>
301-400	14	2.54	<i>analytisk, dokumentation, fortolkning, repræsentation, sandsynlighed, udbredelse</i>
201-300	27	4.89	<i>dokumentere, dynamisk, forholdsvis, hhv. kritisere, markere, påvirkning, resultere, udnytte, ydre</i>
101-200	53	9.60	<i>funktionel, nødvendighed, opløsning, principiel, prioritere, strukturere, udveksling</i>
4-100	422	76.45	<i>afklare, almindeligvis, enkeltstående, formalisere, fællestræk, generalisering, illustration, indskrænke, kategorisering, konsekvent, metodologisk, modellering, perspektivering, ubeskreven, veludviklet</i>
<b>Total</b>	<b>552</b>	<b>100</b>	

### 8.3.2.3. Part of speech distribution of the S-DAWL

The added lemmas were analysed in relation to part of speech in order to reach an understanding of their syntactical properties and to make a comparison to part of speech distribution of the DAWL items. The part of speech analysis was carried out by manually assigning each lemma a part of speech. The AcaDan Corpus and the Danish Dictionary (Det Danske Sprog- og Litteraturselskab, n.d.-a) were consulted in cases where the part of speech could not be determined straight away. Table 8.10 provides the part of speech distribution of the added lemmas including the number of abbreviations together with the part of speech distribution of the DAWL for comparison. As can be seen in Table 8.10, the added lemmas are primarily in the form of nouns and verbs, but also adjectives make up a

significant part of the added lemmas. There are no prepositions and pronouns and only two conjunctions among the added lemmas, echoing the low distribution of these word classes in the DAWL (see Chapter 6). Moreover, very few adverbs were found among the added items. Thus, the added lemmas primarily belong to open word classes in which new lexical items can be created such as compounds. This is part of the explanation for the high number of nouns among the added lemmas compared to the DAWL which contained slightly more verbs than nouns.

*Table 8.10. Part of speech distribution of the added lemmas and of the DAWL*

<b>Part of speech</b>	<b>0.60 added lemmas</b>	<b>0.70 added lemmas</b>	<b>DAWL</b>	<b>S-DAWL</b>
<b>Adjective</b>	60	69	178	308
<b>Adverb</b>	2	5	59	66
<b>Conjunction</b>	0	2	13	15
<b>Noun</b>	205	117	240	562
<b>Numerals</b>	0	0	5	5
<b>Prepositions</b>	0	0	7	7
<b>Pronouns</b>	0	0	12	12
<b>Verb</b>	32	56	247	335
<b>Abbreviation</b>	3	1	9	13
<b>Total</b>	<b>302</b>	<b>250</b>	<b>770</b>	<b>1,322</b>

### 8.3.3. Lexical coverage of the added lemmas and the S-DAWL

In this last section of Section 8.3, I give the results of the lexical coverage analyses of the added lemmas and the S-DAWL. The purpose of measuring the lexical coverage is to find out how the added lemmas contribute in coverage and whether the S-DAWL provides a more comprehensive list of academic vocabulary in comparison with the DAWL. Table 8.11 provides the coverage over the three corpora of both the S-DAWL and the DAWL (for comparison's sake), as well as of the added lemmas. Table 8.12 provides the coverages of the three lists in the four disciplines represented in the AcaDan Corpus.



Table 8.11. The coverage of the S-DAWL and the added lemmas over academic and general language

	<b>The AcaDan Corpus</b>	<b>The Second Academic Language Corpus</b>	<b>The General Language Corpus</b>
<b>S-DAWL</b>	28.19%	29.68%	19.72%
<b>DAWL</b>	26.09%	27.83%	18.96%
<b>The 552 added lemmas</b>	2.1%	1.85%	.76%

Table 8.12. The coverage of the S-DAWL and the added lemmas over the four disciplines in the AcaDan corpus

	<b>Health Science</b>	<b>Humanities</b>	<b>Natural Science</b>	<b>Social Science</b>
<b>S-DAWL</b>	27.23%	27.43%	26.25%	30.48%
<b>DAWL</b>	24.75%	25.65%	24.27%	28.19%
<b>The 552 added lemmas</b>	2.48%	1.78%	1.98%	2.29%

Measured over the AcaDan Corpus the added lemmas contribute with an additional two percent coverage in comparison with the DAWL (see Table 8.11). Their coverage over the General Language Corpus of less than one percent reflects the finding from Section 8.3.2.3 that these lemmas are primarily found outside the general vocabulary of Danish. When we look at the added lemmas' coverage in the four disciplines of Natural Science, Health Science, Humanities, and Social Science (see Table 8.12), we can see that the highest coverage, about 2.5%, is found in Health Science sharply followed by Social Science, about 2.3%. The explanation for this is probably that the Health Science sub-corpus contains a number of research articles from the sub-discipline of Public Health, a cross-disciplinary area of research that draws on methods from both the hard and the soft sciences. On the other hand, the added lemmas' coverage in Natural Science is higher than their coverage in Humanities which may be interpreted as if the added lemmas are the least specific in relation to Humanities. This coverage pattern changes when we look at the coverages of the full S-DAWL in these disciplines. The S-DAWL provides the highest coverage (30.48%) over the Social Science sub-corpus. This is due to the high coverage of the DAWL over this sub-corpus. In general, as can be seen from Tables 8.11 and 8.12, the coverage provided by the S-DAWL independent of language type, corpus, or discipline is greater than that of the DAWL. This is expected considering the increase in number of lemmas in the S-DAWL compared to the DAWL even if the coverage of these 552 added lemmas is relatively small. This in turn highlights the fact that the DAWL comprise almost 50 percent

general high frequency words as mentioned earlier and these words contribute substantially to the coverage of the DAWL and consequently also the S-DAWL. It should be noted, however, that the coverage of the 552 added lemmas is smaller than that of the DAWL lemmas with general high frequency vocabulary excluded. To put it in another way, the DAWL words that fall outside the first 2,000 lemmas in Danish have a higher coverage over the different corpora than the added lemmas. This points at the added lemmas being more specific and the important question is whether the S-DAWL is a better representation of Danish academic vocabulary than the DAWL if taking coverage as a determining factor. This is an issue I elaborate on in Section 8.5.

#### 8.4. Summary

In sum, 552 lemmas were included in the supplemented DAWL, the S-DAWL, which also comprised the 770 DAWL lemmas. In Section 8.3, I have given a description of the morphological nature of the added lemmas, and I have described the full S-DAWL in relation to dispersion, frequency, and part of speech distribution. I also measured the lexical coverage of the added lemmas alone and of the S-DAWL in academic and general language, and compared the results to the DAWL.

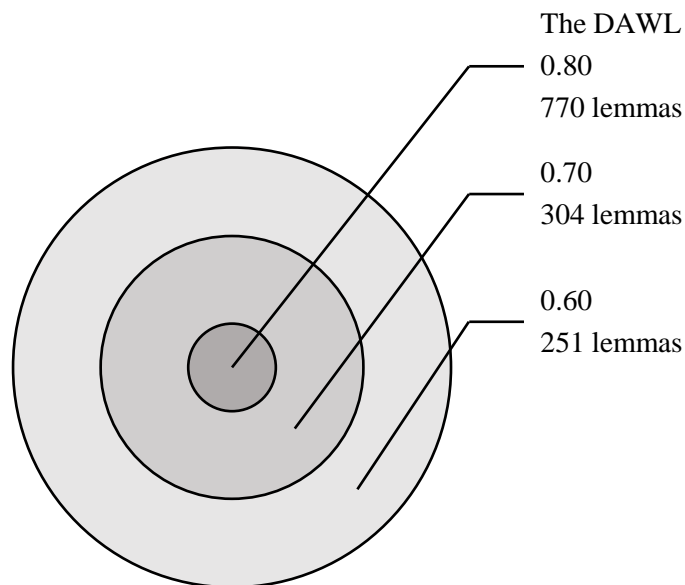
The 552 added lemmas were related to the DAWL lemmas as derivations (388), compounds (161), and abbreviations (4). In relation to the compounds, they tend to be more specific in meaning than the derivations and the DAWL words which has to do with the specifying nature of compounds in general. However, many of them seemed still to be general academic words used for describing academic activities shared across disciplines. In relation to the derivations, the majority of them were nouns, but there were also a substantial number of adjectives of which only a few were antonyms. The added lemmas were also analysed according to overlap with the 2,000 most frequently used lemmas of Danish (Det Danske Sprog- og Litteraturselskab, n.d.-d). This analysis showed that only a small number of them were in fact general high frequency vocabulary. Turning to the S-DAWL, more than half of it occurs with a dispersion above 0.80, which is expected, as the majority of this list is comprised by the DAWL. In terms of frequency of occurrence in the AcaDan Corpus, the majority of the S-DAWL occurs less than a 1,000 times in the AcaDan Corpus. This echoes the frequency distribution of the DAWL in the AcaDan Corpus. The part of speech distribution also resembles the DAWL with verbs and nouns being the most frequent parts of speech. The results of the lexical coverage analyses showed that even though a substantial number of lemmas were added, these lemmas only contributed to the coverage of the DAWL with a few percent in academic language. On

the other side, the coverage in academic language compared to that in general language was substantially higher which suggests that these words are indeed frequently used in academic language.

Thus, in this chapter I have shown that it is possible to add a more qualitative layer to the quantitative vocabulary selection carried out in Study 2 which provides us with additional knowledge of the nature of Danish academic vocabulary. In the three last sections of this chapter, I will discuss three issues: 1) Does the S-DAWL represent Danish academic vocabulary? (Section 8.5). 2) Is it possible to organise the S-DAWL according to how the items it contains relate to each other? (Section 8.6) 3) What are the limitations of Study 4 and what are the implications for further research? (Section 8.7).

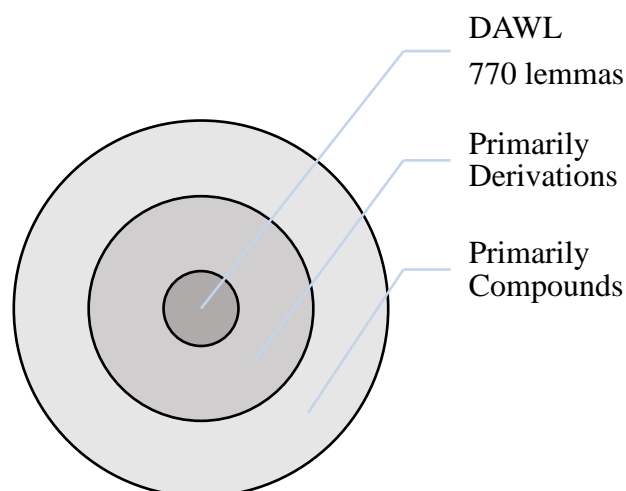
### 8.5. Danish academic vocabulary represented by the S-DAWL

The findings of Study 4 presented in this chapter have important implications for how Danish academic vocabulary can be conceptualised. In Study 2 (Chapter 6), strict criteria of frequency, range and dispersion were applied in the identification of Danish academic words following what other researchers have done in the development of academic word lists in different languages. Study 4 moved beyond this quantitative approach, and attempted to explore those words that did not meet all three criteria employed in Study 2. This means that the identified lemmas that were added to the DAWL in Study 4 may not qualify as a core academic vocabulary in the same degree as the DAWL can be said to represent Danish academic vocabulary. Thus, it can be helpful to envision the S-DAWL of Study 4 as 3 concentric circles with the DAWL, or what can be said to be the core of Danish academic vocabulary at the centre, and the added lemmas surrounding the centre (see Figure 8.1 below). This representation is based on the dispersion values of the lemmas. Specifically, the added lemmas can be divided into two groups according to dispersion in the AcaDan Corpus. The first group are the lemmas with dispersion values between 0.60 and 0.69, and the second group comprises the lemmas with dispersion values between 0.70 and 0.79. As concentric circles, the first group is the outer circle, while the second group is the inner circle surrounding the centre comprising the DAWL lemmas, or the core academic vocabulary of Danish. These lemmas have dispersion values between 0.80 and 0.98.



*Figure 8.1. The S-DAWL represented as a circle based on dispersion values*

In addition, the added lemmas can be categorised according to their morphological relatedness to the DAWL lemmas. In Section 8.3.1.1, we saw that many of the compounds are more specific in meaning than the derivations of which many were nominalisations of DAWL verbs. We also saw that it is not warranted to argue that all the added compounds are more specific in meaning than the derivations. The compounds are, however, in many cases more specific in meaning than the DAWL words they related to due to the specifying function that compounds generally have. This means that the added compounds, irrespective of their dispersion values, can be placed in the outer circle with the derivations in the circle nearest to the centre. The centre of the circle still comprises the DAWL words. Thus, we end up with two circles illustrating the S-DAWL, Figure 8.1 and Figure 8.2 below.



*Figure 8.2. The S-DAWL represented as a circle based on morphological relatedness*

The motivation for representing the S-DAWL in the way of a circle is to illustrate what exactly this supplementation analysis of Study 4 has resulted in. In using morphological relatedness to the already identified academic words as a criterion for inclusion into Danish academic vocabulary, Study 4 has in particular shed light on the nature and role of academic compounds and highlighted how several academic verbs have academic nominal counterparts, an issue that was also given attention in Study 3 of this thesis. As such, Study 4 adds to our knowledge of Danish academic vocabulary not only by expanding the number of words identified as academic but also, in line with Study 3, by shedding light on the nature of these words.

As shown in Table 8.3 and described in Section 8.3, the added lemmas can be organised in groups together with the DAWL lemma(s) they are related to morphologically and semantically. In the next section of this chapter, I discuss how Study 4 can be seen as a first step of organising a Danish academic word list in groups of related items, so-called word groups, similar to the concept of counting words as word families (a headword, its inflections and closely related derivations) (Bauer & Nation, 1993) described in Chapter 2.

## 8.6. Word groups

The DAWL and the subsequent S-DAWL are lemma-based and as such each item in the lists represent the base form of a lemma and its possible inflections. The morphological relatedness between the academic lemmas of Danish highlighted here in Study 4 gives cause to organise the S-DAWL items in groups, resembling to some degree the word family as we know it from English vocabulary

research. As shown in Table 8.1 and also demonstrated in Study 3, many of the DAWL verbs are related through nominalisation to DAWL nouns. Examples of these are listed below:

- *afdekke-afdækning* (uncover-uncovering)
- *beskrive-beskrivelse* (describe-description)
- *definere-definition* (define-definition)
- *konkludere-konklusion* (conclude-conclusion)
- *modificere-modification* (modify-modification)
- *placere-placering* (place-placement)
- *tolke-tolkning* (interpret-interpretation)
- *udforske-udforskning* (explore-exploration)
- *udvikle-udvikling* (develop-development).

This relationship was highlighted in the analysis of Study 4 since several DAWL words were found to have relatives of this kind in the pool of lemmas used for the analysis. For example, the verb *vurdere* (assess) can be added to the DAWL noun *vurdering* (assessment). Moreover, several DAWL lemmas can be grouped together with one or two related DAWL items such as the DAWL lemmas *forstå*, *forståelig*, and *forståelse* forming one DAWL word group. This group of words is expanded in the S-DAWL to include these derivations: *forforståelse*, *indforstået*, *misforståelse*, *uforståelig*, *underforstå*, *forståelsesramme*, *selvforståelse*, *verdensforståelse*. Table 8.13 is an attempt to organise these eight added lemmas in relation to the DAWL lemmas, which are given in bold. The lemma, *forstå*, is chosen as the headword as it is the most frequent one and constitutes the stem for the other lemmas. Expanding the original DAWL word group with these lemmas should be done in a way that reflects how closely related the added lemmas are to the DAWL lemmas. It should be noted here that this grouping of lemmas together is primarily descriptive on the basis of what has been identified in Study 2 and Study 4. Pedagogically, teaching and learning these different items from *indforstået* to *verdensforståelse* include additional focus on the affixes and elements of these derivations and compounds besides the context in which they occur. Moreover, the issue of compounds has particular implications for the word family construct. As mentioned earlier, in Danish, compounds are written as one string of characters equal to so-called transparent compounds in English. The question is if compounds should be counted and listed as separate word families (just like they are counted as separate lemmas), or if they should be added to the word family of one of the compounds.

Table 8.13. Grouping of lemmas related to *forstå*

<i>forstå</i>	
<i>forstå</i> (understand, comprehend)	<i>indforstået</i> (congenial) <i>underforstå</i> (imply)
<i>forståelse</i> (comprehension)	<i>selvforståelse</i> (selfknowledge) <i>forståelsesramme</i> (frame for understanding) <i>misforståelse</i> (misunderstanding) <i>forforståelse</i> (pre-understanding) <i>verdensforståelse</i> (world understanding)
<i>forståelig</i> (comprehensible)	<i>uforståelig</i> (incomprehensible)

While Study 2 gave us a list of lemmas which were described in relation to e.g. part of speech and general high frequency vocabulary, Study 4 provides us with insight into what a Danish academic word list would look like if a word family approach had been taken. This insight is highly usable for pedagogical purposes in that the supplementation analysis has resulted in a number of words that intuitively could be deemed academic. Thus, the question of why is this or that word not included in the DAWL can partly be answered with reference to the S-DAWL. Study 4's description of the morphological relationship between academic words in Danish, both the DAWL lemmas and the added ones, adds to the insight into Danish academic vocabulary provided by Study 2.

## 8.7. Limitations and further research

Study 4 has a number of limitations. Firstly, a potential weakness of Study 4 is the role that semantical relatedness has played in the supplementation analysis. Semantic relatedness was assessed subjectively by an additional researcher and myself, and no strict criteria for this notion were established. Doing so would undoubtedly have contributed to a more thorough analysis, which could also have made use of, for example, the functional framework set forth in Study 3, a procedure which could add to the validity of Study 4. Secondly, another weakness of Study 4, closely connected to the issue of semantical relatedness is the choice to include derivations and compounds that were derived from the added lemmas. For example, the added compound *erkendelsesteori* (cognition theory) was added because it is related to the DAWL word *erkende*. It could be argued that while it would be reasonable to include *erkendelse* (recognition, realisation), had it occurred in the 0.60 or 0.70 lists, adding *erkendelsesteori* would perhaps be taking things too far, also considering the semantical relatedness of this item to the verb *erkende* (recognise, realise). However, using the lemmas from the AcaDan Corpus that met two out of the three criteria applied in the identification and selection of

Danish academic vocabulary as a basis for adding additional words to the DAWL ensured an objective basis for more subjective analyses.

Finally, a potential weakness of Study 4 is related to the limitations of Study 2. The use of Juilland's Dispersion measure (Juilland & Chang-Rodríguez, 1964) was mentioned in Chapter 6 as a potential limitation as it has been found to decrease in sensitivity when used on corpora with many parts (Biber et al., 2016). This means that there is a risk that the added words are less evenly distributed in the AcaDan Corpus as assumed.

## 8.8. Rationale for Chapter 9

This chapter has focused on Study 4 which is the final study reported on in this thesis. The four studies have explored the lexis of professional Danish academic writing with a particular focus on the identification of Danish academic vocabulary. In the next and final chapter of this thesis, Chapter 9, the findings of the four studies will be summarised and the contributions and implications of these findings will be highlighted.



## Chapter 9. Key findings and conclusion

### 9.1. Introduction

The purpose of the research presented in this thesis was to investigate the academic words used in Danish professional academic writing, and in this way address the lack of research-based knowledge of this specific lexical inventory in Danish. In particular, the research has been carried out with two overall aims:

- 1) To identify a Danish academic vocabulary and provide a description of this lexical inventory.
- 2) To investigate the nature of general high frequency vocabulary in academic language.

These aims were accomplished through four studies, each with their own research questions. The data for the four studies consisted of corpora developed specifically for this research project. In particular, a corpus of professional written academic Danish, the AcaDan Corpus (presented in Chapter 4) was developed for the purpose of identifying and exploring Danish academic words.

In Chapter 1, in the introduction to the research presented in this thesis, I outlined research that shows how Danish as an academic language can be a challenge for both Danish L1 students as well as students with Danish as their second or foreign language. I took these findings to suggest that an explicit focus on academic vocabulary would help these students in acquiring the necessary academic language skills of which academic vocabulary is an essential component. As shown in Chapter 2, school-related vocabulary has received some attention in Danish as a second language research, but this research has primarily addressed the language of primary school, and focused on the overlap between general and technical vocabulary through the notion of pre-technical vocabulary. Moreover, in relation to vocabulary, Danish research on this topic has been primarily lexicographically oriented. While the previous research carried out, as described in Chapter 3, has provided us with invaluable knowledge of Danish lexis, it has only to a small degree been applied to the teaching and learning of Danish. I was in particular inspired to investigate Danish academic vocabulary from the applied linguistic research on word lists for the teaching of English for both academic and general purposes. I have, however, stated throughout this thesis that the primary aim of my research was to provide a systemic description of Danish academic vocabulary. Such a description can in turn form a research-based foundation for the development of pedagogical tools and materials for the teaching of Danish as an academic language in both L1 and L2 perspectives. As such, the research carried out in the project should be seen as important first step towards a more explicit focus on vocabulary to the

teaching of Danish. This is underlined by the fact that I, in my dissemination of this PhD project to practitioners, have met several teachers of Danish as a second and foreign language who have asked for more specific directions in regards to what vocabulary to teach their students. However, care should be taken to apply the findings from the project directly to language teaching and learning, an issue that I address in Section 9.2.4 below.

The four individual studies were discussed in the chapters presenting the studies. The focus of this chapter is therefore primarily to give an overview of the **key findings and contributions** of the four studies as well as the pedagogical implications (Section 9.2). In Section 9.3, I give suggestions for **further research** before I end this chapter and thus this thesis with some **concluding remarks** (Section 9.4).

## 9.2. Key findings and contributions

The overview given in this section of the key findings and contributions of the four studies is organised thematically according to these four issues: **1) the nature of Danish general vocabulary, 2) the nature of Danish academic vocabulary, 3) the overlaps between vocabulary categories, and 4) pedagogical implications.**

### 9.2.1. The nature of Danish general vocabulary

The first study, presented in Chapter 5, focused on the relationship between Danish general vocabulary and academic language by exploring the coverage of Danish general high frequency vocabulary, defined as the 2,000 most frequently used words in Danish. By exploring the occurrences of Danish general words in different text types using lexical coverage as the method of analysis, Study 1 moved beyond the frequency-based research on Danish general vocabulary carried out by Bergenholtz (1992), Ruus (1995), and the Danish Language and Literature Society (Det Danske Sprog- og Litteraturselskab, n.d.-d). Study 1 has contributed to our knowledge and understanding of the nature of general vocabulary by 1) providing the lexical coverage of the 2,000 most used lemmas in Danish in different text types and academic disciplines, and 2) applying the framework of Lexical Frequency Profiling (Laufer & Nation, 1995a) to a study of Danish. Related to the first contribution, the results of Study 1 confirmed findings from English and French lexical research that the most frequently occurring 2,000 words in the language cover a large proportion of the words in any text type, also academic texts. However, in contrast to findings from English and French (Cobb & Horst, 2004), the results of Study 1 strongly indicate that the 2,000 most frequent words in Danish provide less coverage independent of text type compared to English and French.

The results also demonstrated that the first 1,000 words in Danish cover a very high proportion of a text compared to the second 1,000 words, which stresses the importance of these very frequent words for language comprehension and production. Moreover, Study 1 demonstrated that general vocabulary behaves differently according to text type, as the coverage of general words in academic language was lower than that in general language, and thus highlighted the nature and importance of high frequency vocabulary in relation to academic language. Specifically, comparing the coverages of Danish general words in texts from the two sub-disciplines of Medicine and Information Science showed that the Medicine texts make less use of these words than the Information Science texts. In this way, Study 1 demonstrated differences in the coverage of general high frequency vocabulary in different academic disciplines echoing the findings of Coxhead (2000). Methodologically, Study 1 employed the Lexical Frequency Profiling framework, described in Chapters 2 and 5, which included the development of a base word list for analysing the vocabulary load of texts. Thus, Study 1 is an important first step in exploring the vocabulary load of different text types in Danish, and the lexical richness and sophistication of learner texts using this framework.

#### 9.2.2. The nature of Danish academic vocabulary.

The primary aim of the research presented in the thesis has been to investigate the nature of Danish academic vocabulary. Study 2 (presented in Chapter 6) comprised a corpus-based identification of 770 Danish academic words (the DAWL) operationalised as those words that occur more frequently in the AcaDan Corpus than in a general language corpus with an even distribution in the AcaDan Corpus. Study 2 also provided us with a description of this lexical inventory, which showed that the most frequent parts of speech in Danish academic vocabulary are verbs and nouns, followed by adjectives. The DAWL was evaluated by measuring its coverage over different corpora, a widely used evaluation method in word list studies (Miller & Biber, 2015). These analyses confirmed that the DAWL is representative of Danish academic vocabulary, as the list's coverage over two academic language corpora was notably higher than its coverage over a general language corpus. There were small differences between the coverages provided by the DAWL in the four academic disciplines of Health Science, Humanities, Natural Science, and Social Science, echoing the coverage results for the Academic Word List developed by Coxhead (2000). Through the identification and description of Danish academic vocabulary, Study 2 is an important contribution to research related to the language and vocabulary of educational settings as it pinpoints the nature of a specific group of Danish words that are essential for producing and comprehending academic discourse. This specification of Danish academic vocabulary was continued in Studies 3 and 4 as described below.

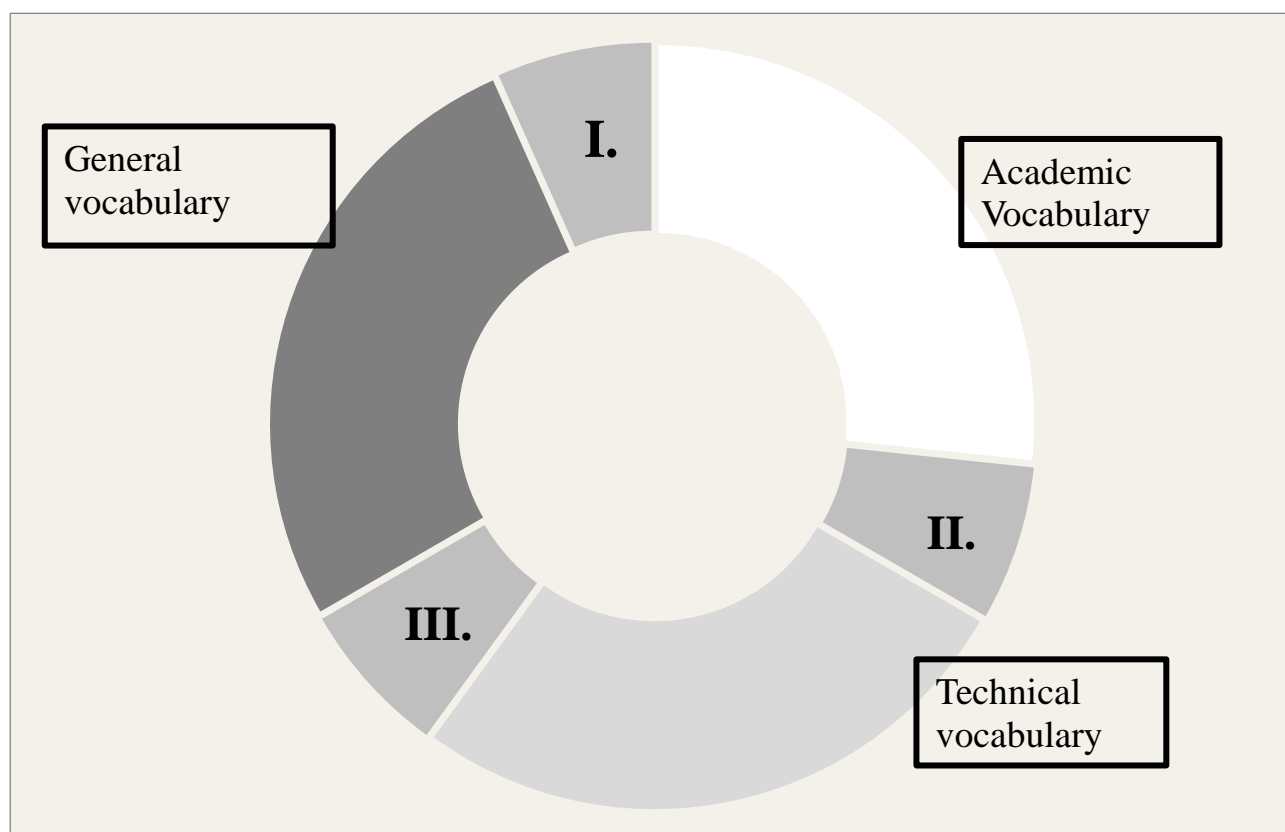
Study 3 contributed to accomplishing the primary research aim of this thesis by exploring the functions of academic vocabulary, and providing a categorisation of the DAWL words. The classification of the academic words in relation to functional use also highlights the important functions these words play in understanding and presenting academic research and the underlying processes. It stresses the need to emphasise this for pupils and students when they encounter academic language in their education, and have to develop their academic literacy skills. Study 3 is the first corpus-based Danish study on the functions of academic vocabulary, and provides an important first step for further analyses of the functions of academic lexis. Furthermore, the study adds to our understanding of the validity of functional categorisations of academic vocabulary, e.g. the analytical framework developed by Hirsh (2004, 2010), by demonstrating that these functions are to a great extent transferable across languages.

Study 4 further investigated Danish academic vocabulary by exploring those words in the AcaDan Corpus that met the first range and frequency criteria established for the identification and vocabulary selection for the DAWL, but which were excluded in the DAWL due to the chosen cut-off value of 0.80 set for the dispersion criterion. Those words with dispersion values between 0.60 and 0.79 in the AcaDan Corpus that were morphologically and semantically related to the DAWL words were identified, described, and added to the DAWL, resulting in a new, expanded word list, the S-DAWL comprising 1,323 words. The S-DAWL proved to have a higher coverage over the academic corpora than the DAWL. Moreover, Study 4 also pointed to the importance of investigating the lexical items that are included or excluded as a result of methodological decisions taken in word list development. Hence, Study 4 presented additional analyses combining frequency-based criteria with semantic and morphological criteria to enable the identification of related lexical items that intuitively belong to academic vocabulary, e.g. pairs of semantically related nouns and verbs describing the same academic processes (e.g. *analyse-analysere, definere-definition*).

### 9.2.3. The overlaps between vocabulary categories

In Chapter 2, I presented the vocabulary circle, which shows the three categories of general, academic, and technical words that can be found in academic texts. The circle also illustrated the overlaps between general and academic vocabulary (overlap zone I.), between academic and technical vocabulary (overlap zone II.), and between general and technical vocabulary (overlap zone III.) by having a grey area between each category. In this section, I will outline how the four studies of this

thesis demonstrate the overlaps or fuzzy boundaries between general and academic vocabulary and between academic and technical vocabulary.



*Figure 9.1. The vocabulary circle – the vocabulary categories of academic language*

In relation to overlap zone I., Study 1 showed that 20 percent of general high frequency words are academic words in that they occurred more frequently in the AcaDan Corpus than in a general language comparison corpus with an even dispersion. Additionally, an overlap analysis carried out in Study 2 between the DAWL and the 2,000 most frequent words showed that the DAWL comprises a high number of general words, 363 to be exact. These findings emphasise that there is a substantial overlap between the two vocabulary categories of general and academic vocabulary. Accordingly, the findings of Study 1 demonstrate the importance of not excluding high frequency vocabulary in the process of extracting an academic word list, i.e. not setting a frequency cut-off point as was done in the development of the Academic Word List (Coxhead, 2000) or using stop lists as was done in the development of the Swedish Academic Word List (Jansson et al., 2012; Ribeck et al., 2014). Moreover, the findings from Study 1 and Study 2 give cause to investigate further the overlap zone between general and academic vocabulary, especially the role of polysemy, and the need to focus on

this when teaching academic literacy skills to L1 and L2 language learners embarking on academic studies throughout the educational system.

In relation to the overlap between academic and technical vocabulary (overlap zone II.), the notion of discipline-dependent polysemy, that is, academic words taking on additional technical senses dependent on the disciplinary context in which they occur (cf. Malmström et al., 2018), was underlined in the findings of Study 4 related to compounds. The majority of the compounds that were added to the DAWL in Study 4 occur with dispersion values between 0.60 and 0.69. These somewhat lower dispersion values suggest that some of these compounds can be regarded as more discipline-specific words. Considering that many compounds are rather specific or technical in meaning (Moon, 1997), there is reason to argue that compounds in Danish academic language should be further investigated. In this way, Study 4 confirms findings from English lexical research that collocations, both academic and technical, are important for academic language use (cf. Ackermann & Chen, 2013; Henriksen & Westbrook, 2017). Also, in Study 3, the notion that academic words can be technical was supported by the findings related to a small group of DAWL words that had both academic and technical senses. In this way, all four studies shed new light on the overlap or fuzzy boundaries between vocabulary categories.

In these three sections on the key findings and contributions of the research presented in this thesis, I have demonstrated how the four studies have advanced our understanding of Danish academic vocabulary and of Danish general vocabulary in relation to academic language, and the two overall aims stated above of the research presented in this thesis have thus been accomplished. In addition, I have highlighted the contributions the studies have made in relation to lexical coverage studies in other languages, to methodological choices made in the extraction of academic vocabulary, to the usefulness of functional categorisation of academic words, and to our understanding of the overlap zones between the categories of general, academic, and technical vocabulary. In the next section, I discuss the pedagogical implications of the four studies.

#### 9.2.4. Implications for pedagogy

The principal aim of the research of this thesis has been to provide a linguistic description of a lexical inventory that we have limited knowledge about in Danish, i.e. a systemic description of Danish academic vocabulary. However, throughout the thesis, and especially in Study 1, I have touched on vocabulary learning issues as the notion of academic word lists is rooted in applied linguistics and

the need to develop wordlists that can be used by language learners, teachers, materials developers and researchers. Therefore, I will briefly present some pedagogical implications of the four studies.

In my dissemination of this PhD project, I have met several teachers of Danish as a second and foreign language who have asked for word lists of both general and academic vocabulary so that they have more specific directions in regards to what vocabulary to teach their students. Study 1 was primarily carried out to come nearer to an understanding of the relationship between general vocabulary and academic language, and the new knowledge provided by this study highlights the importance of general vocabulary and the important role of polysemy for high frequency vocabulary items. This knowledge should be used by teachers and course and material designers. However, as stated in Study 1, much more research is needed before we have pedagogically useful lists of general vocabulary in Danish. Most importantly, the basis for such a list should be both spoken and written Danish similar to Nation's BNC/COCA lists (2012) which was based on both types of language. Moreover, the development of a pedagogical useful general word list for Danish should not only be created using objective, frequency-based criteria. Also subjective criteria such as the ones used for the creation of the *eirfa Graidd* (Morris, 2010; Morris & Meara, 2014), a word list of Welsh core vocabulary (described in Chapter 3), should be employed. Nonetheless, the used list of the 2,000 most frequently used lemmas for the lexical coverage analyses can be considered a first step in developing a pedagogical word list representing general vocabulary in Danish.

The need for a Danish academic word list in a pedagogical perspective is also evident from the research on different groups of tertiary education students' challenges with academic language as reported in Chapter 1. Study 2 identified which words are parts of Danish academic vocabulary, but care should be taken before using the DAWL and the S-DAWL for teaching academic words. This list is representative of professional academic vocabulary, and while tertiary students are expected to read original research, they also read textbooks, and we do not know whether the DAWL or the S-DAWL can prepare students for textbook language. On the other hand, an argument can be made for the usefulness of these two lists in a pedagogical perspective in that 1) both lists have a high number of general words which we can expect to find in many different academic genres, and 2) the non-general high frequency items may be the words that tertiary students need to master for both receptive and productive purposes. Furthermore, related to the overlap between general and academic vocabulary is the issue of polysemy. It may be that many of the general words in the DAWL have specific academic meanings and functions that students need to know to be proficient in

understanding and producing academic language. Lack of knowledge of the specific meaning in academic contexts may lead to misunderstandings and inaccurate renderings of academic content. Finally, in relation to the DAWL, Study 3 provided a functional categorisation of the DAWL words that will be useful in the teaching of Danish academic writing by giving directions in relation to language use. Existing guidelines for Danish academic writing are primarily based on the authors' language intuitions and experience with teaching academic writing. The DAWL, the S-DAWL, and the functional classification of the items in the list provide an empirical basis for refining and expanding this knowledge and the guidelines presented.

### 9.3. Suggestions for further research

The findings of the four studies as well as their limitations offer a number of suggestions for further research, which will be outlined below according to these themes: Research on the nature of Danish academic vocabulary, research on the nature of Danish general vocabulary, and contrastive research on academic vocabulary.

In relation to the nature of **Danish academic vocabulary**, much more research on the lexis of texts from different academic disciplines is needed to further our understanding of this lexical inventory and the relationship between general, academic and technical vocabulary, and language use. Specifically, the AcaDan Corpus can be used for studies on the phraseological and collocational behaviour of Danish academic vocabulary and the use of this type of lexis in different academic disciplines, including the creation of lists of academic formulas and collocations in line with the work carried out in English by Simpson-Vlach and Ellis (2010) and Ackermann and Chen (2013). In addition, Danish spoken academic discourse should also be explored both in relation to the vocabulary of it, but also in relation to the differences between written and spoken academic language in line with what has been done by Biber (2006). Finally, there is also the issue of students' academic language use. In Danish, this has been explored by Blom et al. (2017) who investigated university student papers for linguistic deviations from the orthographic norms of Danish. However, an investigation of the lexis used by students in their papers would contribute to our knowledge of Danish academic language from a student perspective. As such, the DAWL and S-DAWL could be used for measuring the lexical richness of student texts in line with Malmström, Pecorari, and Shaw (Malmström et al., 2018) who analysed Swedish university students' use of words from the Academic Vocabulary List (Gardner & Davies, 2014). Additionally, the functional analysis of Study 3 can be



used for extending the research carried out by Holsting et al. (Holsting et al., 2017) who looked at Danish university students' use of metadiscourse when referring to sources.

In relation to research on the nature of **Danish general vocabulary**, more corpus-based research, including more spoken language corpora, is needed for developing valid and reliable lists of the most frequent words of Danish that can be used for the teaching and learning of Danish including vocabulary testing. This work should also include more qualitative approaches to word lists development as indicated in Section 9.2.4 and by Dang (2017) who suggests interviews with teachers and learners as a way of obtaining information of vocabulary knowledge and vocabulary teaching strategies. Moreover, word lists representing the entire frequency bands of Danish vocabulary in line with Nation's BNC/COCA lists (2006, 2012) can be used to measure the relationship between reading and listening comprehension and vocabulary knowledge. This area has not yet been explored in depth within Danish as a second and foreign language (cf. Chapter 2). Word lists of general Danish together with the DAWL and S-DAWL can also be used for developing a new Danish version of the Vocabulary Levels Test (Nation, 1983, 1990; N. Schmitt et al., 2001) (cf. Albrechtsen et al., 2008).

Related to the issue of general vocabulary is the overlap between this type of vocabulary and academic vocabulary. Using corpora representing different academic disciplines and levels of education could provide us with knowledge on the distributional behaviour of general and academic vocabulary in different academic discourses. Further, the texts used in tertiary teaching should be investigated. We need knowledge about what types of texts university students read and the language of these texts.

The increasing use of English as a medium of instruction in Danish higher education also has implications for the use of Danish academic language. More **contrastive research** is needed to understand the differences between Danish and English academic language, and accordingly prepare students to be proficient in both languages. The work undertaken in setting up the AcaDan corpus, extracting the DAWL and the S-DAWL, and supplying the word class and functional description of the lexical inventory provide data that can be used in future comparisons with English (cf. Shaw & Vassileva, 2009) or with Swedish and Norwegian other languages (cf. Johansson et al., 2017).

#### 9.4. Concluding remarks

Each of the four studies of this thesis has provided new research-based knowledge of Danish academic language, Danish general vocabulary and Danish academic vocabulary. The four studies are also important by adding to our understanding of issues explored in studies of English and other languages and by shedding light on the distinction between general, academic, and technical

vocabulary and the overlaps between these vocabulary categories. While the research reported in this thesis primarily has been occupied with providing a systemic description of Danish academic vocabulary, it has laid the foundation for more corpus-based research with pedagogical perspectives in Danish, including the development of word lists of both general and specialised vocabulary.

## References

- Ackermann, K., & Chen, Y.-H. (2013). Developing the Academic Collocation List (ACL)—A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes*, 12(4), 235–247.
- Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, 31, 81–92.
- Albrechtsen, D., Haastrup, K., & Henriksen, B. (2008). *Vocabulary and writing in a first and second language: Processes and development*. New York: Springer.
- Andersen, L., & Henriksen, B. (2014). Ord og tekst [Words and texts]. In H. L. Andersen, S. S. Fernandez, D. Fristrup, & B. Henriksen (Eds.), *Fremmedsprog i gymnasiet - teori, praksis og udsyn* (pp. 157–167). Samfundslitteratur.
- Anthony, L. (2014). AntWordProfiler (Version 1.4.1). Tokyo, Japan: Waseda University. Retrieved from <http://www.laurenceanthony.net/software>
- Bailey, A. L., Butler, F. A., & Sato, E. (2007). Standards-to-Standards Linkage Under Title III: Exploring Common Language Demands in ELD and Science Standards. *Applied Measurement in Education*, 20(1), 53–78. <https://doi.org/10.1080/08957340709336730>
- Baker, M. (1988). Sub-technical vocabulary and the ESP teacher: An analysis of some rhetorical items in medical journal articles. *Reading in a Foreign Language*, 4, 91–105.
- Bardel, C. (2016). 3 The lexicon of advanced L2 learners. In K. Hyltenstam (Ed.), *Advanced Proficiency and Exceptional Ability in Second Languages*. Berlin, Boston: De Gruyter. <https://doi.org/10.1515/9781614515173-006>

- Bardel, C., & Lindqvist, C. (2011). Developing a lexical profiler for spoken French L2 and Italian L2: The role of frequency, thematic vocabulary and cognates. *EUROSLA Yearbook*, 11(1), 75–93. <https://doi.org/10.1075/eurosla.11.06bar>
- Bauer, L., & Nation, I. S. P. (1993). Word Families. *International Journal of Lexicography*, 6(4), 253–279. <https://doi.org/10.1093/ijl/6.4.253>
- Becher, T. (1989). *Academic tribes and territories*. Bristol: The Society for Research into Higher Education and Open University Press.
- Becher, T., & Trowler, P. (2001). *Academic tribes and territories: intellectual enquiry and the culture of disciplines* (2nd ed). Philadelphia: Open University Press.
- Bergenholtz, H. (1992). *Dansk frekvensordbog. Baseret på tekster fra danske romaner, ugeblade og aviser fra 1987-1990 [Danish frequency dictionary. Based on texts from Danish novels, magazines, and newspapers from 1987-1990]*. Gad.
- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers* (Vol. 23). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371–405.
- Biber, D., Reppen, R., Schnur, E., & Ghanem, R. (2016). On the (non)utility of Juilland's D to measure lexical dispersion in large corpora. *International Journal of Corpus Linguistics*, 21(4), 439–464. <https://doi.org/10.1075/ijcl.21.4.01bib>
- Blom, J. N., Rathje, M., le Fevre Jakobsen, B., Holsting, A., Hansen, K. R., Svendsen, J. T., ... Lindø, A. V. (2017). Linguistic deviations in the written academic register of danish university students. *OSLA Oslo Studies in Language: Academic Language in a Nordic Setting - Linguistic and Educational Perspectives*, 9(3), 169–190.

- BNC Consortium. (2007). British National Corpus [Text]. Retrieved 5 June 2018, from <http://www.natcorp.ox.ac.uk/>
- Bowker, L., & Pearson, J. (2002). *Working with specialized language: a practical guide to using corpora*. London: Routledge.
- Brezina, V., & Gablasova, D. (2015). Is There a Core General Vocabulary? Introducing the New General Service List. *Applied Linguistics*, 36(1), 1–22.  
<https://doi.org/10.1093/applin/amt018>
- Campion, M. E., & Elley, W. B. (1971). *An academic vocabulary list*. Wellington: New Zealand Council for Educational Research.
- Carlund, C., Jansson, H., Kokkinakis, S. J., Prentice, J., & Ribeck, J. (2012). An academic word list for Swedish-a support for language learners in higher education. In *Proceedings of the SLTC 2012 workshop on NLP for CALL* (pp. 20–27). Lund: Linköping University Electronic Press.
- Chen, Y.-H., & Baker, P. (2010). Lexical Bundles in L1 and L2 Academic Writing. *Language Learning & Technology*, 14, 30–49.
- Chung, T. M., & Nation, I. S. P. (2004). Identifying technical vocabulary. *System*, 32(2), 251–263.  
<https://doi.org/10.1016/j.system.2003.11.008>
- CLARIN\_DK. (n.d.-a). Clarin.dk. Retrieved 5 June 2018, from <https://clarin.dk/clarindk/forside.jsp>
- CLARIN\_DK. (n.d.-b). NLP Tools - CLARIN-DK. Retrieved 5 June 2018, from <https://clarin.dk/clarindk/toolchains-wizard.jsp>
- Cobb, T., & Horst, M. (2004). Is there room for an academic word list in French? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a Second Language: Selection, Acquisition, and Testing* (Vol. 10, pp. 15–38). Amsterdam/Philadelphia: John Benjamins Publishing Company.

- Council of Europe. (n.d.). Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR). Retrieved 20 August 2018, from <https://www.coe.int/en/web/portfolio/the-common-european-framework-of-reference-for-languages-learning-teaching-assessment-cefr->
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, 213–238.
- Coxhead, A. (2011). The Academic Word List 10 years on: Research and teaching implications. *TESOL Quarterly*, 45, 355–362.
- Coxhead, A. (2016). Reflecting on Coxhead (2000), “A New Academic Word List”. *TESOL Quarterly*, 50(1), 181–185. <https://doi.org/10.1002/tesq.287>
- Coxhead, A., Yen Dang, T. N., & Mukai, S. (2017). Single and multi-word unit vocabulary in university tutorials and laboratories: Evidence from corpora and textbooks. *Journal of English for Academic Purposes*, 30, 66–78. <https://doi.org/10.1016/j.jeap.2017.11.001>
- Croy, A. B. C. (2016). *Towards establishing a pedagogical word list for young EFL learners - An empirical investigation discerning appropriate methods of vocabulary selection* (Unpublished Master thesis). University of Copenhagen.
- Cummins, J. (1980). The entry and exit fallacy in bilingual education. *NABE Journal*, 4, 25–59.
- Cummins, J. (2008). BICS and CALP: Empirical and Theoretical Status of the Distinction. In N. H. Hornberger (Ed.), *Encyclopedia of Language and Education* (pp. 487–499). Springer US. [https://doi.org/10.1007/978-0-387-30424-3\\_36](https://doi.org/10.1007/978-0-387-30424-3_36)
- Dang, T. N. Y. (2017). *Investigating vocabulary in academic spoken English: Corpora, teachers, and learners* (Unpublished PhD thesis). Victoria University of Wellington, New Zealand.
- Dang, T. N. Y. (2018a). A Hard Science Spoken Word List. *ITL-International Journal of Applied Linguistics*, 169(1), 44–71.

- Dang, T. N. Y. (2018b). The nature of vocabulary in academic speech of hard and soft-sciences. *English for Specific Purposes*, 51, 69–83.
- Dang, T. N. Y., Coxhead, A., & Webb, S. (2017). The Academic Spoken Word List. *Language Learning*, 67(4), 959–997. <https://doi.org/10.1111/lang.12253>
- Dang, T. N. Y., & Webb, S. (2014). The lexical profile of academic spoken English. *English for Specific Purposes*, 33, 66–76.
- Dang, T. N. Y., & Webb, S. (2016). Making an essential word list. In I. S. P. Nation (Ed.), *Making and using word lists for language learning and testing* (Vol. 2016). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Dang, T. N. Y., & Webb, S. (2017). Evaluating lists of high-frequency words. *ITL - International Journal of Applied Linguistics*, 167(2), 132–158. <https://doi.org/10.1075/itl.167.2.02dan>
- Dansk Journalistforbund. (n.d.). Journalisten [The Journalist]. Retrieved 5 June 2018, from <https://journalisten.dk/>
- Dansk Sprognævn. (n.d.). Moderne importord i sprogene i Norden — Dansk Sprognævn [Modern import words in the Nordic languages - The Danish Language Council]. Retrieved 10 August 2018, from <https://dsn.dk/vi-arbejder-ogsaa-med/pavirkning-fra-andre-sprog/moderne-importord-i-sprogene-i-norden>
- Davies, M. (2008). *The Corpus of Contemporary American English: 520 million words, 1990-present*. Retrieved from <https://corpus.byu.edu/coca/>
- Det Danske Sprog- og Litteraturselskab. (n.d.-a). *Den danske ordbog [The Danish Dictionary]*. Online version. Retrieved from <https://ordnet.dk/ddo>
- Det Danske Sprog- og Litteraturselskab. (n.d.-b). Inflected word forms. Retrieved 24 August 2018, from <http://korpus.dsl.dk/flexikon.html>

- Det Danske Sprog- og Litteraturselskab. (n.d.-c). KorpusDK — ordnet.dk. Retrieved 13 April 2018, from <http://ordnet.dk/korpusdk>
- Det Danske Sprog- og Litteraturselskab. (n.d.-d). Most frequently used lemmas in Danish. Retrieved 24 August 2018, from <http://korpus.dsl.dk/lemma-5000.html>
- Det Kongelige Bibliotek. (n.d.). Tidsskrift.dk [Journal.dk]. Retrieved 30 May 2018, from <https://tidsskrift.dk/>
- Durrant, P. (2009). Investigating the viability of a collocation list for students of English for academic purposes. *English for Specific Purposes*, 28(3), 157–169.
- Durrant, P. (2013). Discipline and level specificity in university students' written vocabulary. *Applied Linguistics*, 35, 328–356.
- Eldridge, J. (2008). No, There Isn't an 'Academic Vocabulary,' But...": A Reader Responds to K. Hyland and P. Tse's "Is There an 'Academic Vocabulary'?" *TESOL Quarterly*, 42, 109–113.
- Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *Tesol Quarterly*, 42(3), 375–396.
- Enström, I. (2004). Ordförråd och ordinlärning – med särskilt fokus på avancerade inlärare [Vocabulary and vocabulary learning - with particular focus on advanced learners]. In K. Hyltenstam & I. Lindberg (Eds.), *Svenska som andraspråk - i forskning, undervisning och samhälle* (pp. 171–195). Lund: Studentlitteratur.
- Francis, W. N., & Kučera, H. (1982). *Frequency Analysis of English Usage: Lexicon and Grammar*. Boston: Houghton Mifflin Company.
- Fraser, S. (2003). A statistical analysis of the vocabulary of medical research articles (3): Technical and subtechnical vocabulary. *Integrated Studies in Nursing Science*, 4(2), 27–45.



- Fraser, S. (2006). The nature and role of specialized vocabulary: What do ESP teachers and learners need to know. *Hiroshima University Scholarly Journals*, 2005, 63–75.
- Fraser, S. (2008). Beyond the Academic Word List: Providing ESP learners with the words they really need. In *Proceedings of the BAAL Annual Conference* (pp. 41–44).
- Fraser, S. (2009). Breaking down the divisions between general, academic and technical vocabulary: The establishment of a single, discipline-based word list for ESP learners. *Hiroshima Studies in Language and Language Education*, 12, 151–167.
- Gardner, D. (2007). Validating the Construct of Word in Applied Corpus-based Vocabulary Research: A Critical Survey. *Applied Linguistics*, 28(2), 241–265.  
<https://doi.org/10.1093/applin/amm010>
- Gardner, D., & Davies, M. (2014). A New Academic Vocabulary List. *Applied Linguistics*, 35(3), 305–327. <https://doi.org/10.1093/applin/amt015>
- Gellert, A. S. (2003). Ordkendskab og læseforståelse [Word knowledge and reading comprehension]. *Psykologisk Pædagogisk Rådgivning*, 40(1), 55–70.
- Ghadessy, P. (1979). Frequency counts, word lists, and materials preparation: a new approach. In *English Teaching Forum* (Vol. 17, pp. 24–27).
- Gilner, L. (2011). A primer on the General Service List. *Reading in a Foreign Language*, 23(1), 65–83.
- Gimbel, J. (1995). Bakker og udale [Hills and U-shaped valleys]. *Sprogforum - Tidsskrift for Sprog- Og Kulturpædagogik*, 3, 28–34.
- Gimbel, J. (1998). Tyrkiske børns fagrelevante danske ordforråd i femte klasse [Turkish children's discipline-specific Danish vocabulary in 5th grade]. In P. Quist, A. Holmen, & J. N. Jørgensen (Eds.), *Tosproget udvikling* (Vol. 4, pp. 95–111).

- Golden, A. (1984). Fagord og andre ord i o-fagsbøker for grunnskolen [Discipline-specific words and other words in primary school textbooks]. In A. Hvenekilde & E. Ryen (Eds.), "*Kan jeg få ordene dine, lærer?*" (pp. 170–175). Oslo: Landslaget for Norskundervisning (LNU). J. W. Cappelens Forlag.
- Golden, A. (2016). Ord i Nord. Forskning på ordforråd, ordbruk og ordlæring i et andrespråksperspektiv i Norden [Words in the North. Research in vocabulary, vocabulary use and vocabulary learning in the Nordic languages as second languages]. In *Svenskans beskrivning 35. Förhandlingar vid trettiofemte sammankomsten* (pp. 1–18). Göteborg: Institutionen för svenska språket, Göteborg Universitet. Retrieved from <https://gupea.ub.gu.se/handle/2077/52211>
- Golden, A., & Hvenekilde, A. (1983). *Rapport fra prosjektet Lærebokspråk [Report from the project Textbook language]*. Sentret for Språkpedagogikk, Universitetet i Oslo.
- Gries, S. T. (2008). Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics*, 13(4), 403–437. <https://doi.org/10.1075/ijcl.13.4.02gri>
- Gries, S. T. (2010). Useful statistics for corpus linguistics. *A Mosaic of Corpus Linguistics: Selected Approaches*, 2010(66), 269–291.
- Hagen, K., Johannessen, J. B., & Saidi, A. (2016). Constructing a Norwegian Academic Wordlist. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)* (pp. 1457–1462).
- Halliday, M. A. (1976). *System and function in language: Selected papers*. (G. Kress, Ed.). London: Oxford University Press.
- Hansen, E., & Heltoft, L. (2011). *Grammatik over det Danske Sprog [Danish Language Grammar]* (Vol. 1–3). Det Danske Sprog- og Litteraturselskab.

- Hauksdóttir, A. (2012). *Dansk som fremmedsprog i en akademisk kontekst [Danish as a foreign language in an academic context]*. (A. Holmen, Ed.) (Vol. 68). København: Københavns Universitet. Humanistisk fakultet.
- Hazenberg, S., & Hulstijn, J. H. (1996). Defining a minimal receptive second-language vocabulary for non-native university students: An empirical investigation. *Applied Linguistics*, 17(2), 145–163.
- Henriksen, B. (2014). *The distinction between general, academic and domain-specific collocations*. Presented at the PhD Applied Linguistics Annual Conference, Cardiff, Wales.
- Henriksen, B. (2015). Det oversete ordforråd? [The ignored vocabulary?]. In A. S. Gregersen (Ed.), *Sprogfag I Forandring* (2., pp. 201–225). Samfundslitteratur.
- Henriksen, B., & Danelund, L. (2015). Studies of Danish L2 learners' vocabulary knowledge and the lexical richness of their written production in English. In P. Pietilä, K. Doró, & R. Pípalová (Eds.), *Lexical issues in L2 writing* (pp. 29–56). Newcastle upon Tyne: Cambridge Scholars Publishing.
- Henriksen, B., & Westbrook, P. (2017). Responding to Research Challenges Related to Studying L2 Collocational Use in Professional Academic Discourse. *Vocabulary Learning and Instruction*, 6(1), 32–47.
- Hirsh, D. (2004). *A functional representation of academic vocabulary* (Unpublished PhD thesis). Victoria University of Wellington, New Zealand.
- Hirsh, D. (2010). *Academic vocabulary in context* (Vol. 118). Bern: Peter Lang.
- Hirsh, D., & Nation, I. S. P. (1992). What Vocabulary Size Is Needed to Read Unsimplified Texts for Pleasure? *Reading in a Foreign Language*, 8(2), 689–696.

- Holmen, A. (2016). Dansk som akademisk sprog for nordiske studerende [Danish as an academic language for Nordic students]. *Sprog i Norden. Tema: Forståelse Og Kommunikationsstrategier*, 37–48.
- Holsting, A., Svendsen, J. T., Lindø, A. V., Iversen, D. L., Blom, J. N., Rathje, M., ... le Fevre Jakobsen, B. (2017). En undersøgelse af henvisnings-former i nyindskrevne universitets-studerendes akademiske opgaver [An investigation of how first-year university students make references in their academic papers]. *RASK - International Journal of Language and Communication*, 46 (Autumn), 21–67.
- Hornby, A. S. (2010). *Oxford Advanced Learner's Dictionary*. (J. Turnbull, D. Lea, & D. Parkinson, Eds.) (8th ed.). Oxford: Oxford University Press.
- Hultgren, A. K. (2013). *Parallelsproglighed på danske universiteter: en statusrapport 2013*. [Parallel lingualism at Danish Universities: a status report 2013] (Vol. C5). Københavns Universitet. Humanistisk fakultet.
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.  
<https://doi.org/10.1017/CBO9781139524773>
- Hyland, K. (1999). Disciplinary discourses: Writer stance in research articles. In C. N. Candlin & K. Hyland (Eds.), *Writing: Texts, processes and practices* (Vol. 2014, pp. 99–121). London & New York: Routledge.
- Hyland, K. (2004). *Disciplinary discourses. Social interactions in academic writing*. Ann Arbor: University of Michigan Press.
- Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, 7(2), 173–192.

- Hyland, K. (2008). Academic clusters: text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18(1), 41–62. <https://doi.org/10.1111/j.1473-4192.2008.00178.x>
- Hyland, K., & Tse, P. (2007). Is There an ‘Academic Vocabulary’? *TESOL Quarterly*, 41(2), 235–253. <https://doi.org/10.2307/40264352>
- Institute of Formal and Applied Linguistics. (n.d.). CoNLL-2009 ST Task Description. Retrieved 5 June 2018, from <http://ufal.mff.cuni.cz/conll2009-st/task-description.html>
- Jakobsen, A. S. (2017). From Implicit Norms to Explicit Skills – Focusing on Danish Academic Vocabulary. *OSLA Oslo Studies in Language: Academic Language in a Nordic Setting - Linguistic and Educational Perspectives*, 9(3), 59–76.
- Jakobsen, A. S., Coxhead, A., & Henriksen, B. (2018). General and academic high frequency vocabulary in Danish. *Nordand – Nordisk Tidsskrift for Andrespråksforskning*, 01/2018, 64–89. <https://doi.org/10.18261>
- Jakubíček, M., Kilgariff, A., Kovář, V., Rychlý, P., & Suchomel, V. (2013). The tenten corpus family. In *7th International Corpus Linguistics Conference CL* (pp. 125–127).
- Jansen, M. (1973). *De almindeligste ord i begynderundervisningen i dansk: En oversigt [The most common words in early primary school teaching: An overview]* (Vol. 76). Munksgaard.
- Jansson, H., Kokkinakis, S. J., Ribeck, J., & Sköldberg, E. (2012). A Swedish Academic Word List: Methods and Data. In *Proceedings of the 15th EURALEX International Congress* (pp. 7–11). Oslo: University of Oslo.
- Järborg, J. (2007). Om ord och ordkunskap [On words and vocabulary knowledge]. In I. Lindberg & S. Johansson Kokkinakis (Eds.), *OrdiL – en korpusbaserad kartläggning av ordförrådet i läromedel för grundskolans senare år* (pp. 61–100). Göteborg: Institutet för svenska språket, Göteborg Universitet. Retrieved from <https://gupea.ub.gu.se/handle/2077/20503>

- Johansson, S. (2017). Bedömning av en- och flerspråkiga elevers ordförståelse i naturvetenskapliga ämnen [Assessment of mono- and multilingual students' vocabulary knowledge in natural science subjects]. Presented at the NORDAND: Nordens språk som andraspråk 13, Vasa, Finland.
- Johansson, S., Hagen, K., & Johannessen, J. B. (2017). A bilingual academic word list: The merging of a norwegian and a swedish list. *OSLA Oslo Studies in Language: Academic Language in a Nordic Setting - Linguistic and Educational Perspectives*, 9(3), 147–168.
- Jørgensen, J. N. (1984). Fremmedarbejdersbørns danske ordforråd. En sociolingvistisk undersøgelse. [The Danish vocabulary of immigrant children. A socio-linguistic investigation]. In R. Ringgaard & V. Sørensen (Eds.), *The Nordic Languages and Modern Linguistics*. Århus: Nordisk Institut, Aarhus Universitet.
- Juilland, A., & Chang-Rodríguez, E. (1964). *Frequency dictionary of Spanish words*. London: Mouton.
- Kanebrant, E., Mühlenbock, K. H., Kokkinakis, S. J., Jönsson, A., Liberg, C., Af Geijerstam, \AAsa, ... Falkenjack, J. (2015). T-MASTER-A Tool for Assessing Students' Reading Abilities. In *Proceedings of the 7th International Conference on Computer Supported Education CSEDU (1)* (pp. 220–227). Lisbon, Portugal.  
<https://doi.org/10.5220/0005410902200227>
- Kennedy, G. (2003). Amplifier collocations in the British National Corpus: Implications for English language teaching. *Tesol Quarterly*, 37(3), 467–487.
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., ... Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicographica*, 1(1), 7–36.
- Kilgarriff, A., Rychlý, P., Smrž, P., & Tugwell, D. (2004). *The Sketch Engine* (Information Technology Research Institute Technical Report Series No. ITRI-04-08). University of

Brighton. Retrieved from

[https://www.researchgate.net/profile/Adam\\_Kilgarriff/publication/260387608\\_ITRI-04-](https://www.researchgate.net/profile/Adam_Kilgarriff/publication/260387608_ITRI-04-08_the_sketch_engine/links/54e0d1210cf24d184b0de48f/ITRI-04-08-the-sketch-engine.pdf)

[08\\_the\\_sketch\\_engine/links/54e0d1210cf24d184b0de48f/ITRI-04-08-the-sketch-engine.pdf](https://www.researchgate.net/profile/Adam_Kilgarriff/publication/260387608_ITRI-04-08_the_sketch_engine/links/54e0d1210cf24d184b0de48f/ITRI-04-08-the-sketch-engine.pdf)

Kjærulff Nielsen, B. (1988). *Engelsk-Dansk Ordbog [English-Danish Dictionary]* (6th ed.).

København: Gyldendal.

Knudsen, S. (2009). Har du et problem? En undersøgelse af universitetsstuderendes forståelse, anvendelse og kommunikation af problemorienterede vidensproblemer [Do you have a problem? A study on university students' comprehension, use, and communication of problem-oriented knowledge problems]. *Dansk Universitetspædagogisk Tidsskrift*, 4(7), 49–57.

Københavns Universitet. (2008, July 18). KULEX ordbog [KULEX dictionary]. Retrieved 2 July 2018, from <https://ordbog.ku.dk/>

Kristiansen, B. (2010). Tekstproduktion og vidensproduktion [Text productions and knowledge production]. *Dansk Universitetspædagogisk Tidsskrift*, 5, 50–54.

Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? In C. Laurén & M. Nordman (Eds.), *Special Language: From Humans Thinking to Thinking Machines* (pp. 316–323). Clevedon: Multilingual Matters.

Laufer, B., & Nation, I. S. P. (1995a). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16, 307–322.

Laufer, B., & Nation, P. (1995b). Vocabulary Size and Use: Lexical Richness in L2 Written Production. *Applied Linguistics*, 16(3), 307–322. <https://doi.org/10.1093/applin/16.3.307>

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, 22(1), 15–30.

- Laursen, H. P. (2006). Den sproglige dimension i naturfagsundervisningen [The linguistic dimension in the teaching of natural science subjects]. *MONA*, 2, 27–46.
- Laursen, K. Å. (2013). *'Det er sprogligt-selv hvor du ikke lægger mærke til det': En empirisk undersøgelse af de sproglige og faglige vanskeligheder hos farmaceutstuderende med dansk som andetsprog på Københavns Universitet* ["It's all about language even when it's not": An empirical study on the linguistic and study-related difficulties in pharmacist students with Danish as a second language at the University of Copenhagen] (Vol. C4). Københavns Universitet. Humanistisk fakultet.
- Leech, G., & Rayson, P. (2014). *Word frequencies in written and spoken English: Based on the British National Corpus*. Routledge.
- Lindberg, I. (2007). Forskning om läromedelsspråk och ordförrådsutveckling [Research on textbook language and vocabulary development]. In I. Lindberg & S. J. Kokkinakis (Eds.), *OrdiL - En korpusbaserad kartläggning av ordförrådet i läromedel för grundskolans senare år* (pp. 13–60). Göteborg: Institutet för svenska språket, Göteborg Universitet.
- Lindberg, I., & Johansson Kokkinakis, S. (2007). *OrdiL – en korpusbaserad kartläggning av ordförrådet i läromedel för grundskolans senare år* [OrdiL - a corpus-based survey of the vocabulary of textbooks used in lower secondary school] (Rapporter om svenska som andraspråk (ROSA) No. 8). Institutet för svenska som andraspråk, Göteborg Universitet. Retrieved from <https://gupea.ub.gu.se/handle/2077/20503>
- Lund, K. (2016). Komplexiteten i akademisk sprogbrug belyst ved modelanalyser [The complexity of academic language use illustrated by model analyses]. *Sprogforum - Tidsskrift for Sprog- Og Kulturpædagogik*, 2016(63), 80–90.
- Lund, K., & Bertelsen, E. (2008a). At mestre et uddannelsesfagligt fremmed/andetsprog – en forudsætning på en videregående uddannelse [Mastering of an educational second/foreign



- language - a requirement of higher education]. *Sprogforum - Tidsskrift for Sprog- Og Kulturpædagogik*, (44), 17–24.
- Lund, K., & Bertelsen, E. (2008b). *Fra Studieprøven til de videregående uddannelser: En undersøgelse af de nødvendige og tilstrækkelige kompetencer [From Studieprøven to higher education: A study on the necessary and adequate skills]*. Ministeriet for Flygtninge, Indvandrere og Integration.
- Lundskær-Nielsen, T., & Holmes, P. (2010). *Danish: A comprehensive grammar* (2nd ed.). London & New York: Routledge.
- Lynn, R. W. (1973). Preparing word-lists: a suggested method. *RELJ Journal*, 4(1), 25–28.
- Malmström, H., Pecorari, D. E., & Gustafsson, M. (2016). Coverage and development of academic vocabulary in assessment texts in English medium instruction. In S. Göpferich & I. Neumann (Eds.), *Developing and Assessing Academic and Professional Writing Skills*. (pp. 45–69). New York: Peter Lang.
- Malmström, H., Pecorari, D., & Shaw, P. (2018). Words for what? Contrasting university students' receptive and productive academic vocabulary needs. *English for Specific Purposes*, 50, 28–39.
- Martin, A. V. (1976). Teaching academic vocabulary to foreign graduate students. *Tesol Quarterly*, 91–97.
- Matthews, J., & Cheng, J. (2015). Recognition of high frequency words from speech as a predictor of L2 listening comprehension. *System*, 52, 1–13.
- McEnery, T., & Hardie, A. (2011). *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.
- McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. London & New York: Taylor & Francis.

- McQuillan, J. (2016). What can readers read after graded readers? *Reading in a Foreign Language*, 28(1), 63–78.
- Meyer, P. G. (1990). Non-technical vocabulary in technical language. Presented at the AILA, Thessaloniki.
- Miller, D., & Biber, D. (2015). Evaluating reliability in quantitative vocabulary studies: The influence of corpus design and composition. *International Journal of Corpus Linguistics*, 20(1), 30–53. <https://doi.org/10.1075/ijcl.20.1.02mil>
- Møller, P. H. (2015). *Sproglige og studiemæssige udfordringer hos studerende med dansk som andetsprog-pilotprojekt [Linguistic and educational challenges of students with Danish as their second language - a pilot study]*. Aarhus: Aarhus Universitet.
- Moon, R. (1997). Vocabulary connections: Multi-word items in English. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 40–63). Cambridge: Cambridge University Press.
- Morris, S. (2010). Core Vocabulary for Adult Learners of Welsh- A celtic template? *Journal of Celtic Language Learning*, 111–127.
- Morris, S., & Meara, P. (2014). *Welsh Words. Core vocabulary with phrases*. Ceredigion: Y Lolfa.
- Nagy, W., & Townsend, D. (2012). Words as Tools: Learning Academic Vocabulary as Language Acquisition. *Reading Research Quarterly*, 47(1), 91–108. <https://doi.org/10.1002/RRQ.011>
- Nation, I. S. P. (1983). Testing and teaching vocabulary. *Guidelines*, 5(1), 12–25.
- Nation, I. S. P. (1990). *Teaching and Learning Vocabulary*. Boston, MA: Heinle & Heinle.
- Nation, I. S. P. (2004). A study of the most frequent word families in the British National Corpus. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (Vol. 10, pp. 3–13). Amsterdam/Philadelphia: John Benjamins Publishing Company.

- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63, 59–82.
- Nation, I. S. P. (2012). The BNC/COCA word family lists. Retrieved 24 August 2018, from [https://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Information-on-the-BNC\\_COCA-word-family-lists.pdf](https://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Information-on-the-BNC_COCA-word-family-lists.pdf)
- Nation, I. S. P. (2013). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nation, I. S. P. (2016). *Making and using word lists for language learning and testing*. John Benjamins Publishing Company.
- Nation, I. S. P., & Coxhead, A. (2014). Vocabulary size research at Victoria University of Wellington, New Zealand. *Language Teaching*, 47(3), 398–403.  
<https://doi.org/10.1017/S0261444814000111>
- Nation, I. S. P., Coxhead, A., Chung, T. M., & Quero, B. (2016). Specialized word lists. In I. S. P. Nation (Ed.), *Making and Using Word Lists for Language Learning and Testing* (pp. 145–151). Amsterdam: John Benjamins Publishing Company.
- Nation, I. S. P., Heatly, A., & Coxhead, A. (2002). *Range: A program for the analysis of vocabulary in texts*. Retrieved from <https://www.victoria.ac.nz/lals/about/staff/paul-nation#links>
- Nation, I. S. P., & Parent, K. (2016). Homoforms and polysemes. In I. S. P. Nation (Ed.), *Making and Using Word Lists for Language Learning and Testing* (pp. 41–53). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Nation, I. S. P., & Sorell, J. (2016). Corpus selection and design. In I. S. P. Nation (Ed.), *Making and Using Word Lists for Language Learning and Testing* (pp. 95–105). Amsterdam/Philadelphia: John Benjamins Publishing Company.

- Nation, I. S. P., & Webb, S. A. (2011). *Researching and analyzing vocabulary*. Heinle, Cengage Learning.
- Nesi, H. (2002). An English spoken academic word list. In *Proceedings of the Tenth EURALEX International Congress* (Vol. 1, pp. 351–358). EURALEX Copenhagen, Denmark.
- Nguyen, T. M. H., & Webb, S. (2017). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research*, 21(3), 298–320.
- Nielsen, A.-M. V., Daugaard, H. T., & Juul, H. (2017). Ordkendskabsundervisning på mellemtrinnet [Vocabulary knowledge teaching in middle school]. *Skandinaviske Sprogstudier*, 8(1).
- Odgaard, S. M. (2014). ‘*Det er ikke et sprog, man bare kan samle op på gaden*’: en undersøgelse af behovet for kurser i akademisk dansk for studerende med dansk som andetsprog på Institut for Tværkulturelle og regionale Studier på Københavns Universitet [*It's not a language that you can just pick up on the street": an investigation of the need for courses in academic Danish for students with Danish as a second language at the Department of Cross-Cultural and Regional Studies at the University of Copenhagen*] (Vol. C6). Københavns Universitet. Humanistisk fakultet.
- Organisation for Economic Co-operation and Development. (2007). *OECD Revised Field of Science and Technology (FOS) Classification in the Frascati Manual* (No. DSTI/EAS/STP/NESTI(2006)19/FINAL). Organisation for Economic Co-operation and Development. Retrieved from <http://www.oecd.org/sti/inno/38235147.pdf>
- Paquot, M. (2010). *Academic vocabulary in learner writing: From extraction to analysis*. Bloomsbury Publishing.

- Pinchbeck, G. G. (2014). Lexical frequency profiling of a large sample of Canadian high school diploma exam expository writing: L1 and L2 academic English. *Roundtable Presentation at American Association of Applied Linguistics, Portland, OR, USA*.
- Praninskas, J. (1972). *American university word list*. London: Longman Group Limited.
- Quero, B., & Coxhead, A. (2018). Using a Corpus-Based Approach to Select Medical Vocabulary for an ESP Course: The Case for High-Frequency Vocabulary. In Y. Kırkgöz & K. Dikilitaş (Eds.), *Key Issues in English for Specific Purposes in Higher Education* (pp. 51–75). Springer, Cham. [https://doi.org/10.1007/978-3-319-70214-8\\_4](https://doi.org/10.1007/978-3-319-70214-8_4)
- Ranney, S. (2012). Defining and teaching academic language: Developments in K-12 ESL. *Language and Linguistics Compass*, 6(9), 560–574.
- Revision of the Frascati Manual - OECD. (n.d.). Retrieved 30 May 2018, from <http://www.oecd.org/sti/inno/frascati-manual-revision.htm>
- Ribeck, J., Jansson, H., & Sköldbberg, E. (2014). Från aspekt till övergripande: en ordlista över svensk akademisk vokabulär [From aspect to overall: a word list of Swedish academic vocabulary]. In *DIVA* (Vol. 13, pp. 370–384). Nordisk forening for leksikografi.
- Richards, J. C. (1974). Word lists: problems and prospects. *RELC Journal*, 5(2), 69–84.
- Rienecker, L., & Jørgensen, P. S. (2012). *Den gode opgave [The good paper]* (4. udgave). Samfundslitteratur.
- Ruus, H. (1992). Anmeldelse af Henning Bergholtz: Frekvensordbog, Århus 1989, og Henning Bergholtz: Dansk Frekvensordbog, G.E.C.Gads Forlag 1992 [Review of Henning Bergholtz: Frequency dictionary, Århus 1989, and Henning Bergholtz: Danish frequency dictionary, G.E.C.Gads Forlag 1992]. *Danske Studier*, 1992, 154–158.
- Ruus, H. (1995). *Danske kerneord: centrale dele af den danske leksikalske norm [Danish core words: central elements of the Danish lexical norm]*. København: Museum Tusculanum.

- Savický, P., & Hlaváčová, J. (2002). Measures of Word Commonness. *Journal of Quantitative Linguistics*, 9(3), 215–231. <https://doi.org/10.1076/jqul.9.3.215.14124>
- Schmitt, D., & Schmitt, N. (2005). *Focus on vocabulary: Mastering the Academic Word List*. Pearson Education.
- Schmitt, D., Schmitt, N., & Mann, D. (2011). *Focus on vocabulary 1: Bridging vocabulary*. Longman.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Springer.
- Schmitt, N., Jiang, X., & Grabe, W. (2011). The Percentage of Words Known in a Text and Reading Comprehension. *The Modern Language Journal*, 95(1), 26–43.
- Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, 47(4), 484–503. <https://doi.org/10.1017/S0261444812000018>
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18(1), 55–88.
- Scott, M. (1997). PC analysis of key words—and key key words. *System*, 25(2), 233–245.
- Scott, M., & Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education* (Vol. 22). John Benjamins Publishing.
- Shaw, P., & Vassileva, I. (2009). Co-evolving academic rhetoric across culture; Britain, Bulgaria, Denmark, Germany in the 20th century. *Journal of Pragmatics*, 41, 290–305.
- Simonsen, I. (2015). Handler” teksten” om noget” eller” undersøger” den” hvordan”?: En undersøgelse af tyske L2-sprogbrugeres beherskelse af” formelagtigt sprog [Is ‘the text’ about something or does it ‘investigate’ ‘how’?: A study on German L2 users’ mastering of formulaic language]. In 15. *Møde om Udforskningen af Dansk Sprog (MUDS 15)* (pp. 345–363). Aarhus Universitet.

- Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31, 487–512.
- Sköldberg, E., & Johansson Kokkinakis, S. (2012). A och O om akademiska ord-Om framtagning av en svensk akademisk ordlista [A and O on academic words - on the extraction of a Swedish academic word list]. *Nordiske Studier i Leksikografi*.
- Skov, S. (2006). Hvordan bedømmer man det sproglige i universitetsopgaver? [How to assess the language of university papers?]. *Dansk Universitetspædagogisk Tidsskrift*, 1, 18–25.
- Skov, S. (2013). Progression i de studerendes skrivekompetence [Progression in the writings skills of students]. *Spor - et Tidsskrift for Universitetspædagogik*, (1), 1–8.
- Snow, C. E. (2010). Academic language and the challenge of reading for learning about science. *Science*, 328, 450–452.
- Snow, C. E., & Uccelli, P. (2009). The challenge of academic language. *The Cambridge Handbook of Literacy*, 112–133.
- Stray Jørgensen, P. (2004). Videnskabelig ord – sproglig rådgivning i videnskabelighed [Scientific words – language guidance in scientific character]. In H. Jørgensen & P. Stray Jørgensen (Eds.), *På godt dansk. Festskrift til Henrik Galberg Jacobsen i anledning af hans 60 års fødselsdag 4. februar 2004* (pp. 181–190). København; Århus: Wessel og Huitfeldt.
- Stray Jørgensen, P. (2007). Teksten i vejledning - modeller for vejledning i videnskabelig opgaveskrivning [The role of the text in guidance - models of guidance in academic writing]. In *Skrive for nåtid og fremtid* (pp. 158–166). Trondheim: Tapir Akademisk Forlag.
- SwePub. (n.d.). Retrieved 25 June 2018, from <http://swepub.kb.se/>
- Treffers-Daller, J., Parslow, P., & Williams, S. (2018). Back to Basics: How Measures of Lexical Diversity Can Help Discriminate between CEFR Levels. *Applied Linguistics*, 39(3), 302–327. <https://doi.org/10.1093/applin/amw009>

- Uddannelses- og Forskningsministeriet. (n.d.). Den Danske Forskningsdatabase [Danish National Research Database]. Retrieved 29 May 2018, from <http://www.forskningsdatabasen.dk/da>
- Ulriksen, L., Murning, S., & Ebbensgaard, A. B. (2009). *Når gymnasiet er en fremmed verden: Eleverfaringer-social baggrund-fagligt udbytte [When upper secondary school is a unfamiliar world: student experiences, social background, learning]*. Samfundslitteratur.
- Universitetet i Oslo. (n.d.). NBAO - Norsk Bokmål Akademisk Ordliste. Retrieved 20 August 2018, from <http://www.tekstlab.uio.no:4000/about>
- van Zeeland, H., & Schmitt, N. (2013). Lexical Coverage in L1 and L2 Listening Comprehension: The Same or Different from Reading Comprehension? *Applied Linguistics*, 34(4), 457–479. <https://doi.org/10.1093/applin/ams074>
- Wang, K. M.-T., & Nation, P. (2004). Word meaning in academic English: Homography in the academic word list. *Applied Linguistics*, 25(3), 291–314.
- Webb, S. A., & Chang, A. C.-S. (2012). Second language vocabulary growth. *RELC Journal*, 43(1), 113–126.
- Webb, S., & Macalister, J. (2013). Is Text Written for Children Useful for L2 Extensive Reading? *TESOL Quarterly*, 47(2), 300–322.
- Weber, E. (2009). At skelne det før-faglige ordforråd [Identifying pre-technical vocabulary]. *Sprogforum - Tidsskrift for Sprog- Og Kulturpædagogik*, (47), 52–58.
- Wenger, E. (1998). Communities of practice: Learning as a social system. *Systems Thinker*, 9(5), 2–3.
- West, M. (1953). *A General Service List of English Words* (Vol. 1953). London: Longman, Green and Co.



- Wray, A. (2004). 'Here's one I prepared earlier': Formulaic language learning on television. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use*. Amsterdam: John Benjamins Publishing Company.
- Wray, A. (2008). *Formulaic language: Pushing the boundaries*. Oxford: Oxford University Press.
- Xu, J. (2009). Log-likelihood ratio calculator. Retrieved 6 April 2018, from <http://www.bfsu-corpus.org/static/BFSUTools/BFSULL.xls>
- Xue, G., & Nation, I. S. P. (1984). A university word list. *Language Learning and Communication*, 3(2), 215–229.

## Abstract

Academic vocabulary lists developed for English, Norwegian and Swedish have advanced our understanding of the lexical inventories in these languages. This thesis presents four studies aimed at identifying and describing Danish academic vocabulary. To accomplish this aim, a corpus of 3 million words of written academic Danish, the AcaDan Corpus, was compiled together with two smaller corpora of written academic and general Danish.

Drawing on these corpora, Study 1 investigated the lexical coverage of Danish general high frequency vocabulary defined as the 2,000 most frequent words of Danish in general and academic language. The results showed that these 2,000 words cover 76 percent of general language texts and 63 percent of academic language texts. Moreover, the first 1,000 words have substantially higher coverage than the second 1,000 words. Study 1 also explored the potential overlap between general and academic vocabulary by using criteria for word list development. In this way, 402 general high frequency words were identified as academic.

In Study 2, Danish academic words were identified using quantitative criteria of range, frequency, and dispersion resulting in a Danish Academic Word List (the DAWL). This list contains 770 academic lemmas and provides a coverage of 26-28 percent in academic language represented by two academic language corpora. In contrast, its coverage over a general language corpus was 19 percent, which confirms the academic nature of the DAWL words. Study 2, in line with the findings from Study 1, highlighted that academic and general vocabulary overlap in that 363 of the DAWL words belong to Danish general high frequency vocabulary.

Study 3 analysed the words of the DAWL according to the functions these words perform in the academic text thus providing a functional categorisation of Danish academic vocabulary. In Study 4, the DAWL was expanded by investigating the words excluded from the DAWL because they only satisfied the first two criteria established for the DAWL in Study 2. Words morphologically and semantically related to DAWL words were added and thus a new list, the S-DAWL emerged.

The findings of these four studies advance our understanding of academic vocabulary, in general, and of Danish academic vocabulary, in particular. Moreover, this research has pedagogical implications for the teaching of academic Danish to both L1 and L2 students and learners as it offers a basis for the development of pedagogical tools and teaching material with an explicit focus on vocabulary.

## Abstract in Danish

Akademiske ordlister udviklet for engelsk, svensk og norsk har bidraget betydeligt til forståelsen af akademisk ordforråd i disse sprog. Denne afhandling præsenterer fire undersøgelser hvis overordnede mål det har været at identificere og beskrive dansk akademisk ordforråd. Til dette formål blev der indsamlet og dannet et skriftligt akademisk korpus på 3 millioner ord, AkaDan-korpusset, samt to mindre korpusser indeholdende akademisk og alment skriftligt dansk.

Den første undersøgelse indkredsede hvordan det almene højfrekvente ordforråd, defineret som de 2.000 hyppigste ord i dansk, dækker alment og akademisk sprogbrug. Resultaterne viste at disse 2.000 ord dækker 76 procent af ordforrådet i almene tekster og 63 procent af ordforrådet i akademiske tekster. Derudover har de første 1.000 hyppigste ord en væsentlig højere dækningsgrad end de næste 1.000 hyppigste ord. Overlappet mellem alment og akademisk ordforråd blev endvidere undersøgt ud fra kriterier der anvendes inden for udvikling af ordlister. Ved at anvende denne metode blev 402 almene højfrekvente ord identificeret som akademiske ord.

I den anden undersøgelse blev dansk akademisk ordforråd identificeret ved at ordene i AkaDan-korpusset blev målt ud fra deres rækkevidde, frekvens og spredning. På baggrund heraf blev en dansk akademisk ordliste, DAO, etableret. Denne liste indeholder 770 akademiske lemmaer og dækker 26-28 procent af akademisk sprog og 19 procent af ordforrådet i alment sprog hvilket dokumenterer at den repræsenterer dansk akademisk ordforråd. Ligesom den første undersøgelse bekræfter den anden at akademisk og alment ordforråd overlapper i og med at 363 af ordene i DAO også er at finde blandt de 2.000 mest hyppige ord i dansk.

I afhandlingens tredje undersøgelse blev ordene i DAO analyseret ift. deres funktioner i akademiske tekster, og der blev ud fra denne analyse etableret en funktionel kategorisering af dansk akademisk ordforråd. I den fjerde undersøgelse blev DAO udvidet til en ny ordliste S-DAO, der indeholder 1.322 akademiske lemmaer. Listen blev udviklet ved at analysere de ord, som ikke opfyldte kriteriet for spredning i undersøgelse 2, i forhold til morfologiske og semantiske relationer til ordene i DAO. De ord, der ud fra denne analyse havde morfologiske og semantiske relationer DAO blev således tilføjet.

Resultaterne af de fire undersøgelser bidrager til vores forståelse af akademisk ordforråd generelt og af dansk akademisk ordforråd i særdeleshed og har betydning for undervisningen og tilegnelsen af akademisk dansk for både L1 og L2 studerende og lærende idet resultaterne kan anvendes til udvikling af pædagogiske redskaber og undervisningsmaterialer med et eksplicit fokus på ordforråd.

## Appendices

Appendix A. Texts in the AcaDan Corpus arranged according to academic discipline .....	274
Appendix B. Texts in the Second Academic Corpus arranged according to academic discipline ..	298
Appendix C. Texts in the General Language Corpus arranged according to type.....	300
Appendix D. Co-author Declaration and Confirmation.....	316
Appendix E. 402 academic lemmas .....	319
Appendix F. The DAWL .....	329
Appendix G. The S-DAWL .....	347

## Appendix A. Texts in the AcaDan Corpus arranged according to academic discipline

### *Health Science*

Agergaard, S. (2004). Idræt som folkelig dannelse og idrætsfaglig uddannelse. Idrottsforum.org.

Agerholm, A., Bukholt, S. E., Golubovic, S., Hansson, L., Ibsen, J. R., Niemann, H. L., ... & Vestergaard, S. (2015). National klinisk retningslinje for udvalgte sundhedsfaglige indsatser ved rehabilitering til patienter med type 2 diabetes.

Andkjær, S. (2004). Pædagogiske perspektiver på udviklingen i dansk friluftsliv. Focus - Tidsskrift for Idræt, 3, 17–23.

Arvin, E., Bardow, A., Bruvo, M., Rygaard, M., & Spliid, H. (2010). Caries hos børn og unge påvirkes af drikkevandskvaliteten. Dansk vand, 78(6), 16–19.

Avlund, K., Bruunsgaard, H., Christensen, U., Fiehn, N.-E., Hansen, Å. M., Holm-Pedersen, P., ... Lund, R. (2009). CAMB - Copenhagen Aging and Midlife Biobank. Perspektiver for fremtidig forskning. Miljø Og Sundhed, 15(suppl. nr. 1), 81–88.

Belstrøm, D. (2016). Mikrobiologisk diagnostik ved infektioner i mundhulen nu og i fremtiden. Tandlægebladet, 120(5), 418–425.

Biong, S., & Thylstrup, B. (2016). Verden vælter: Pårørendes erfaringer med narkotikarelaterede dødsfald-The world collapses: Relatives' experiences with drug related deaths. Klinisk Sygepleje, 30(02), 75–86.

Bjørndal, L., Bruun, G., Pedersen, S. D. N., & Langemark, C. (2014). Den biomekaniske udrensning -rationale, effekt og kliniske principper. Tandlægebladet, 118(7), 528–38.

Blom, J., Finne, M. K., Reibel, J., & Kragelund, C. (2014). Keratocyster: Overensstemmelse mellem tentativ klinisk/radiologisk diagnose og histopatologisk diagnose. Tandlægebladet, 118(8), 618–623.

Bruun, G., & Bjørndal, L. (2013). Behandling af profund caries-baseret på evidens fra nyere kliniske undersøgelser. Tandlægebladet, 117, 322–9.

Budolfsen, K. B., Andersen, K. S., Bakke, M., & Jensen, R. H. (2012). Klyngehovedpine (Hortons hovedpine) – karakteristik og fællestræk med tandsmerter. Tandlægebladet, 116(8), 592–598.

Buus, N. (2011). Forbedring af adherence til medicinsk depressionsbehandling med udgangspunkt i brugerperspektivet.

Bygum, A., & Rye Rasmussen, E. (2013). Arveligt angioødem i odontologien. Tandlægebladet, 117, 648–650.

### *Health Science*

Carlsen, A., & Hillerup, S. (2010). Mislykket kæberekonstruktion med samtidig implantatindsættelse. Tandlægebladet, 114(14), 1122–1125.

Clausen, T., Bern, S. H., Carnelro, I. G., Sejbæk, C. S., Borg, V., & Aust, B. (2012). Forskelle og ligheder i oplevelsen af arbejdsmiljøet i ældreplejen i ni kommuner fra 2005-2008.

- Dabelsteen, E. (2010). Hvad skal vi se efter i mundhulen - og hvad kan vi gøre? *Ugeskrift for Læger*, 172(44), 3016–3018.
- Dalhoff, K. P. (2012). Interaktioner mellem mad og lægemidler. *Diætisten*, 119(5-8).
- Dalsten, H., & Kreiborg, S. (2014). Sammenhængen mellem astma og carieserfaring. *Tandlægebladet*, 118(8), 632–7.
- Diderichsen, F., Nygaard, E., & Bonde, A. (2009). Samarbejde mellem forskning og praksis på forebyggelsesområdet.
- Eichberg, H. (2006). Idrætsopfattelsen, et foranderligt fænomen. *Idrottsforum.org*.
- Eliassen, M., Skov-Ettrup, L., Christiansen, A. H., Pedersen, M. G., Mikkelsen, S. S., Grønbæk, M., ... & Becker, U. (2012). Alkohol, rygning og postoperative komplikationer.
- Engel, L. (2004). På sporet af bevægelse - imagination og kreative processer. *Idrottsforum.org*.
- Folmann, N., & Jørgensen, T. (2006). Etniske minoriteter-sygdom og brug af sundhedsvæsenet: Et registerstudie.
- Friis-Hasché, E. (2010). Patienters beskrivelse af akutte orale smerter i relation til angst. *Tandlægebladet*, 114(3), 216–23.
- Friis-Hasché, E., & Øzhayat, E. B. (2014). Psykologiske observationer og diagnoser i odontologisk klinik. *Tandlægebladet*, 119(2), 112–19.
- Gotfredsen, K. (2011). Protetiske aspekter indenfor traumatologien. *Tandlægebladet*, 115(15), 1118–1125.
- Greisen, G. (2008). De mindste børn i pædiatrien. *Bibl Læger*, 200, 516-26.
- Haj Brade, L., Lundell, E., & Mohr, S. (2013). We are here, we are queer and we are so used to it. *Kvinder, Køn and Forskning*, 22(1), 3–9.
- Hansen, J. (2008). Fra OL i Moskva til OL i Beijing. *Idrottsforum.org*.
- Heegaard, S., & Villumsen, J. (2009). Isbjørn - *Ursus maritimus*. *Oftalmolog*, 4, 3–5.
- Helweg-Larsen, K., & Larsen, H. B. (2006). Overgreb på børn: hvorfor sker det og hvordan kan det undgås? Kbh.: Statens Institut for Folkesundhed, SDU.
- Helweg-Larsen, K., Kastrup, M., Baez, A., & Flachs, E. M. (2007). Etniske Forskelle i kontaktmønsteret til psykiatrisk behandling. Et registerbaseret studie. København: Videnscenter for Transkulturel Psykiatri.
- Hess, P., & Kreiborg, S. (2014). Forebyggelse af uønskede tandvandring efter tidligt mælketandstab. *Tandlægebladet*, 118(11), 902–9.
- Hjelmar, U., Bhatti, Y., Rostgaard, T., & Petersen, O. H. (2016). Kvalitet på offentlige og private plejecentre i Danmark. Forskningsprojektet "Dokumentation af effekter ved konkurrenceudsættelse af offentlige opgaver". Del-rapport, 6.
- Hvidtfeldt, U. A., Vinther-Larsen, M., Petersen, C. B., Thygesen, L. C., & Grønbæk, M. (2006). Ældre og alkohol: Sammenhæng mellem socioøkonomiske faktorer og alkoholstorforbrug.

- Håkansson, K., Aanæs, K., Buchwald, C. von, & Johansen, H. K. (2012). Inflammatoriske og infektiøse tilstande i næse-bihuler. *Tandlægebladet*, (7), 506–509.
- Ibsen, K. K. (2008). Syge børn og børnelæger gennem 100 år. *Bibliotek for Læger* nr. 4, 2008. 387-426.
- Iversen, P. (2011). PSA-baseret screening for prostatacancer. *Klinisk Biokemi i Norden*, Nr. 3, vol. 23, 10-13
- Jacobsen, N., & Dahl, J. E. (2013). Osteoporose og marginal parodontitis. *Tandlægebladet*, 117(12). 994-1000.
- Jakobsen, A. S., Laursen, L. C., Schou, L. H., Emme, C., & Phanareth, K. (2012). Vekslende effekt af telemedicin ved behandling af kronisk obstruktiv lungesygdom. *Ugeskrift for Læger*, 174(14), 936–42.
- Jakobsen, C., Skou, R., Hertz, J. M., Gjørup, H., Bäckman, B., & Lovschall, H. (2013). Amelogenesis imperfecta: Gener, proteiner og fænotyper. *Tandlægebladet*, 117(7), 574–582.
- Jensen, R. M., Jensen, T., Lelkaitis, G., & Neumann-Jensen, B. (2013). Traumatisk knoglecyste i collum mandibulae. *Tandlægebladet*, 117(5), 400–404.
- Johansen, M. E., & Jensen, J.-U. S. (2010). Infektionsbiomarkøren procalcitonin til diagnose, prognose og behandlingsvejledning hos intensivpatienter. *Klinisk Biokemi i Norden*, 22(4), 10–16.
- Juul Nielsen, A., & Vallgård, S. (2014). Tværfaglighed som ideal og praksis. *Dansk Universitetspædagogisk Tidsskrift*, 9(16), 113–123.
- Jørgensen, T. (2013). De sygdomsfremmende virkninger. *Bibliotek for Læger*, 205, 184–201.
- Kjær, P., Junker, K., Kongsted, A., Fournier, G., Hartvigsen, J., Kirkeskov, L., ... & Palsson, T. S. (2016). National klinisk retningslinje for ikke-kirurgisk behandling af lumbal nerverodspåvirkning (lumbal radikulopati).
- Kjær, P., Junker, K., Kongsted, A., Schiøttz-Christensen, B., Møller, C., Hansen, I. R., ... & Melbye, M. (2015). National klinisk retningslinje for ikke-kirurgisk behandling af nyopstået rodspåvirkning i nakken.
- Kongstad, J., Fiehn, N.-E., Bindslev, P. H., & Larsen, T. (2014). Biocider i tandlægepraksis. *Tandlægebladet*, 117(9), 750–758.
- Kongstad, J., Fiehn, N.-E., Bindslev, P. H., & Larsen, T. (2014). Biocider i tandlægepraksis. *Tandlægebladet*, 117(9), 750–758.
- Larsen, T., Ciofu, O., Moesby, L., Kirkevang, L.-L., & Poulsen, A. H. (2013). Anvendelse af antibiotika i tandlægepraksis. *Tandlægebladet*, 117(9), 718–729.
- Larsen, T., Ciofu, O., Moesby, L., Kirkevang, L.-L., & Poulsen, A. H. (2013). Anvendelse af antibiotika i tandlægepraksis. *Tandlægebladet*, 117(9), 718–729.
- Larsen, T., Holmstrup, P., Fiehn, N.-E., & Dahlén, G. (2016). Orale bakterier og sygdomme udenfor mundhulen. *Tandlægebladet*, 120(5), 436–441.
- Lippert, M. L. (2011). Kvalitetsmålinger i almen praksis: ny teknologi og gamle logikker. *Tidsskrift for Forskning I Sygdom Og Samfund*, 8(15), 97–120.

- Maindal, H. T., & Vinther-Jensen, K. (2016). Sundhedskompetence (Health literacy)–teori, forskning og praksis-Health Literacy–Theory, Research and Practice. *Klinisk Sygepleje*, 30(01), 3-16.
- Martiny, F., Malmqvist, J., Lykke Bie, A. K., Lykke Toft, E., Enggaard Kaae, S., & Brodersen, J. (2017). Screening for tarmkræft. *Practicus*, 40(236), 38–40.
- Mathiasen, R., Hansen, B. M., Løkke, A., & Greisen, G. (2008). Behandling af tidligt fødte børn på Rigshospitalet i perioden 1955–2007. *Bibliotek for læger*.
- Mikkelsen, K. H., Knudsen, S. U., & Jørgensen, L. N. (2013). Rejser og venøs tromboemboli. *Ugeskrift for Læger*, 175(44), 2628–2631.
- Modvig, J. (2014). Lægers medvirken til tortur. *Bibliotek for Læger* nr. 2, 2014. 147-160.
- Nielsen, G., & Stelter, R. (2011). Mellem social kreativitet og sportslige kompetenser. *Idrottsforum.org*, 1–16.
- Nielsen, L., Curtis, T., Grønbæk, M., & Nielsen, N. R. (2007). Forebyggelse og behandling af stress i Danmark. Statens Institut for Folkesundhed.
- Nielsen, N. K. (2007). Topografi, krop og idræt - eller noget om at få en røget fisk som tak for kampen. *Www.idrottsforum.org./Articles*, 1–9.
- Nymark, A. (2016). Spondylodiscitis—et pilotprojekt med fokus på patienternes oplevelser-Spondylodiscitis—a pilot study with focus on the patients' perspective. *Klinisk Sygepleje*, 30(02), 101-113.
- Nørgaard, L., Kier-Swiatecka, E., Oliver, A., Ingerslev, J., Urth Hansen, K., & Thorn, J. J. (2016). Det kliniske spektrum ved orale planocellulære karcinomer. *Tandlægebladet*, 120 (7), 612-619
- Pedersen, A. M. L. (2015). Saliva som diagnostisk redskab-muligheder og begrænsninger. *Norske Tannlegeforenings Tidende*, 125(2), 112-118.
- Pedersen, S. D. N., Dabelsteen, S., & Bjørndal, L. (2014). Inflammation og hårdtvævsdannelse i pulpa. *Tandlægebladet*, 118(6), 430–8.
- Pedersen, T. K., Dalstra, M., Cattaneo, P. M., Blomlöf, P. J. S., Buhl, J., Nørholt, S. E., & Melsen, B. (2014). Anvendelse af digitale dentale og 3-d-virtuelle modeller i ortodontiske, ortokirurgiske og rekonstruktive kirurgiske behandlinger. *Tandlægebladet*, 118(1), 14–20.
- Pisinger, C. (2013). Tobaksindustrien: "Spred tvivl om stærk videnskabelige evidens og offentligheden vil ikke vide, hvad den skal tro". *Bibliotek for Læger*, 205, 156–181.
- Provencal, P. (2014). Sygehusene i den islamiske verden i den klassiske tid. *Bibliotek for Læger*, 206(1), 28–41.
- Ramsgaard-Jensen, T. F., Horskjær, M., Jensen, L. B., Due, K. M., & Grønkjær, M. (2014). n-CPAP behandling forårsager alvorlige nasale hudproblemer hos de mindste præmature børn. *Klinisk Sygepleje*, 28(1), 35–45.
- Rasmussen, E. H. R., Fast, S., Tagesen, J., & Bygum, A. (2013). Angioødem med og uden urticaria-odontologisk perspektiv. *Tandlægebladet*, 117(6), 474-480.
- Rehfeld, J. F., Bardram, L., Gøtze, J. P., & Hilsted, L. M. (2011). Gastrin som markør for gastrinomer: Et lærestykke om immunanalyse-kits. *Klinisk Kemi I Norden*, 23(3), 20–23.



- Roos, E. M., Hartvigsen, J., Bliddal, H., Mølgaard, C., Christensen, R. D. K., Søgaard, K., & Zebis, M. K. (2013). Forebyggelse af skader og sygdomme i muskler og led.
- Rud, K., Egerod, I. E., & Brodersen, J. (2014). Patientoplevelse af accelererede brystkræftoperationer belyst ved spørgeskemaundersøgelse. *Klinisk Sygepleje*, 28(1), 46–62.
- Røder, M. A., & Iversen, P. (2012). PSA screening for prostatacancer. *Lægemagasinet*, 5, 14–16.
- Schjøtler, G., & Delmar, C. (2016). Angst, sårbarhed og skam—operationspatienters sanseerfaringer—Anxiety, Vulnerability and Shame—Surgical Patients' Sensory Experiences. *Klinisk Sygepleje*, 30(01), 17–30.
- Schou, S., & Hillerup, S. (2010). Videreuddannelse til specialtandlæge i tand-, mund- og kæbekirurgi. *Tandlægebladet*, 114(13), 1046–1048.
- Staun, L., Nyvad, B., & Richards, A. (2013). Aktivt fluorid i fluortandpastaer på det danske marked. *Tandlægebladet*, 117(8), 638–644.
- Stelter, R. (2012). Tredje generations coaching. *Coaching Psykologi - the Danish Journal of Coaching Psychology*, 2(1), 11–25.
- Stelter, R., Nielsen, G., & Wikman, J. M. (2011). Narrativ-samskabende gruppecoaching udvikler social kapital. *Coaching Psykologi - the Danish Journal of Coaching Psychology*, 1(1), 67–77.
- Storm, R. K., & Nielsen, K. (2010). Dansk eliteidræts konkurrenceevne. *Nordic Sport Studies Forum*, 1(1), 27–50.
- Sudergaard, M. M., Jørgensen, B. D., & Lorenzen, M. D. (2017). Skærpet fokus på hurtigt at fjerne kateteret. *Sygeplejersken*, 2017(4), 62–67.
- Wadmann, S. (2013). Informeret samtykke i kliniske forsøg: teknikaliteter, tillid og tætte relationer. *Etikk I Praksis*, 7(2), 31–46.
- Winther, H. (2013). Først og bedst over målstregen med Body-sds? *Idrottsforum.org*, 1–9.

### *Humanities*

- Andersen, M. C. (2014). Tilsynekomsten af den engelske Assam-te. *Kulturstudier*, Nr 2. 2014, 84–98.
- Andersen, N. M. (2010). Talesprog og sproglig polyfoni. *Tijdschrift Voor Skandinavistiek*, 31(2), 3–23.
- Auken, S. (1997). Stjernernes Morgensang. Om N.F.S. Grundtvigs historiske salme: Hyrderne ved Bethlehem. *Grundtvig-Studier*, (1997), 212–217.
- Auken, S. (2011). Verdensklasse eller blålys. *Akademisk Kvarter*, 2, 13–30.
- Bandak, A. (2011). Refræner i Damaskus – kristne og nationale. *Kult*, (8).
- Blicher, H. (2014). Senklassicistisk omgang med tekster. Om åbenbare og skjulte citater i Jens Baggesens Labyrinten. *European Journal of Scandinavian Studies*, 44(2), 270–285.
- Bloch, D. (2015). Tekstkritiske udgaver og klassiske filologer: En analyse af udgaverne af de platoniske breve. *Aigis*, 15, 1–18.
- Bloch, D. K. (2008). Græsk filosofi og dødsriget. *Aigis*, 8(1), 1–12.

- Boisen, J. (2008). Var Marquis de Sade sadist? *Passage*, (59), 103–114.
- Boisen, J. (2015). "Vær frie - det er en ordre!" Verden vs. Charlie Hebdo. *Fransk Nyt*, (268), 32–39.
- Bolt, M. (2014). Grænser for deltagelse. *K And K*, 43(118), 193–204.
- Borgnakke, K. (2015). Evidensbevægelsen i spændingsfeltet mellem sikker viden og ikke-viden. *Cebra-Striben*, (17), 22–31.
- Borum, P. (2016). En bergsoniansk pædagogik? *Studier I Pædagogisk Filosofi*, Årg. 3, Nr 2 (2014)(22449140), 61–86. <http://doi.org/10.7146/spf.v3i2.7358>
- Christensen, C. L. (2008). Livsstil som tv-underholdning. *Mediekultur*, (45), 24–36
- Christensen, C. L. (2013). Sundhed på tv. Fra læge til sundhedsguru. *Mediekultur*, 29(54), 104–122.
- Christensen, L. H. (2012). Dansk på stedet. *Dansk Universitetspædagogisk Tidsskrift*, (13), 27–40.
- Christensen, M. E. (2013). Den svære kunst – refleksioner over begrebet kunst i en arkæologisk kontekst. *Arkæologisk Forum*, Vol 29, 15-18
- Christensen, T. K., & Jensen, T. J. (2014). Det er klart at ledsætninger har da ledsætningsordstilling- eller har de?. *Nyt Fra Sprognævnet*, 2014(1), 1-7.
- Cold, C. (2014). Den ansigtsløse torturbøddel - Magt og bøddel i et historisk-sociologisk perspektiv. *Bibliotek for læger*, juni 2014, 112-142
- Dahl, C. (2017). Topos og motiv. *K And K*, (123), 23–36.
- Dalby, M., Elbro, C., & Mårbjerg, S. (2009). Langtidsprognosen for sprogindlæringsvanskeligheder. *Ugeskrift for Læger*, 171(1), 37–41.
- Dam Christensen, H. (2011). Skjulte forbindelser? *Nordisk Museologi*, (2), 3–18.
- Damsholt, T. (2016). "At overskue, tilfredsstille og lyksaliggjøre". *K And K*, 121.
- Daugaard, L. M., & Johansen, M. B. (2012). Tosprogede børns møde med metafiktion i en postmoderne billedbog. *Nordic Journal of Childlit Aesthetics*, 3.
- Due, B. L. (2010). Rhizom-møde-modellen. *Chara - Tidsskrift for Kreativitet, Spontaneitet Og Læring*, 1(4), 493–512.
- Due, B. L. (2012). Den narrative konstruktion af en idé. *Språk Och Interaktion*, (3), 45–89.
- Due, B. L. (2015). Idéudviklingens trepartstruktur og den katalytiske funktion af kritik. *Språk Och Interaktion*, 4(2), 41–62.
- Elbeshausen, H., Jensen, A. L., & Riis, R. (2013). Fra kollegial supervision til fællesfaglig udvikling - et eksperiment med entreprenørskabsundervisning i faget Kulturformidling. *Dansk Universitetspædagogisk Tidsskrift*, årg 8(14), 66–75.
- Elbro, C., & Farø, K. J. (2013). Tegnsætning med mening. *Mål Og Mæle*, 36(1), 17-23.
- Engberg, S. G. (2008). Profetologion og tre-læsningsteorien. *Patristik*, 9.
- Farø, K. (2011). Onomasiologiskhed og leksikografi. *LexicoNordica*, (18).

- Farø, K. (2013). Inventariografi–al leksikografis og grammatikografis mo (r) der?. *LexicoNordica*, (20).
- Farø, K. J. (2011). Partileksikografi dansk-tysk: Problemtyper og et koncept. *Leda-nyt*, (52).
- Fastrup, A. (2007). Kulturel liminalitet i Don Quixote. *K And K : Kultur Og Klasse : Kritik Og Kulturanalyse*, 103, årg. 35(1), 80–107.
- Fjeldsøe, M. (2010). Vitalisme i Carl Nielsens musik. *Danish Musicology Online*, 1, 33–55.
- Frederichsen, K. (2011). Venskabshuset: Et sovjetisk kulturcenter i København, 1977-1992. *Arbejderhistorie. Tidsskrift for Historie, Kultur Og Politik*, (2), 68-86.
- Frederichsen, K. (2014). De russiske statsformer og krigsfangenskab under og efter Første Verdenskrig, 1914-1922. *Magasin Fra Det Kongelige Bibliotek*, 27(3), 57–66.
- Frederiksen, B. O. (2007). Om "Buk fra Faar at skille" *Renæssanceforum : Tidsskrift for Renæssanceforskning*, 3.
- Funch, B. S. (2003). Den fænomenologiske metode i museologisk forskning. *Nordisk Museologi*, (1), 17.
- Galal, E. (2013). Nationale tv-strategier under det arabiske forår. *Tidsskrift for Islamforskning*, 7(1), 53–68.
- Gram-Hansen, L. B., & Gram-Hansen, S. B. (2013). Fra Geocaching til Persuasiv Læring. *Munkiana*, 17(52), 16–35.
- Gravengaard, G. (2008). "Det var, hvad vi ikke valgte at bringe". *Mediekultur*, 44, 50–57.
- Gregersen, N. H. (2015). Bekendelsens briller: Om hvorfor ortodoksien ikke er for fastholdere. *Dansk Tidsskrift for Teologi Og Kirke*, 42(3), 247–261.
- Gremaud, A.-S. N. (2014). Krisegeografier i islandsk fotografi. *Ekfrase Nordisk Tidsskrift for Visuell Kultur*, (2), 61–77.
- Grønholdt-Pedersen, J. (2008). Ændrede talevilkår–bedre interviews. *MedieKultur: Journal of media and communication research*, 24(44), 13.
- Gudme, A. K. de H. (2015). Tomhændet må ingen se mit ansigt. *Dansk Teologisk Tidsskrift*, 78/4, 300–319.
- Gaarden, M. (2013). Den empiriske fordring til homiletikken. *Halvårsskrift for Praktisk Teologi*, 30(2), 3–20.
- Hald, M. M., & Skuldbøl, T. B. B. (2012). Hvorfor graver vi? *Arkæologisk Forum*, 26, 37–40.
- Handberg, K. (2014). Populuxe Pop: Retroficeringen af 1950'erne i albumcovere. *Ekfrase*, 4(02), 94-109.
- Hansen, H. (2011). Aristoteliske relative: En oversættelse af *Metafysikken V 15*. *Aigis, Vol. Supplementum I*, 1-11
- Hastrup, F., & Brichet, N. S. (2016). Antropocæne monstre og vidundere. *Kartofler, samarbejdsformer og globale forbindelser i et dansk ruinlandskab. Tidsskriftet Kulturstudier*, 2016(1).

- Hedin, G. (2011). Forfatteren som eksperimentator. *Edda*, 98(4), 330–344.
- Hesselager, J. (2008). Analytiske tilgange til Verdis musikdramaturgiske formsprog. *Musik and Forskning*, 31 2007, 19–43.
- Holm, I. W. (2010). Æstetik og politik. *K And K : Kultur Og Klasse : Kritik Og Kulturanalyse*, (110), 147–155.
- Holm, I. W. (2012). Mennesker vs. zombier. *Kritik*, 204, 122–136.
- Holm, I. W., & Tygstrup, F. (2008). Livets rum, erindringens form. *Passage*, (58), 97–116.
- Holst, S., Christensen, P., Kristensen, S., Norsker, A., Larsen, A., Siegismund, K., & Ehrensvärd, M. G. (2015). Det Syriske Adams Testamente. *Dansk Teologisk Tidsskrift*, 78(3), 163–184.
- Hultgren, A. K. (2011). Fup og fakta i debatten om domænetab. *Nyt Fra Sprognaevnet*.
- Hvithamar, A. (2015). Jehovas Vidner. *Religionsvidenskabeligt Tidsskrift*, 62, 125–137.
- Hvithamar, A., & Schaffalitzky de Muckadell, C. (2013). Case-baseret undervisning. *Dansk Universitetspædagogisk Tidsskrift*, 8(14), 41–50.
- Isager, C. (2015). Haderen, elsker og læseren: Holdning, underholdning og pædagogisk potentiale i essayistisk skrivelitteratur. *Journalistica*, 2015(1), 4–23.
- Jacobsen, B. A. (2008). Muslimer i mandtal. *Tidsskrift for Islamforskning*, 3(1), 84–110.
- Jacobsen, B. A. (2015). Hellige bygninger på grænsefladen i dansk kommunalpolitik. *Religionsvidenskabeligt Tidsskrift*, 62, 91–105.
- Jensen, E. S. (2013). Ordklasseproblemer, tilfældet sådan. *LexicoNordica*, (20).
- "Jensen, S. E. (2013). Udstillingen som kommunikation – om at kommunikere med arkæologiske genstande i udstillinger. *Arkæologisk Forum*, Vol 29, 19-25"
- Jensen, T. J. (2009). Non-realitetsmarkører i dansk talesprog. *Studier I Nordisk*, 75–101.
- Jensen, T. J. (2009). Refleksivt anvendte pronominer i moderne dansk. *Ny Forskning I Grammatik*, 16, 131–151.
- Jensen, T. J. (2011). Ordstilling i ledsætninger i moderne dansk talesprog. *Ny forskning i grammatik*, (18).
- Johansen, B. S., & Johansen, K. S. (2008). Islam i dansk psykiatri. *Tidsskrift for Islamforskning*, 3(1), 30–43.
- Johansen, M. B. (2014). Dic cur hic. *Etikk I Praksis*, 8(2), 50–68.
- Johansen, M. B. (2015). "Jeg har forstået den sådan, at den ikke skal forstås." *Acta Didactica Norge - Nasjonalt Tidsskrift for Fagdidaktisk Forsknings- Og Utviklingsarbeid*, 9(1), 1–20.
- Johansen, M. B., & Frederiksen, L. A. L. (2013). Teori og praksis i de danske professionsuddannelser. *Nordisk Barnehageforskning*, 6(5), 1–12. <http://doi.org/10.7577/nbf.578>

- Johansen, M. B., & Morsing, O. (2014). "En sværm af skyer, som skal tænkes." *Studier I Pædagogisk Filosofi*, 3(2), 1–20. <http://doi.org/10.7146/spf.v3i2.18132>
- Jørgensen, C. (2012). *Snyd med sproghandlinger*. *Sakprosa*, 4(2).
- Jørgensen, J. L. (2011). Litteraturens skabende viden. *Edda. Nordisk Tidsskrift for Litteraturforskning*, 2011(3), 226–237.
- Kappel, K. (2014). Sandheden og den usynlige hånd. *Politik*, 17(1), 15–22.
- Kirilova, M. K. (2014). "Det kan være svært" – om sprog og kultur i andetsprogsdanske ansættelsessamtaler. *Nordand : Nordisk Tidsskrift for Andrespråksforskning*, 19(1), 9–36.
- Kirkegaard, A. (2011). Om at fare vild i verdensmusikkens buskads. *Danish Musicology Online*, 2, 53–78.
- Knuth Federspiel, B. (2003). Konserveringsbegrebet i det 20. århundrede. *Nordisk Museologi*, 1, 3–16.
- Kock, C. E. J. (2001). *Imitatio* : en bro mellem tekstlæsning og skriftlig fremstilling. *Dansk*, (2), 3–10.
- Kock, C. E. J. (2008). Fornuftig uenighed. *Rhetorica Scandinavica*, (48), 64–83.
- Kock, C. E. J. (2008). Kan man opstille adfærdsregler for offentlig debat? *Mediekultur*, 44, 58–65.
- Kock, C. E. J. (2009). Kynismesyndromet. *Rhetorica Scandinavica*, (49/50), 51–71.
- Kock, C. E. J. (2012). Retorik. *Dansk Noter*, (2), 14–19.
- Kock, C. E. J. (2012). Retoriske uskikke i offentlig debat - illustreret med eksempler fra prostitutionsdebatten. *Rhetorica Scandinavica*, (59), 94–107.
- Kondrup, J. (2012). Den postmoderne biografi. *Folia Scandinavica Posnaniensia*, 14, 34–44.
- Kragh, K. A. J., & Strudsholm, E. (2015). Deiksis i sprog og kontekst. *Skandinaviske Sprogstudier*, 6(3), 134–157.
- Kramer, M. (2006). Partnervalg på film. *Mediekultur*, (40), 53–63.
- Kristensen, N. N. (2009). Det er et spørgsmål om at gøre det nemt, men ikke at føre pennen. *N O R D I C O M - Information*, 31(1-2), 81–112.
- Kristensen, N. N., & Lejre, C. (2014). En kvantitativ metode til analyse af radio. *Mediekultur*, 30(56), 151–169. <http://doi.org/10.7146/mediekultur.v30i56.15998>
- Kristiansen, T. (2009). Har Norge et standardtalemål? -. *Norsk Lingvistisk Tidsskrift*, 27(1), 79–94.
- Kristiansen, T. (2009). Åbne og skjulte holdninger til engelskindflydelsen. *Språk I Norden*, 95–112.
- Kristiansen, T. (2010). Om betydningen af den senmoderne danske (K/M/L)-variation som henholdsvis Pige-sprog og Dreng-sprog. *Skandinaviske Sprogstudier*, 1(1).
- Kristiansen, T. (2010). Sociolingvistikken og (sprog)politikken. *Danske Talesprog*, 10, 1–33.
- Kristiansen, T. (2011). Den herskende sprogideologis almægtighed. *Københavnstudier I Tosprogethed*, 58.

- Lassen, J. M. M. (2012). Nationalitetskonstruktion i dansk public service. *Nordisk Kulturpolitisk Tidsskrift*, (02), 235–252.
- Lorentzen, H., & Theilgaard, L. (2012). Netordbøger - hvordan finder brugerne frem til dem, og hvad gør de, når det har fundet dem? *Skrifter Udgivet Af Nordisk Forening for Leksikografi*, 418–430.
- Lysemose, K. (2010). Fra tilskuer til skuespiller. *Res Cogitans*, 1(7), 1–32.
- Madsen, C. (2016). Retorik og lykke. *K And K*, Årg. 44(121), 45–61.
- Moberg, B. R. (2007). Op af historiens grøft. Edda. *Nordisk Tidsskrift for Litteraturforskning*, 107(1), 39–57.
- Mortensen, S. S. (2011). Problemer og udveje i dansk ledsætningsklassifikation. *Ny Forskning I Grammatik*, 18, 225–248.
- Møller, J. F. (2012). Katolikkernes stilling og antikatolicismen i Danmark 1800-1849. *Kirkehistoriske Samlinger*, 2012, 113–135.
- Møller, M. (2013). Latter kan mere end at latterliggøre. *Tidsskrift for Medier, Erkendelse Og Formidling*, 1(2), 23–38.
- Nielsen, B. F. (2014). Den Praktiske Teologis grundlagsspørgsmål. *Dansk Teologisk Tidsskrift*, 77(2), 162–169.
- Nielsen, M. F. (2008). Foghs forslag. *Mediekultur*, 44, 16–26.
- Nielsen, M. L. (2012). "Og efter sin elskede"? *Danske Studier*, 2012, 5–23.
- Nielsen, N. J. (2013). Grænseløse arbejdere-en diskussion af identitet og selvbevidsthed med udgangspunkt i polske migrantarbejdere. *Arbejderhistorie: Tidsskrift for historie, kultur og politik*, 2, 44–60.
- Nielsen, T. H. (2006). Herodot og hemerodromoi. *Aigis*, 6.1(6.1., 2006), 1–10.
- Nielsen, T. H. (2015). Hvad betragtede en klassisk græker som sit fædreland? *Aigis*, 15(1), 1–25.
- Nyboe, J. Ø. (2009). Pest eller kernekraft. *Nordica*, 26, 145–168.
- Nyborg, O. (2011). Hvorfor benægter verden, at den sande kærlighed skal findes i fuldkommenhedens bånd? *Tidsskrift for Teologi Og Kirke*, 2, 280–297.
- Olden-Jørgensen, S. (2011). "Herremænd i kongeklæder" eller "af den gamle skole". *Aigis, Supplementum(I)*, 1–14.
- Olsen, N. L. (2009). Lucretia og Dido. *Aigis*, 9(1), 1–23.
- Padovan-Özdemir, M. (2012). Migratoriske dannelsesprocesser. *Dansk Pædagogisk Tidsskrift*, (2), 28–37.
- Petersen, A. R. (2012). Kulturel erindring i migrationens splintrede spejl. *Passepartout. Skrifter for Kunsthistorie*, 18. årgang(33), 16–30.
- Petersen, A. R. (2013). Film som kunst. *Passepartout*, 18(34), 66–86.
- Petersen, C., & Engerer, V. P. (2014). 'Den lange rejse ...' - metaforiske betydningslag og branding i filmmediet. *Mediekultur*, 30(57), 154–175.

- Plum, M. (2013). De pædagogiske læreplaners reformering. *Nordisk Barnehageforskning*, 5(12), 1–12.
- Rasmussen, A. J. (2015). Peter Seebergs poetik hinsides modernisme og realisme. *European Journal of Scandinavian Studies*, 45, 144–160.
- Rasmussen, K. S. G. (2013). Når lyrikken tager form. *Passage*, (69), 35–48.
- Riede, F. (2014). Brommeproblemet 2.1. *Arkæologisk Forum*, 31, 39–45.
- Rosenbeck, B. (2012). Mange veje, nye retninger. *Scandia: Tidsskrift Foer Historisk Forskning*, 78(2 Supplement), 70–76.
- Rösing, L. M. (2014). Topos og trope hos Josefine Klougart. *Spring - Tidsskrift for Moderne Dansk Litteratur*, 35, 68–81.
- Sandberg, A. (2007). Utopi og rejse hos Jens Baggesen. *Tijdschrift Voor Skandinavistiek*, (1), 21–40.
- Sandvik, K. (2005). Politik for computerspil? *Nordisk Kulturpolitisk Tidsskrift*, (1), 48–84.
- Sandvik, K. (2007). Kroppen, musen og den fantastiske avatar. *Turbulens. Kroppen: Hinsides Posthumaniteten*, (10).
- Sandvik, K. (2009). Medieleg af 1. og 2. grad. *Buks - Tidsskrift for Børne- and Ungdomskultur*, (53), 75–87.
- Schwartz, A. (2004). Ovidius cruentus: elementer af vold og gru i Metamorfoserne. *Aigis*, 4.1.
- Schwartz, A. (2004). Xenophons lederskab i Anabasis. *Aigis*, 4.1, 1–25.
- Schwartz, A. (2005). Græker mod græker - hoplitslaget i klassisk tid. *Aigis*, 5.2, 1–11.
- Schwartz, A. (2011). Hvordan man ikke skal skrive historie: Lukian som historiografi- og litteraturkritiker. *Aigis*, 11.2, 1–22.
- Schweppenhäuser, J. (2008). Krydsild. *Kritik*, (189), 73–84.
- Schøsler, L., & Skovgaard-Hansen, M. (2007). Undersøgelse over komplekse nominal-syntagmer i latin. *Renæssanceforum : Tidsskrift for Renæssanceforskning*, (3).
- Simonsen, J. B. (2011). Minoriteter og nationalstaten i Mellemøsten. *Kult*, (8).
- Simonsen, J. B. (2012). Johannes Pedersens ophold i Cairo 1920-1921. *Orientalia Suecana*, LXI(Supplement).
- Sindbæk, T., & Dedovic, I. (2014). Første Verdenskrig som erindring i Bosnien, Kroatien og Serbien. *Nordisk Oestforum*, 28(2), 99–118.
- Skot-Hansen, D. (2013). Fransk kultur i undtagelsestilstand - franske kunst- og kulturarenaer i en digital æra. *Nordisk Kulturpolitisk Tidsskrift*, 16(02), 201–216.
- Skovgaard-Petersen, J. (2013). Syriens kendteste TV-shaykh bakker stædigt op bag styret. Hvorfor? *Tidsskrift for Islamforskning*, 7(1), 69–87.

- Søndergaard, H. (2014). Når en fælles fjende er den bedste ven i nøden. *N O R D I C O M - Information*, 36(1), 65–72.
- Sørensen, S. A. (2012). Kassation på de danske museer. En modedille, et velovervejede paradigmeskift, eller en praktisk løsning? *Arkæologisk Forum*, Vol 27, 13-17
- Sørensen, S. M. (2011). Musens kammerpige. *Danish Musicology Online*, 2, 79–111.
- Sørensen, S. M. (2015). Lutvirtuos og engageret kunstner. *Danish Musicology Online*, 7, 57–80.
- Tegtmejer, T. (2013). Rousseau i 2012: De-konstrueret og re-konstrueret. *Studier i Pædagogisk Filosofi*, 2(1), 70-84.
- Theilgaard, L. (2013). Kära nån, hvad er der med de interjektioner? *Lexiconordica*, 20, 127–143.
- Thisted, K. (2015). Imperiets genfærd–Profeterne i Evighedsfjorden og den dansk-grønlandske historieskrivning. *Nordlit*, (35), 105-121.
- Thøgersen, J. (2013). Vil De prøve at sige A? *Nys. Nydanske Studier Og Almen Kommunikationsteori*, 43, 101–132.
- Thøgersen, U. (2013). Rousseau vivant: en aktualisering af Émile med fokus på det lidenskabelige fænomenfelt. *Studier i Pædagogisk Filosofi*, 2(1), 34-43.
- Troelsgård, C. (2011). Den sandfærdige løgn: Eustathios af Thessalonikis (ca. 1115 – 1195 e.Kr.) Forord til Odyssékommentarerne. *Aigis, Supplementum*(1.51), 1–6.
- Trolle, A. K. (2015). Migrantkirker. *Religionsvidenskabeligt Tidsskrift*, 62(Temanummer Den danske religionsmodels grænseflade), 61–75.
- Ulbæk, I. (2012). Tekstlingvistikkens problemer. *Skandinaviske Sprogstudier*, 3(2), 94–113.
- Warburg, M. (2011). Baha'i i et mellemøstligt perspektiv. *Kult*, 8, 23–38.
- Warburg, M. (2015). De danske udlandskirker. *Religionsvidenskabeligt Tidsskrift*, 62, 45–60.
- Warburg, M. (2015). Den danske religionsmodels grænseflade. *Religionsvidenskabeligt Tidsskrift*, 62, 5–14.
- Wiil, P. H. (2015). Atmosfære på museum. *Nordisk Museologi*, 2015(1), 40–55.
- Øland, T. (2014). Mobilisering af barnets potentiale for en bedre fremtid. *Barnelitterært Forskningskrift*, Vol. 5. <http://doi.org/10.3402/blft.v5.23736>
- Øland, T., & Larsen, V. (2011). Integrationisme i pædagogisk forskning og professionalisme. *Praktiske Grunde. Tidsskrift for Kultur- Og Samfundsvidenskab*, 5(1), 5–16.
- Natural Science*
- Jensen, A. M. D., Mortensen, B., & Paaske, K. (2012). Pesticidforbrug og pesticidbelastning på golfbaner. *Miljøstyrelsen*.
- Agerholm, J. S., Holm, W., & Aalbæk, B. (2014). Perifert gangræn hos kalve: beskrivelse af tre tilfælde hos 1-3 måneder gamle kalve. *Dansk Veterinærtidsskrift*, 97(16), 24-28.
- Andersen, H. T., Samson, J., & Winther, L. (2010). Kunsten at sælge et sted–stedsidentitet og branding. *Samfundsøkonomen*, (6), 28-32.



- Andersen, J. M. (2015). HÈrregården i landskabet. Udviklingen på grevskabet Knuthenborg 1840-1880. *Landbohistorisk Tidsskrift*, 10(2), 36-74.
- Andersen, M. C., & Rugh, S. V. (2011). Naturlove og muligheden for liv. Del 1. *Kvant*, 2, 30-35."
- Bassø, L., Rasmussen, M., & Mikkelsen, P. S. (2014). AMOK—er det bare sund fornuft?: Avanceret online Måling af OverløbsKvalitet. *Eva: Erfaringsudveksling I Vandmiljøteknikken*, 27(4), 10-13.
- Behrens, J., & Buchmann, K. (2015). Parasit påvirker torskens hjerterytmer. *Aktuel Naturvidenskab*, (1), 32-35.
- Berg, J., Boysen, L., Fredholm, M., & Proschowsky, H. F. (2013). Ichthyosis hos golden retriever i Danmark. *Dansk Veterinaertidsskrift*, 96(7), 12-14.
- Biltoft-Jensen, A. A., Ejlerskov, K. T., Fagt, S., Knudsen, V. K., Matthiessen, J., Budtz, A., ... & Fødevareinstituttet, D. T. U. (2014). Danskernes kost-og aktivitetsvaner under lup. E-artikel Fra DTU Fødevareinstituttet, 2014(3), 1-7.
- Bolet, L., Agerholm, N., Tradisauskas, N., & Lahrmann, H. (2012). Brugen af Geografiske Informationssystemer til effektundersøgelse af Intelligent Fartilpasning i Danmark. *Geoforum Perspektiv*, 11(22).
- Bonde, N. C., Madsen, H., Schultz, B., Sylvestersen, R., & Jakobsen, S. L. (2010). Moléret i Nordjylland-diatomitten og fossilerne. *Dansk Naturhistorisk Forenings Årsskrift*, 19, 25-40.
- Bregnballe, T., & Jørgensen, H. E. (2013). Udvikling i ynglebestanden af Fjordterne i Danmark 1970-2012. *Dansk Ornitologisk Forenings Tidsskrift*, 107, 261-280.
- Bregnballe, T., & Lyngs, P. (2014). Udviklingen i ynglebestanden af Sølvmåger i Danmark 1920-2012. *Dansk Orn. Foren. Tidsskr*, 108, 187-198.
- Bregnballe, T., Christensen, H., & Drachmann, J. (2015). Udviklingen i ynglebestanden af HÈttemÅger i Danmark 1970-2010. *Dansk Ornitologisk Forenings Tidsskrift*, 109, 179-192.
- Bregnballe, T., Thorup, O., Jacobsen, L. B., Kjeldsen, J. P., & Hansen, M. (2015). Udviklingen i ynglebestanden af Klyder i Danmark 1970-2014. *Dansk Orn. Foren. Tidsskr*, 109, 121-133.
- Broholm, M. M., Janniche, G. S., Damgaard, I., Olsson, M. E., Asmussen, O. W., Fjordbøge, A. S., & Kern-Jespersen, H. (2014). Anvendelse af MIP ved kildekarakterisering og vurdering af nedbrydning af chlorerede opløsningsmidler. *Jordforurening. info*, (14), 17-27.
- Bügel, S. G., Bredie, W., Jensen, A. B., Jensen, J. D., Petersen, C. B., Leth, K., ... & Langer, J. W. (2015). SMAG-Skønne Måltider til Alle Gamle: Hvidbog om nærende måltider med kulinarisk kvalitet til ældre.
- Bødker, R. (2014). Nye aktive redskaber til passiv overvågning. *Dansk Veterinaertidsskrift*, 2014(04), 30-32.
- Baaner, A. L. (2014). Klimatilpasning—Udledning af regnvand fra veje og befæstede arealer. *Nordisk Miljörättslig Tidskrift/nordic Environmental Law Journal*, 2014(2), 43-59.
- Baaner, L. (2013). Autoritative geografiske grunddata på miljøområdet. *Geoforum Perspektiv*, 2013(23), 25-37.

- Caspersen, O. H., & Jensen, F. S. (2011). Færdsel langs danske vandløb. Arbejdsrapport Skov & Landskab.
- Christensen, B. A. (2013). Mere end repræsentationer, i praksis. Om at være forpligtet og engageret. *Geoforum Perspektiv*, 12(23).
- Christensen, F. K., & Sørensen, M. T. (2014). Grundejerforeninger og lokalplanen. *Tidsskrift for Kortlægning og Arealforvaltning*, 122(47), 35-50.
- Christiansen, H. H. (2009). Permafrost og klima-om at tage temperaturen på permafrosten. *GeologiskNyt*, 19(6).
- Dahl-Jensen, D., & Steffensen, J. P. (2013). Indlandsisen på Grønland: et levende field fortæller klimahistorie. *Det Kongelige Danske Videnskabernes Selskab*, 55-66.
- Dam, P. (2013). Overdrev - fra ekstensivt landbrug til intensiv natur?. *Landbohistorisk Tidsskrift*, 10(2), 9-35.
- Damgaard, J., Grosen, B., Martinez, K., Nielsen, O. F., Kern-Jespersen, H., Janniche, G. S., ... & Jørgensen, T. H. (2013). Geologisk og hydrogeologisk konceptuel modeludvikling for moræner og kalk. *Jordforurening. info*, (2), 9-17.
- Didaktik, N., & von Stemann, J. H. Udfordringer ved undervisning i enzymer. *MONA*, 49.
- Dolmer, A. P., Poulsen, L. K., Blæsbjerg, M., Kristensen, P. S., Geitner, K., Christoffersen, M., & Hoffmann, E. (2009). Konsekvensvurdering af fiskeri af østers i Nisum Bredning 2010/2011.
- Dupont, N. H., Fertner, M. E., Pedersen, K., Toft, N., & Stege, H. (2014). Rapportering af danske svins antibiotikaforbrug: hvor stor betydning har beregningsmetoden?. *Dansk Veterinærtidsskrift*, 97(14), 33-38.
- Fjordbøge, A. S., Kjeldsen, P., Petersen, P. A., & Durant, N. D. (2007). Oprensning af forureningen på depotet ved Høfde 42 ved hjælp af nul-valent jern.
- Graudal, L., Nielsen, U. B., Schou, E., Thorsen, B. J., Hansen, J. K., Bentsen, N. S., & Johannsen, V. K. (2014). Muligheder for bæredygtig udvidelse af dansk produceret vedmasse 2010-2100: perspektiver for skovenes bidrag til grøn omstilling mod en biobaseret økonomi. Institut for Geovidenskab og Naturforvaltning, Københavns Universitet.
- Graversen, J. T., & Nørgaard, N. H. (2000). Netværkssamarbejde i svineproduktion: etablering af kontrakter omkring alt ind-alt ud produktion.
- Hald, A. B., Petersen, J. S., Forkman, B., & Andersen, T. (2013). Fluer og fluelarver som nyttedyr i fjerkræproduktion-bedre dyrevelfærd, nye fodertyper og affaldsreduktion. *Dansk Erhvervsfjerkræ*, 42(11), 662-663.
- Hansen, S. H., Pedersen, K., Sollerman, J., Andersen, A. C., Fynbo, J. P. U., Hjorth, J., ... & Watson, D. J. (2009). Mørkt stof – vi ved så meget og dog så lidt. *Kvant*, (2), 14-17.
- Hasler, B., Hansen, L. B., Andersen, H. E., & Konrad, M. (2015). Modellerings af omkostningseffektive reduktioner af kvælstoftilførslerne til Limfjorden: Dokumentation af model og resultater.
- Haustein, S., & Nielsen, T. A. S. (2015). Delebilisme: nye koncepter, brugersegmenter og effekter. *Trafik and Veje*, 92(8).

- Hjulsager, C. K., Krog, J. S., Hansen, M. S., Chriél, M., & Larsen, L. E. (2015). Fugleinfluenzavirus H10N7 spredte sig blandt danske sæler i 2014. *Dansk Veterinærtidsskrift*, (9), 42-42.
- Houmark-Nielsen, M. (2012). Hvad fortæller vore store vandreblokke om alderen af det danske istidslandskab: Kosmogen eksponeringsdatering af kæmpesten. *Geologisk Tidsskrift*, 2012, 1-13.
- Jacobsen, A. L., Skov, C., Berg, S., Koed, A., & Larsen, P. F. (2008). Udsætning af geddeyngel som bestands-ophjælpning i danske brakvandsområder—effektivrurdering og perspektivering. DTU Aqua-rapport, (196-08).
- Jacobsen, C., Holmer, I., Jensen, J. S., Madsen, S. S., Lund-Larsen, M., & Mejborn, H. (2016). Effekten af stegetid og-temperatur på kvaliteten af spiseolier. DTU Fødevareinstituttet.
- Janniche, G. S., Kern-Jespersen, H., Christensen, A. G., Grosen, B., & Broholm, M. M. (2013). Anvendelse af Water FLUTE multi-level vandprøvetagning til DNAPL-karakterisering. *Jordforurening. info*, (2), 4-8.
- Jensen, A. H. J. (2009). Matematik er kunst uden pensel, kunst er matematik uden kridt. *Kvant*, 25-31.
- Jespersen, A., Hammer, A. S. V., Jensen, H. E., & Larsen, P. F. (2013). Sår hos mink: nyt forskningsprojekt skal udvikle et værktøj, der kan danne fagligt grundlag for vurdering af sår i praksis. *Dansk Veterinærtidsskrift*, 96(6), 12-15.
- Jørgensen, H. E. (2015). Ynglebestanden af Toppet Lappedykker i Østdanmark 1970-2010. *Dansk Orn. Foren. Tidsskr*, 109, 11-23.
- "Klinkby, E. B. (2008). ATLAS-detektoren ved LHC. *Kvant*, 8-11.
- Knudsen, A. V. K., & Fødevareinstituttet, D. T. U. (2014). Danskernes forbrug af kosttilskud. E-artikel Fra DTU Fødevareinstituttet, (2), 1-6.
- Kristensen, P. S., & Hoffmann, E. (2004). Bestanden af blåmuslinger i Limfjorden 1993 til 2003. *Danmarks Fiskeriundersøgelser*.
- Kærgård, N. (2014). Markederne, økonomerne og moralen. *Samfundsøkonomen*, 2014(3), 5-9.
- Kærgård, N., & Dalgaard, T. (2014). Dansk landbrugs strukturudvikling siden 2. verdenskrig. *Landbohistorisk Tidsskrift*, 11(1-2), 9-33.
- Langerhuus, A. T., Lindholst, S., Andersen, H. R., Chhetri, R. K., Hansen, K. M. S., Sundmark, K., ... & Kragelund, C. (2016). Effektiv bionedbrydning af lægemidler i spildevand ved biofilmbaseret teknologi. *Spildevandsteknisk Tidsskrift*, 44(2), 29-30.
- Lassen, A. D., Vognsen, K. M., & Gross, G. (2014). Nøglehullet på spisesteder: Erfaringer med certificerede spisesteder. E-artikel Fra DTU Fødevareinstituttet, 2014(5), 1-12.
- Lerche, A. B. M. (2010). Fang CO2 med aminosyrer. *Aktuel Naturvidenskab*, 6, 24-27.
- Lund, H., & Mathiesen, B. V. (2009). IDA's Klimaplan 2050-Fagligt Notat: Konsekvensanalyse af tilføjelse af CCS-anlæg til IDA's Klimaplan 2050.
- Mejborn, H., Ygil, K. H., Fagt, S., Trolle, E., Kørup, K., & Christensen, T. (2014). Danskernes fuldkornsindtag 2011-2013. E-artikel fra DTU Fødevareinstituttet, 2014(4), 1-7.

- Nielsen, C. S., Ivarsson, A., & Schramm, J. (2010). Anvendelse af koldpresset rapsolie på varebiler: Rapport til Trafikstyrelsen, Center for Grøn Transport.
- Nielsen, J. S., & Lassen, T. B. L. (2015). Video som metode til undersøgelser af fødebiologi hos Stor Hornugle. Dansk Ornitologisk Forenings Tidsskrift, 109, 161-166.
- Nielsen, R., Thøgersen, T., Andersen, J. L., Dalskov, J., & Kusier, R. (2015). Situationsbeskrivelse af den danske fiskeri-, akvakultur-og fiskeindustri sektor.
- Nørgaard, P., Nielsen, M. V., Helander, C., Eknæs, M., & Nadeau, E. (2015). Drægtige fårs foderoptagelse aftager med stigende tyggetidsindeks: model eksempel til estimering af drægtige fårs grovfoder optagelse. Faar, 80(1), 8-11.
- Orellana, F. (2008). databehandling af LHC-data med GRID. Kvant, 3, 32-34.
- Pedersen, A. B., & Jensen, A. (2016). Brugerundersøgelse og videreudvikling af rådgivning i Kloge Fødevareindkøb: Videnskabelig rapport fra DCE-Nationalt Center for Miljø og Energi nr. 180.
- Pedersen, G. K., Pedersen, S. A. S., Bonde, N. C., Heilmann-Clausen, C., Larsen, L. M., Lindow, B. E. K., ... & Storey, M. (2011). Molerområdet geologi-sedimenter, fossiler, askelag og glacialtektonik. Geologisk Tidsskrift, 2011(12), 41-135.
- PEDERSEN, N. L. (2006). Hvordan påvirker temperaturen den bakterielle omsætning?. Vand Og Jord, 13(3), 88-92.
- Rasmussen, B., & Hoffmeyer, D. (2015). Lydisolation mellem boliger i etagebyggeri-Kortlægning og forbedringsmuligheder. SBI forlag, 2015
- Rasmussen, S. E. (2009). Analyse ved røntgendiffraction-Rietveld-metoden. GeologiskNyt, 19(6).
- Rasmussen, T. B. (2015). Rabies 2014-Rabies hos dyr. EPI-Nyt, (Uge 14/15).
- Ravensbeck, L., Frost, H., & Andersen, P. (2013). Fiskeri, økosystemtjenester og økonomi. Nationaløkonomisk Tidsskrift, 151, 259-277.
- Reeler, N. E. A., Nielsen, O. F., Sauer, S. P., Kjærgaard, H. G., Borring, N., Filtenborg, T., ... & Wadum, J. (2013). Raman af hvide pigmenter. Dansk Kemi, 94(12), 30-34.
- Rousing, T., Thomsen, P. T., Sørensen, J. T., Otten, N. D., & Houe, H. (2013). Nødvendigt med flere mål for at vurdere dyrevelfærden i en malkekvægsbesætning. Dansk Veterinærtidsskrift, 96(2), 14-16.
- Skov, H. (2012). Udkantsdanmark skal kortlægges. Geoforum Perspektiv, 11(22).
- Skov-Petersen, H., Nielsen, T. A. S., Nyed, P. K., Senstius, J., & Jensen, C. (2015). Bystruktur og cyklisme fase I: Betydningen af regional placering, detaljeret bystruktur, cykelstier, parkering og kollektiv transport for cykelture til/fra boliger og arbejdspladser.
- Skovsgaard, T. L. (2014). Procesvalg, rekvisition og forberedelse-regelgrundlaget for proces før skelforretningen. Tidsskrift for Kortlægning og Arealforvaltning, 122(47), 1-10.
- Skovsgaard, T. L. (2014). Sagens oplysning-regelgrundlaget for proces i skelforretningssager. Tidsskrift for Kortlægning og Arealforvaltning, 122(47), 11-22.
- Spodsberg, E. M. H. (2013). Behandling af felin hyperthyroidisme. Dansk Veterinærtidsskrift, 96(14), 30-33.

Stender, M. (2015). Kan steders identitet designes?. *Byplan*, 67(1 (April)), 39-47.

Švitra, G., Vilhelmsen, F., & Karsholt, O. (2012). Udbredelsen af *Eupithecia innotata* (Hufnagel, 1767) og *E. ochridata* Schütze & Pinker, 1968 i Danmark. *Lepidoptera*, 10(4), 119-129.

"Sørensen, J. J. W.H., Pedersen, M. K., & Sherson, J. F. (2014). Spillere hjælper kvantecomputeren på vej. *Kvant*, 31-35.

Tofft, J. (2003). Tranens indvandring til Sønderjylland 2002-13. *ret*, 1(2002).

Wøhlk, R., & Danielsen, S. K. (2012). Analyse af GIS-baseret borgerinddragelse i danske kommuner. *Geoforum Perspektiv*, 11(21).

Ørum, J. E., & Samsøe-Petersen, L. (2014). Bekæmpelsesmiddelstatistik 2013: behandlingshyppighed og belastning. Orientering fra Miljøstyrelsen nr. 6, 2014. Miljøstyrelsen, Copenhagen, Denmark. 1-66.

### *Social Science*

Andersen, B. S. (2008). Ironiens kultur og kulturens ironi. *Dansk Sociologi*, 19(1), 9–30.

Andersen, H. (2012). Forskere - rekruttering, karriere og social baggrund. *Dansk Sociologi*, 23(2), 9–38.

Andersen, J. S., Drenck, N.-E., & Keiding, H. (2007). Diagnoserelaterede grupper på intensivafdelinger. *Ugeskrift for Læger*, 169(8), 727–729.

Andersen, J., & Arnfjord, S. (2016). Grønlandsk socialt arbejde i et empowermentperspektiv. *Uden for Nummer*, 16(33).

Andersen, S. K., Hansen, L. L., Navrbjerg, S. E., & Mailand, M. (2011). Europæiske arbejdsmarkedsforskere samlet til IIRA kongres i København. *Tidsskrift for Arbejdsliv*, 13(1), 103–107.

Andersson, C., Bach, A. S., & Breengaard, M. H. (2014). Familier og forældreskaber anno 2014. *Kvinder, Køn and Forskning*, 23(1-2), 3–8.

Arnoldi, J. D. (2003). Aktør-Netværkteori. *Dansk Sociologi*, 14(3), 9–23.

Arnoldi, J. D. (2005). (Medieskabt) Ekspertise i medierne. *Dansk Sociologi*, 16(3), 9–24.

Auken, S., & Emmeche, C. (2010). Mismåling af forskningskvalitet. Sandhed, relevans og normativ validitet i den bibliometriske forskningsindikator. *Kritik*, (197), 2–12.

Balling, G., & Grøn, R. (2012). Formidling af læseoplevelser? *Dansk Biblioteksforskning*, nr.3, 7–17.

Balling, G., Dahl, T. A., Mangel, A., Nilsson, S. K., Lund, H., & Höglund, L. (2014). E-bogen. Skandinaviske perspektiver på forskning og uddannelse. *Nordisk Tidsskrift for Informationsvidenskab Og Kulturformidling*, 3(1), 5–19.

Bech, H. (1999). Sociologiens tredje. *Dansk Sociologi*, 10(3), 69–92.

Bech, H. (2002). Det onde i Danmark. *Dansk Sociologi*, 13(3), 49–74.

Beck Jørgensen, T., & Vrangbæk, K. (2013). Den gode forvaltning - på basis af hvilke værdier? *Nordisk Administrativ Tidsskrift*, 90(3), 115–134.

- Bergqvist, C. (2010). Konkurrenceloven og konkurrencepraksis 1998 - 2009. *Erhvervsjuridisk Tidsskrift*, (3), 149–159.
- Berthelsen, J. (2006). Blandt de blinde er den enøjede konge. *Psyke Og Logos*, 27(1).
- Bhatti, Y., Hansen, K. M., & Olsen, A. L. (2012). Betyder ændringen i skatteprocenten noget for tilslutningen til borgmesterpartiet? *Politica*, 44(4), 563–581.
- Björneborn, L. (2008). Serendipitetsfaktorer og brugeradfærd på det fysiske bibliotek. *Dansk Biblioteksforskning*, 4(2), 41–54.
- Bjørst, L. R. (2011). Klima som sila. *Tidsskriftet Antropologi*, (64), 89–99.
- Bloch, C. (1998). Kroppens stemthed. *Dansk Sociologi*, 9(2), 61–78.
- Bloch, C. (2002). Følelser og sociale bånd i Akademia. *Dansk Sociologi*, 13(4), 43–61.
- Blok, A., Skrydstrup, M. C., & Wahlberg, A. (2012). KOVIKO - et nyt bud på "undervisningsbaseret forskning" i et tværfagligt digitalt lærings- og forskningsfællesskab? *Dansk Universitetspædagogisk Tidsskrift*, 7(13), 120–130.
- Blume, P. E., & Herrmann, J. R. (2009). Registrering af biometriske og biologiske personoplysninger. *Juristen*, 91(3), 73–80. <http://doi.org/10.1000/182>
- Boelsbjerg, H. B. (2013). Det Hellige Rum. *Tidsskrift for Forskning I Sygdom Og Samfund*, (18), 67–86.
- Borch, C. (2006). Massen. *Dansk Sociologi*, 17(2), 43–57.
- Brichet, N. (2014). Grønlandsk og globalt guld. Grundstoffer i antropologisk analyse. *Kulturstudier*, 5(2), 48–64.
- Brodersen, L., Staunstrup, J. K., Møller, J., & Knak-Nielsen, S. (2011). Den rådne banan på kort. Visual literacy - fra billeder til viden. *Dansk Biblioteksforskning*, 7(2/3), 69–82.
- Bruun, H. H. (2003). En klassiker - død eller levende? *Dansk Sociologi*, 14(4), 61–75.
- Baarts, C. (2006). Druk, bajere og løgnehistorier. *Dansk Sociologi*, 17(1), 67–83.
- Crone, M. (2002). Autenticitet og kritisk sprogfællesskab hos Charles Taylor. *Dansk Sociologi*, 13(4), 27–40.
- Dahlgaard, J. O., Hansen, J. H., Hansen, K. M., & Larsen, M. V. (2015). Hvordan påvirkes vælgerne af meningsmålinger? Effekten af meningsmålinger på danskernes stemmeadfærd og sympati for partierne. *Politica*, 47(1), 5–23.
- Dahlkild, N. (2007). De første folkebiblioteksbygninger. *Dansk Biblioteksforskning*, 3(3), 41–55.
- Dahlkild, N. (2013). Biblioteket som rum i byens rum. *Dansk Biblioteksforskning*, 2(1), 21–36.
- Dam Christensen, H. (2012). Informativ affektivitet. Ekfrase *Nordisk Tidsskrift for Visuell Kultur*, 2(1), 4–22.
- Dammeyer, J. H. (2012). Tegnsprogsforskning. *Skandinaviske Sprogstudier* (2011), 3(2), 31–46.
- Dragsted, B., Kristensen, M. M., Nielsen, S., Pauli, T. L., Steinmejer, A., & Oxlund, B. (2011). Bevægelse som universalløsning. *Kulturstudier*, (2), 92–118.

- Due, J. J., & Madsen, J. S. (2013). 20 år med den danske model. *Tidsskrift for Arbejdsliv*, 15(1), 94–103.
- Dyrbye, M. (2011). På sporet af den nye medietid. *Dansk Biblioteksforskning*, 7(1), 45–55.
- Dyrbye, M. (2014). En flytning i "Vinterens Hjerter". *Det Kongelige Bibliotek. Magasin*, 27(1), 47–60.
- Elkjær, H. K., Mortensen, E. L., Poulsen, S. B., Kristensen, E., & Lau, M. E. (2012). (U)overensstemmelse mellem behandler- og patientvurdering af effekt af gruppeterapi. *Matrix*, 29(1), 23–38.
- Elling, R. C. (2014). Oliebyen midt i en jazztid. *Tidsskriftet Kulturstudier*, (2), 24–47.
- Engerer, V. P. (2013). Faglig vs. informationsrelateret læring. *Nordisk Tidsskrift for Informationsvidenskab Og Kulturformidling*, 2(2), 47–57.
- Engerer, V. P. (2013). Kan forskningsbibliotekernes e-materialer være e-læringsmaterialer? *Læring Og Medier*, 6(11), 1–37.
- Engerer, V. P. (2014). Læringscykluser. *Dansk Universitetspædagogisk Tidsskrift*, 9(16), 72–83.
- Eriksson, R. (2012). Hvad er det der skal formidles? *Nordisk Tidsskrift for Informationsvidenskab Og Kulturformidling*, 1(3), 31–40.
- Eriksson, R. (2014). Faglitterære fortællinger: Hvad gør de, hvad kan de? *Nordisk Tidsskrift for Informationsvidenskab Og Kulturformidling*, 2(3), 7–18.
- Gad, U. P. (2008). Når mor/barn-relationen bliver teenager. *Politica - Tidsskrift for Politisk Videnskab*, 40(2), 111–133.
- Gad, U. P. (2009). Hvem taler? Hvem lytter? *Babylon. Tidsskrift Om Midtøsten Og Nord-Afrika*, 7(1), 88–101.
- Gad, U. P., Hannibal, I., Holst, K., & Adler-Nissen, R. (2011). EUs oversøiske lande og territorier. *Politik*, 14(1), 15–24.
- Gammelgård, J. (2014). Freud og mester-diskursen. *Agora*, (1-2), 92–111.
- Gammeltoft, T., Rasch, V., & Knudsen, L. B. (2007). Provokeret abort og stratificeret reproduktion i Danmark. *Tidsskrift for Forskning I Sygdom Og Samfund*, 7(7), 71–85.
- Grünenberg, K., Walker, H. K., & Knudsen, J. S. (2009). Mindfulness meditation. *Dansk Sociologi*, 20(2), 89–109.
- Grøn, R., & Balling, G. (2012). Litteraturformidling og bibliotekaren som faglig-personlig formidlingsautoritet. *Dansk Biblioteksforskning*, 1(3), 51–61.
- Gudiksen, J. (2008). Metodologiske aspekter ved studiet af brugeradfærd. *Dansk Biblioteksforskning*, 4(2), 57–70.
- Gundelach, P. (2001). National identitet i en globaliseringstid. *Dansk Sociologi*, 12(1), 63–80.
- Gundelach, P. (2009). Sekulariseringens mange ansigter. *Dansk Sociologi*, 20(1), 25–44.
- Gundelach, P., & Nørregård-Nielsen, E. C. (2002). Hvornår er man ung? *Dansk Sociologi*, 13(3), 27–46.

- Gøtze, M., & Bergqvist, C. (2014). Når myndigheder kommer på besøg. *Juristen*, 2014(6), 221–230.
- Hammershøj, L. G. (2001). Selvdannelse og nye former for socialitet - technofesten som eksempel. *Dansk Sociologi*, 12(2), 23–44.
- Hansen, J. A., & Hammerslev, O. (2010). Bourdieu og staten. *Praktiske Grunde. Tidsskrift for Kultur- Og Samfundsvidenskab*, (1-2), 11–33.
- Hansen, J. A., & Hansen, N. W. (2009). Polakker på det danske arbejdsmarked - frie fugle eller ny randgruppe? *Tidsskrift for Arbejdsliv*, 11(2), 24–40.
- Hansen, S. J. (2011). Om fisk, familie og færdigheder. *Praktiske Grunde. Tidsskrift for Kultur- Og Samfundsvidenskab*, 2-3, 77–86.
- Hariri, J. G. (2012). Kausal inferens i statskundskaben. *Politica*, 44(2), 184–201.
- Hariri, J. G. (2014). Statskundskabens sammenfildrede virkelighed og et bud på en løsning: IV-estimation. *Politica - Tidsskrift for Politisk Videnskab*, 46(1), 79–94.
- Hastrup, K. (2004). Videnskabelighed. *Tidsskrift for Forskning I Sygdom Og Samfund*, 1, 21–37.
- Heinskou, M. B. (2007). Seksualitet mellem risiko og chance. *Dansk Sociologi*, 1(18), 55–75.
- Henriksen, A., & Schack, M. (2013). Lovligheden af en humanitær intervention i Syrien. *Ugeskrift for Retsvæsen, Den Litterære Afdeling*, 147(44), 393–399.
- Herrmann, J. M. R. (2005). Det etiske forbehold. *Ugeskrift for Retsvæsen*, 385–393.
- Hjørland, B. (2010). Visioner for forskningsbiblioteker. *Dansk Biblioteksforskning*, 5(2/3), 21–35.
- Hjørland, B. (2011). Evidensbaseret praksis i videnskabsteoretisk belysning. *Dansk Biblioteksforskning*, 6(2/3), 35–47.
- Hjørland, B. (2011). Kompetencer i forskningsbibliotekerne i historisk og aktuel belysning. *Dansk Biblioteksforskning*, 7(1), 5–30.
- Hjørland, B. (2012). Om problemformuleringer i biblioteks-, dokumentations- og informationsvidenskab. *Dansk Biblioteksforskning*, 1(3), 63–74.
- Holm, A., & Jæger, M. (2004). Penge, (ud)dannelse, forbindelser eller brains? *Dansk Sociologi*, 15(3), 67–84.
- Holm, A., & Jæger, M. M. (2008). Livsformer i Danmark. *Dansk Sociologi*, 19(1), 31–53.
- Høst, J. E. (2014). Kvote og fisk på det frie marked. *Kulturstudier*, 5(2).
- Ilsøe, A. (2012). Den danske model til eksamen - et casestudie af CSC-konflikten. *Tidsskrift for Arbejdsliv*, 14(2), 31–48.
- Isager, J. M., & Olesen, M. N. (2013). Dannelse som basis for udvikling af moderne matematikundervisning. *Dansk Universitetspædagogisk Tidsskrift*, 8(15), 125–134.
- Jensen, A. M. B. (2008). Når døde børn er livets gave. *Kvinder, Køn Og Forskning*, 17(1-2), 51–59.
- Jensen, A. M. B. (2009). Mistede Liv og Nye Chancer. *Tidsskrift for Forskning I Sygdom Og Samfund*, 11, 31–50.



- Jensen, C. S. (1999). Mod en europæisk velfærds- og arbejdsmarkedsmodel. *Dansk Sociologi*, 10(2), 37–55.
- Jensen, C. S. (2004). Faglig organisering under forandring. *Tidsskrift for Arbejdsliv*, 6(3), 7–25.
- Jensen, C. S. (2010). Flexicurity og de danske arbejdsmarkedsrelationer. *Dansk Sociologi*, (1), 27–53.
- Jensen, C. S. (2012). Arbejdsgivere i Danmark og Storbritannien. *Tidsskrift for Arbejdsliv*, 14(1), 41–55.
- Jensen, H. N., & Juul Jensen, C. (2015). Specialestuderendes læringsudfordringer i vejledningen - nudging som handlemulighed. *Dansk Universitetspædagogisk Tidsskrift*, 10(18), 25–34.
- Jensen, H., & Jensen, H. N. (2011). Specialeskrivning på seks måneder. *Dansk Universitetspædagogisk Tidsskrift*, 6(10), 25–30.
- Jochumsen, H., Hvenegaard, C., & Skot-Hansen, D. (2013). Biblioteket som mødested – Sociologisk legitimitet og inspiration fra byplanlægningen. *Dansk Biblioteksforskning*, Årgang 2(1), 51–59.
- Johannsen, C. G. (2010). Folkebibliotekernes brugerbilleder. *Dansk Biblioteksforskning*, 5(1), 5–16.
- Johannsen, C. G. (2011). Kritik af evidensbevægelsen. *Dansk Biblioteksforskning*, 6(Nr.2/3), 17–33.
- Johansen, K. S., & Ludvigsen, K. L. B. (2011). Antropologiske evalueringer af misbrugsbehandling - viden om kompleksitet, relationer og hverdagspraksis. *Tidsskrift for Forskning I Sygdom Og Samfund*, (15), 77–96.
- Jørgensen, H. L., Larsen, B., Ingwersen, P., & Rehfeld, J. F. (2008). Forskningsaktiviteten for speciallæger i klinisk biokemi. *Ugeskrift for Læger*, 170(36), 2798–2802.
- Jørgensen, S. K. (2010). Interviewet og hvad så? En kritisk analyse af Asmaa Abdol-Hamids betingelser for selvfremsstilling. *Tidsskrift for Islamforskning*, 4(2), 77–103.
- Kann-Christensen, N. (2010). Institutionelle logikker i biblioteksvæsenet. *Dansk Biblioteksforskning*, 5(1), 17–28.
- Kann-Rasmussen, N., & Balling, G. (2015). Ikke-læsning som 'problem' i dansk kulturpolitik. *Nordisk Kulturpolitisk Tidsskrift*, 18(2), 250–266.
- Karpantschof, R. (2003). Højreradikalismen i Danmark. *Dansk Sociologi*, 14(3), 25–41.
- Karpantschof, R., & Mikkelsen, F. (2014). Petitioner og demokrati i Danmark i det 19. århundrede. *Arbetarhistoria*, 149-150(1-2), 18–24.
- Kirkegaard, B., & Borlund, P. (2006). Brugen af tv-udsendelser i Statens Mediesamling. *Dansk Biblioteksforskning*, 2(3), 15–27.
- Kirkegaard, B., & Borlund, P. (2008). Karakteristika ved brugernes informationsbehov i en tv-udsendelseskontekst. *Dansk Biblioteksforskning*, 4(3), 45–56.
- Knudsen, T. (2011). Den politiske djøficering. *Samfundsøkonomen*, (3), 37–41.
- Kongsted, H. C., & Konnerup, M. (2011). At måle en effekt. *Samfundsøkonomen*, (1), 11–17.
- Konzack, L. (2015). Meme Fame. *Nordisk Tidsskrift for Informationsvidenskab Og Kulturformidling*, 4(2), 45–52.

- Kosiara-Pedersen, K. (2014). Partimedlemmernes deltagelse og syn på partidemokrati 2000-2012. *Politica - Tidsskrift for Politisk Videnskab*, 46(3), 274–295.
- Kristensen, J., & Larsen, J. E. (2007). Fattigdom, social eksklusion og boligforhold. *Dansk Sociologi*, 18(4), 9–28.
- Kristiansson, M. (2010). Systematisk løbende refleksion. *Dansk Biblioteksforskning*, 5(1), 29–40.
- Kurrild-Klitgaard, P. (2011). Kontraktpolitik, kulturkamp og ideologi 2001-2011. *Økonomi and Politik*, 84(3), 47–62.
- Kurrild-Klitgaard, P. (2012). Frihed mellem fornuft og skepsis. *Kritik*, (206), 31–39.
- Larsen, A. G., & Ellersgaard, C. H. (2012). Status og integration på magtens felt for danske topdirektører. *Praktiske Grunde. Tidsskrift for Kultur- Og Samfundsvidenskab*, (2-3), 9–30.
- Larsen, C. H. (2014). Politisk vold som transnationalt fænomen i Danmark 1917-1939. *Arbetsarhistoria*, 1-2(149-150), 25–32.
- Larsen, T. B. (2011). Hvilket lands lov finder anvendelse, når ophavsrettigheder krænktes på Internettet? *N I R*, 3(2011), 222–229.
- Larsen, T. P. (2009). Tidsbegrænset ansatte. *Tidsskrift for Arbejdsliv*, 11(2), 56–72.
- Larsen, T. P. (2012). Køns betydning for lokalaftaler om ligeløn og familie-arbejdsliv. *Kvinder, Køn and Forskning*, 21(4), 7–17.
- Lassen, A. J. (2011). Insulin som tricket. *Dansk tidsskrift for kulturhistorie, etnologi, folkloristik og lokalhistorie*, 45.
- Lassen, A. J., & Jespersen, A. P. (2015). Ældres hverdagspraksisser i aldringspolitikken. *Tidsskriftet Kulturstudier*, 2015:1.
- Leilund, S. D. (2012). Det havde været mere rationelt straks at lægge pengene i skraldespanden.. *Kulturstudier*, (# 2).
- Levin, K. (2012). Kunstens Psykologi og Sansningens Primat. *Psyke Og Logos*, 33(2), 269–291.
- Liebst, L. S. (2009). Etikken i den maskerede by. *Dansk Sociologi*, 20(1), 7–23.
- Malchow-Møller, N., Munch, J. R., & Skaksen, J. R. (2009). Indvandringen til Danmark. *Søkelys På Arbejdslivet*, 26(1), 135–146.
- Martinsen, D. S. (2013). Danmarks implementering af EU-politik. *Politica*, 45(4), 437–456.
- Mogensen, H. O. (2007). Den dobbelte ensomhed og det omsorgsfulde system. *Tidsskrift for Forskning I Sygdom Og Samfund*, 7, 87–103.
- Mørck, Y. (2002). Multikulturalismernes kønsblinde øje. *Dansk Sociologi*, 13(3), 7–25.
- Nicolaisen, M. A. (2014). I kakaoens grænseland. *Kulturstudier*, November 2014(2), 120–138.
- Nicolaisen, M. S. (2014). Analog tekst i digital kontekst. *Nordisk Tidsskrift for Informationsvidenskab Og Kulturformidling*, 2(3), 41–53.
- Nielsen, A. S., & Mogensen, H. O. (2007). Introduktion: Kulturmøder i sundhedsvæsenet-migration, kategorisering og kompleksitet. *Tidsskrift for Forskning i Sygdom og Samfund*, 4(7).

- Nielsen, J. H. (2014). Eksperimenter og politologisk forskning. *Tidsskriftet Politik*, 17(3), 21–28.
- Nielsen, M. R. (2007). Partnervalgets grænse. *Dansk Sociologi*, 18(3), 25–47.
- Nielsen, N. J. (2013). Arbejderen mellem praksis og ideologisering 1850-2000. *Kulturstudier*, 4(1), 58–81.
- Nissen, M. (2014). Brugerdrevne standarder som konkret utopi. *Psyke and Logos*, 35(1), 164–192.
- Nordentoft, C. (2011). Evidens og kreativitet. *Dansk Biblioteksforskning*, 6(2/3), 49–63.
- Nørup Sørensen, C. T. (2015). Aggressiv under visse betingelser: en neoklassisk realistisk analyse af udviklingen i kinesisk udenrigs- og sikkerhedspolitik. *Politica*, 47(1), 66–83.
- Olesen-Bagneux, O. (2014). Da biblioteket blev til arkitektur. *Nordisk Tidsskrift for Informationsvidenskab Og Kulturformidling*, 2(2), 5–15.
- Olwig, K. F. (2002). Det etnografiske feltarbejde. *Norsk Antropologisk Tidsskrift*, 13(3), 111–123.
- Olwig, K. F. (2008). Integration mellem forestillede og glemte fællesskaber. *Norsk Antropologisk Tidsskrift*, 19(4), 233–242.
- Olwig, K. F. (2010). Migrationsnarrativer. *Kulturstudier*, 2(December), 119–139.
- Oxlund, B. (2012). At leve gennem tal. *Tidsskrift for Forskning I Sygdom Og Samfund*, 10(17), 101–118.
- Pedersen, F. T. (2003). Menneskerettigheder, afgudsdyrkelse og international politik ifølge Michael Ignatieff. *Slagmark*, (37), 155–168.
- Pedersen, F. T. (2006). En grænseløs ytringsfrihed i et grænseløst samfund? *Dansk Sociologi*, 17(2), 73–81.
- Pedersen, I. K. (1999). Bagom talentbegrebet. *Dansk Sociologi*, 10(1), 7–19.
- Pedersen, I. K. (2009). I grænselandet mellem optimering og helbredelse. *Tidsskrift for Forskning I Sygdom Og Samfund*, (11), 87–103.
- Plambech, S. (2005). "Postordrebrude" i Nordvestjylland. *Dansk Sociologi*, 16(1), 92–110.
- Poder, P. (2010). Når medarbejdere håndterer hinandens følelser. *Tidsskrift for Arbejdsliv*, 12(3), 72–86.
- Popp-Madsen, B. A. (2014). Debatterende eller besluttende offentlighed? *Slagmark*, 69, 35–51.
- Pors, N. O. (2008). Personlighed og Informationsadfærd. *Dansk Biblioteksforskning*, 4(2), 5–16.
- Pors, N. O. (2011). Evidens om bibliotekernes brugere. *Dansk Biblioteksforskning*, 6(2/3), 65–81.
- Rasmussen, C. H. (2015). Brugerinddragelse og kulturpolitisk kvalitet. *Nordisk Kulturpolitisk Tidsskrift*, 18(1), 76–95.
- Social Science*
- Rasmussen, M. B. (2014). Statens nye tilstedeværelse? *Norsk Antropologisk Tidsskrift*, 25(2), 94–107.

- Rasmussen, M. B. (2014). Statens nye tilstedeværelse? Norsk Antropologisk Tidsskrift, 25(2), 94–107.
- Ravnbøl, C. I. (2015). Hjem og arbejde for hjemløse og arbejdsløse EU medborgere. Socialmedicinsk Tidsskrift, 92(3), 326–334.
- Rostbøll, C. F. (2013). Social retfærdighed og økonomiske incitament. Politik, 16(2), 23–30.
- Rubow, C., & Tind Johannessen-Henry, C. (2010). Variationer af liv i døden. Tidsskrift for Forskning I Sygdom Og Samfund, 12, 135–153.
- Røgilds, F. (2002). Den nye racisme. Dansk Sociologi, 13(3), 101–111.
- Rønsholdt, S. (2008). Om god forvaltningsskik. Ugeskrift for Retsvæsen, B, 242–248.
- Raahauge, K. M. (2007). Ved vejen – I komplekset. Dansk Sociologi, 18(4), 31–49.
- Salamon, K. L. (2015). Forretningslivet som etnografisk felt: en praktisk videnskab. Jordens Folk, 50(1/2), 78–85.
- Schreiber, T. (2011). Informationskompetence: Set ud fra to forskellige praksisteorier. Dansk Biblioteksforskning, 7(2/3), 95–107.
- Simonsen, D. G., & Mordhorst, C. (2012). Skrald! Kulturstudier, (2), 6–13.
- Simonsen, M. M. (2008). Partsinddragelsen i beskæftigelsespolitiske reformer under VK-regeringen. Tidsskrift for Arbejdsliv, 10(3), 39–56.
- Skouvig, L. H. C. (2014). Efter indhentet efterretning. Nordisk Tidsskrift for Informationsvidenskab Og Kulturformidling, 3(2/3), 7–17.
- Søe, S. O. (2014). Information, misinformation og disinformation. Nordisk Tidsskrift for Informationsvidenskab Og Kulturformidling, 3(1), 21–30.
- Sørensen, P. B. (2010). Nedturen i dansk økonomi. Samfundsøkonomen, 2010(1), 29–34.
- Villadsen, K. F. (2002). Michael Foucault og liberalisme kritik. Dansk Sociologi, 13(3), 77–97.
- Wøldike, M. E. (2007). Kvinders smag for mænd, mænds smag for kvinder. Kvinder, Køn Og Forskning, (4), 9–20.

## Appendix B. Texts in the Second Academic Corpus arranged according to academic discipline

### *Health Science*

Bonde, H. (2015). Jernhårde ladies? Idrottsforum.org, 1-27.

Sandahl, C., & Winther, H. (2016). Familieidræt: Bevægelse, udvikling og livgivende øjeblikke for store og små. Idrottsforum.org, 1-13.

Vinther, S. B., & Thing, L. F. (2015). Krop og træning i Kilo Killers. Idrottsforum.org, 1-21.

### *Humanities*

Daugaard, L. M., Jensen, N. H., Kristensen, K. S., Laursen, H., Slåttvik, A. B., & Wolf, G. (2016). *Flersprogethed i dagtilbud og skole. Lingvistisk etnografiske analyser af sprogpædagogisk praksis*. Københavnerstudier i tosprogethed, Københavns Universitet, Humanistisk Fakultet, bind 73.

Hultgren, A. K. (2013). *Parallelsproglighed på danske universiteter: en status rapport 2013*. Københavnerstudier i tosprogethed, Studier i parallelsproglighed, Københavns Universitet, Humanistisk Fakultet, bind C5.

Hvas, S. V. *Når psykiateren taler dansk som andetsprog. En konversationsanalyse af forståelsesproblemer og identitet i psykiatriske samtaler med L2-talende psykiatere*. Københavnerstudier i tosprogethed, Københavns Universitet, Humanistisk Fakultet, bind 74.

Jakobsen, A. S. (2010). *"Ellers er det lige ud af landevejen": En interviewundersøgelse af ti underviseres holdninger til og erfaringer med englesksproget undervisning ved Det Biovidenskabelige Fakultet, KU*. Københavnerstudier i tosprogethed, Studier i parallelsproglighed, Københavns Universitet, Humanistisk Fakultet, bind C2.

Kappelgaard, S. B., & Hjorth, H. B. (2017). *Det stærkeste køn. En undersøgelse af genusrealisering i dansk blandt teenagere i flersprogede miljøer i Køge og på Amager*. Københavnerstudier i tosprogethed, Københavns Universitet, Humanistisk Fakultet, bind 75.

Kirilova, M., & Holmen, A. (Eds.) (2016). *Kulturlæring*. Københavnerstudier i tosprogethed, Studier i parallelsproglighed, Københavns Universitet, Humanistisk Fakultet, bind C9.

Laursen, K. Å. (2013). *"Det er sprogligt-selv hvor du ikke lægger mærke til det": en empirisk undersøgelse af de sproglige og faglige vanskeligheder hos farmaceutstuderende med dansk som andetsprog på Københavns Unviersitet*. Københavnerstudier i tosprogethed, Studier i parallelsproglighed, Københavns Universitet, Humanistisk Fakultet, bind C4.

### *Natural Science*

Boklund, A., Hisham Beshara Halasa, T., Struve, T., Østergaard, J., Clausen, J., & Chriél, M. (2015). Simulering af kontrolforanstaltninger til bekæmpelse af plasmacytose i minkfarme. *Dansk Veterinaertidsskrift*, (6), 24-30.

### *Social Science*

Dinesen, P. T., & Sønderskov, K. M. (2012). Hvorfor stiger tilliden? *Politica - Tidsskrift for Politisk Videnskab*, 44(1), 87-110.

Hjelmar, U., Møller, A. M., & Graulund, A. S. (2015). Læring fra en evaluering af Vidensportalen - fra forskningsbaseret viden til praktisk handlen? *Ceptra-Striben*, (17), 46-53.

Kristensen, R. A. (2008). At være eller ikke være diabetiker? *Tidsskrift for Forskning I Sygdom Og Samfund*, (9), 53-70.

Steffen, V., & Andersen, S. L. (2013). Sygdom, normalitet og egenomsorg. *Tidsskrift for Forskning I Sygdom Og Samfund*, 2013(19), 121-140.

## Appendix C. Texts in the General Language Corpus arranged according to type

Title	Source	Type	Author(s)	Publication date
Big Data i sundhed rammer de svageste	Politiken.dk	Feature article	Troels Krarup Hansen, Jens Winther Jensen, Kaj Grønbæk og Carsten Obel	2017-0518
Mor og pædagog: Tør du træde ind i dit barns digitale univers?	Politiken.dk	Feature article	Camilla B. Nielsen	2017-0502
Stop nu anglificeringen. Dansk er vores modersmål - lad os dog holde fast i det	Politiken.dk	Feature article	Flemming Juhl	2017-0504
Drop industrigrisene og sats på frie, glade grise - så slipper man også for MRSA	Politiken.dk	Feature article	Johanne Gabel	2017-0502
Specialkonsulent: Fri os fra sundhedsplatform og offentlige it-systemer	Politiken.dk	Feature article	Helene Brochmann	2017-0515
Museer handler ikke bare om oplevelser og publikum. De skal bevare vores kulturarv	Politiken.dk	Feature article	Poul Bache	2017-0428
Danskerne forstår ikke at religion og videnskab hænger sammen	Politiken.dk	Feature article	Niels Kærgård	2017-0514
Undervisningsledere: Hvordan bliver de unge uddannet til fremtiden?	Politiken.dk	Feature article	René van Laer og Jesper Jans	2017-0516
København må ikke blive en lukket fest for de rige. Lighedskampen er i dag også en kamp på boligmarkedet	Politiken.dk	Feature article	Frank Jensen, Christian Grønnemark, Jan Trojaborg og Vibeke Westh	2017-0430
Nick Hækkerup: Centrum-Venstre har svigtet arbejderne og mellemklassen i Europa	Politiken.dk	Feature article	Nick Hækkerup	2017-0506
Gymnasier dyster i dragebåde	Fyens Stiftstidende	Newspaper article	Jesper Mads Eriksen	2016-0411
KL: Giv plejehjem faste læger	Fyens Stiftstidende	Newspaper article	Ritzau	2016-0411
Har Karrusel nogen værdi?	Fyens Stiftstidende	Newspaper article	Svend Conrad	2016-0913
Da Louise flyttede hjemmefra	Fyens Stiftstidende	Newspaper article	Rasmus Lundberg	2016-0913

Dansk stoledesign på den olympiske scene	Fyens Stiftstidende	Newspaper article	Miriam Kjer	2016- 0819
Mødrene holder fast -også efter et år	Fyens Stiftstidende	Newspaper article	Stefan Brix	2016- 0411
Når boblen brister i 2021	Fyens Stiftstidende	Newspaper article	Morten Skak	2016- 1218
Bornholm blev glemt	Fyens Stiftstidende	Newspaper article	Carl Otto Dethlefsen	2016- 0417
Flygtninge mangler boliger	Fyens Stiftstidende	Newspaper article	Henrik Juel Skovrider	2016- 0314
Skoleleder fyret efter kun et år	Fyens Stiftstidende	Newspaper article	Henrik Juel Skovrider	2016- 0217
Nekrolog	Fyens Stiftstidende	Newspaper article	[empty]	2016- 0519
Asylfrygt gjort til skamme	Fyens Stiftstidende	Newspaper article	Kasper Løvkvist	2016- 0615
Et sted for dem på vej videre	Fyens Stiftstidende	Newspaper article	Nikolaj Kennov Rasmussen	2016- 0615
Sidste chance for Hårslev	Fyens Stiftstidende	Newspaper article	Magnus Ørum Harkjær	2016- 1014
Topfolk gider ikke mere skrammel	Fyens Stiftstidende	Newspaper article	Karsten L. Sørensen	2016- 0514
På tide at tale om kvinders rettigheder	Fyens Stiftstidende	Newspaper article	Kiri Kim Lassen	2016- 0513
100 almennyttige boliger på vej	Fyens Stiftstidende	Newspaper article	Søren Gottwald	2016- 0116
Høj leje sender folk udenbys	Fyens Stiftstidende	Newspaper article	Søren Gottwald	2016- 0116
København lider under romalejre og tiggeri	Fyens Stiftstidende	Newspaper article	Jens Ejning	2016- 0911
Vores fjord også statens ansvar	Fyens Stiftstidende	Newspaper article	Hans Luunbjerg, Morten Andersen og Jane Jegind	2016- 0911
Skattereglerne gør forældre køb attraktive	Fyens Stiftstidende	Newspaper article	Lise Nytoft Bergmann	2016- 0918
Slipper for svamp men smides ud	Fyens Stiftstidende	Newspaper article	Torsten Cilleborg	2016- 0215
Ingen lun stue i kulden	Fyens Stiftstidende	Newspaper article	Thomas Juul	2016- 0113



Oplev Inkariget til fods	Fyens Stiftstidende	Newspaper article	[empty]	2016- 0612
Det bliver sjovere med et 20-årigt	Fyens Stiftstidende	Newspaper article	Lise Nytoft Bergmann	2016- 0612
Pas på det perfekte	Fyens Stiftstidende	Newspaper article	Susanne Crawley Larsen	2016- 0612
Fedes tomme liv får indhold	Fyens Stiftstidende	Newspaper article	Jesper Mads Eriksen	2016- 0218
Borgernes retssikkerhed ikke i orden	Fyens Stiftstidende	Newspaper article	[empty]	2016- 0711
Fattigdom. Nej tak til fattigdom i Danmark	Fyens Stiftstidende	Newspaper article	[empty]	2016- 1115
På Nordfyn findes næsten alt, hvad jeg har brug for	Fyens Stiftstidende	Newspaper article	[empty]	2016- 0210
De fynske boligsælgere er blevet mere optimistiske	Fyens Stiftstidende	Newspaper article	Lise Nytoft Bergmann	2016- 0410
Antallet af tvangsauktioner falder ikke længere	Fyens Stiftstidende	Newspaper article	Lise Nytoft Bergmann	2016- 0214
Produktionsskolen er espens springbræt	Fyens Stiftstidende	Newspaper article	Maja Korsgaard Svenningsen	2016- 1214
Fiskeskolen fanger flygtninge	Fyens Stiftstidende	Newspaper article	Nils Mogensen	2016- 1010
Et bidrag til en broget branche	Fyens Stiftstidende	Newspaper article	Thomas Bach-Laursen	2016- 0418
Fedtskeder og luksusfælden	Fyens Stiftstidende	Newspaper article	Peter Rasmussen	2016- 1119
Ægtepar kan ikke få for Energi Fyn fred	Fyens Stiftstidende	Newspaper article	Peter Ammitzbøll	2016- 1017
Kirsten har fået nok af kommunen	Fyens Stiftstidende	Newspaper article	Anne Vegeberg Hansen	2016- 0813
Fotografen, der er fascineret af fugle	Fyens Stiftstidende	Newspaper article	Anne Jørgensen	2016- 0814
Vi vil bare gerne sikre vores fremtid	Fyens Stiftstidende	Newspaper article	Helle Nordström	2016- 0212
"Mini-sygehuse" skal forbedre hjælpen til de svageste ældre	Fyens Stiftstidende	Newspaper article	Flemming Steen Pedersen	2016- 0412
Flis-huggeri giver støj og støv	Fyens Stiftstidende	Newspaper article	Henrik Mohr	2016- 0111

Fynboer vil forbedre boligen	Fyens Stiftstidende	Newspaper article	Brian Peterson	2016- 0619
Flygtninge flytter ind i baghaven	Fyens Stiftstidende	Newspaper article	Annette Bredmose	2016- 0619
I Middelfart får husejerne ikke glæde af finansloven	Fyens Stiftstidende	Newspaper article	Lise Nytoft Bergmann	2016- 0110
Stor huslejestigning på plejehjem	Fyens Stiftstidende	Newspaper article	René Johansen	2016- 1015
Boligen er den helt store pengesluger for pensionister	Fyens Stiftstidende	Newspaper article	Lise Nytoft Bergmann	2016- 1016
Kasper smider skovlen, når det brænder	Fyens Stiftstidende	Newspaper article	Ole Grube	2016- 1212
Kommentar. Sanktionering og flere regler får arbejdsløse i arbejde	Fyens Stiftstidende	Newspaper article	Christoffer Lilleholt	2016- 1111
Statslige flygtningelandsbyer i stedet for kommunale udfordringer	Fyens Stiftstidende	Newspaper article	[empty]	2016- 0114
I onsdags blev det forår	Fyens Stiftstidende	Newspaper article	Jesper Mads Eriksen	2016- 0318
Spørgsmålene hang i luften	Fyens Stiftstidende	Newspaper article	Simon Staun	2016- 0819
Specialskele lukker og slukker	Fyens Stiftstidende	Newspaper article	Helle Kryger	2016- 0917
Billige boliger til de trængende	Fyens Stiftstidende	Newspaper article	Ulrik Sass	2016- 0419
Terrorcelle i Belgien ville ramme Paris	Fyens Stiftstidende	Newspaper article	Kit Lindhardt	2016- 0411
Et velovervejede træk til Tåsinge	Fyens Stiftstidende	Newspaper article	Julie Ruby Bødiker	2016- 0411
Dansk crowdfunding buldrer frem	Fyens Stiftstidende	Newspaper article	Mikkel Walentin Mortensen	2016- 1218
Højtid. Skilsmisse-jul	Fyens Stiftstidende	Newspaper article	Charlotte Lind	2016- 1218
De Vestindiske Øer er parat til festår	Fyens Stiftstidende	Newspaper article	Vincent Byakika	2016- 1218
Nostalgien ødelagt af håbløs lyd	Fyens Stiftstidende	Newspaper article	Simon Staun	2016- 0417

IS-propagandamaskine skal knuses	Fyens Stiftstidende	Newspaper article	Christian Braad Petersen og Sine Schack Vestergaard	2016- 0417
Tyrkiet truer EU med åbne flygtningesluser	Fyens Stiftstidende	Newspaper article	Anders Aarkrog Jepsen	2016- 0514
Gode vine og gamle bygninger	Fyens Stiftstidende	Newspaper article	Torben Svane Christensen	2016- 0514
1300 motionister til halfest i Ørbæk	Fyens Stiftstidende	Newspaper article	Kasper Riggelsen	2016- 1113
Surmul ikke -gør noget	Fyens Stiftstidende	Newspaper article	Søren Thorup	2016- 1113
Birger har bygget broen	Fyens Stiftstidende	Newspaper article	Ole Grube	2016- 1113
Ingrid taler for de handicappede	Fyens Stiftstidende	Newspaper article	Nanna Bundgaard Bruun	2016- 1113
Her er det nye Carlslund	Fyens Stiftstidende	Newspaper article	Karsten Hüttel	2016- 0211
Sukker fortsat efter ferieture	Fyens Stiftstidende	Newspaper article	Daniel Kofoed	2016- 0211
Storgartner har betalt millioner i overpris for el	Fyens Stiftstidende	Newspaper article	Jesper Beenfeldt Nielsen	2016- 0519
National-egoisme	Fyens Stiftstidende	Newspaper article	Flemming Kjersgaard Johansen	2016- 0519
Merkel tillader sag mod komiker	Fyens Stiftstidende	Newspaper article	Troels Heeger	2016- 0416
Se, nu stiger solen	Fyens Stiftstidende	Newspaper article	[empty]	2016- 0513
Nekrolog	Fyens Stiftstidende	Newspaper article	Ritzau	2016- 0513
Gentleman-sportens oase	Fyens Stiftstidende	Newspaper article	niels Abildtrup	2016- 1116
Alt for mange af mine venner er døde	Fyens Stiftstidende	Newspaper article	Ulrik sass	2016- 1116
S og V: Vi har brug for politiets afhøringer i Holst-sagen	Fyens Stiftstidende	Newspaper article	Jonas Ancher Nyeng	2016- 1215
Borgmester er ærgerlig	Fyens Stiftstidende	Newspaper article	Martin Kloster	2016- 1215

Røde Kors: Det er ikke brødnid	Fyens Stiftstidende	Newspaper article	Bent Warneke	2016- 0116
Iran -her kan du færdes trygt	Fyens Stiftstidende	Newspaper article	Charlotte Pedersen og Jakob Haugaard Christiansen	2016- 0116
Gustav vil score kassen som tjener på Oslobåden	Fyens Stiftstidende	Newspaper article	Bent Warneke	2016- 0618
Kronborg styrer pryglet GOG	Fyens Stiftstidende	Newspaper article	Karsten L. Sørensen	2016- 0918
Advokat: Der skal meget til for at dømmes for injurier	Fyens Stiftstidende	Newspaper article	Nanna Bundgaard Bruun	2016- 0911
Mordersken	Fyens Stiftstidende	Newspaper article	Torben Hangaard	2016- 0911
Terror ramte Tyrkiets turistcentrum	Fyens Stiftstidende	Newspaper article	Finn Jørgensen	2016- 0113
Søges: Kvinder af den rette støbning	Fyens Stiftstidende	Newspaper article	Kiri Kim Lassen	2016- 0812
Sidste swing for pinsejazz på Strynø	Fyens Stiftstidende	Newspaper article	Alexander Dornwirth	2016- 0516
Kong Lears arvesag	Fyens Stiftstidende	Newspaper article	Jens Kristian Elvstrøm	2016- 0516
Vintersul til Romsøs bukke	Fyens Stiftstidende	Newspaper article	Helle Kryger Hynke	2016- 0218
DF'S formand gik ikke frivilligt	Fyens Stiftstidende	Newspaper article	Lise Wolf	2016- 0313
På besøg hos Agatha Christie	Fyens Stiftstidende	Newspaper article	Silja Smith	2016- 0710
Pionererne faldt straks til på Kastanievej	Fyens Stiftstidende	Newspaper article	Palle Søby	2016- 1219
Mindeord	Fyens Stiftstidende	Newspaper article	Helene og Henrik Vorager	2016- 1219
Uheldig ledelse	Fyens Stiftstidende	Newspaper article	Finn Wiedemann	2016- 0210
God PR for odense i Japan	Fyens Stiftstidende	Newspaper article	Jesper Mads Eriksen	2016- 0210
En fornemmelse af tomhed indeni	Fyens Stiftstidende	Newspaper article	[empty]	2016- 0210

En glad terrier fra Nigeria	Fyens Stiftstidende	Newspaper article	Leif Rasmussen	2016- 0210
Vilks tilbage i København	Fyens Stiftstidende	Newspaper article	Jens Holt	2016- 0214
En afrikansk odysseé	Fyens Stiftstidende	Newspaper article	Lars U. Thomsen	2016- 0214
Slut med kæmpekontrakt til ét firma	Fyens Stiftstidende	Newspaper article	Charlotte Pedersen	2016- 0613
Rugbrød vinder over romkuglen til hverdag	Jyllands- Posten	Newspaper article	Katrine Stampe	2017- 0324
Keep calm and carry on	Jyllands- Posten	Newspaper article	Mads Bonde Broberg	2017- 0324
Rugbrødets og toilettets historie er absolut også relevante emner for os	Jyllands- Posten	Newspaper article	Peter Schollert	2017- 0324
Var angrebet i London forløberen for en endnu større massakre?	Jyllands- Posten	Newspaper article	Solveig Gram Jensen	2017- 0324
Jeg sagde mit job op, fordi jeg havde et ubehag ved arbejdslivet generelt	Jyllands- Posten	Newspaper article	Naja Dandanell	2017- 0222
Politisk medvind til at lade domstole afgøre skilsmisser	Jyllands- Posten	Newspaper article	Hanne Fall Nielsen	2017- 0222
Kronik: KRONIK: Filosofi med børn er svaret	Jyllands- Posten	Newspaper article	Lærke Groth og Dorete Kallesøe	2017- 0222
Min privatøkonomi? Den rager ikke vælgerne	Jyllands- Posten	Newspaper article	Christian Sehested Rasmus Bendtsen	2017- 0222
Børn i Myanmar flygter alene	Jyllands- Posten	Newspaper article	Muneeza Naqvi	2017- 0929
En hyldest til farverne i alle husets rum	Jyllands- Posten	Newspaper article	Mettemaje Skøtt	2017- 0820
Barcelona slikker sine sår på stranden, på hospitalerne og på skrift	Jyllands- Posten	Newspaper article	Martin Kaae	2017- 0820
Ansigtet og stemmen har han fra sin far, snakketøjet er mors	Jyllands- Posten	Newspaper article	Peter Schollert	2017- 0820
AKTUELT PORTRÆT: " Sikke et cirkus" er spækket med Ramasjang-stjerner.	Jyllands- Posten	Newspaper article	Sofie Ejdrup Larsen	2017- 0929
Den seksuelle revolutions formand Mao er død	Jyllands- Posten	Newspaper article	David Jacobsen Turner	2017- 0929
Trappetendenser	Jyllands- Posten	Newspaper article	Cathrine Errboe	2017- 0820

I København kan man stadig se sporene af Kulturbry 96: Hvilke aftryk kommer Aarhus til at sætte?	Jyllands-Posten	Newspaper article	Lars Dalsgaard	2017-0820
To ministre drog til Afrika - men tusinder vil hellere den modsatte vej	Jyllands-Posten	Newspaper article	Jesper Hvass	2017-0820
USA's stillehavsø Guam holder vejret	Jyllands-Posten	Newspaper article	Jørgen Ullerup	2017-0820
Livet venter dig i rigt mål	Jyllands-Posten	Newspaper article	Torkil Rasmussen	2017-0624
Udlandsdanskere og pensionistreglen	Jyllands-Posten	Newspaper article	Tommy V. Christiansen	2017-0624
Kritikere: Christiansborg spiller hasard med dansk fjernvarme	Jyllands-Posten	Newspaper article	Michael Stenvei	2017-0324
Forældet beredskab overser biologiske trusler	Jyllands-Posten	Newspaper article	Erik Heegaard, Jesper Nielsen og Mikael Rørdam Andersen	2017-0324
Når klimaet samler familien	Jyllands-Posten	Newspaper article	Thomas Bech Hansen	2017-0324
Airbnb vil tjene penge, hvor Uber bed i asfalten	Jyllands-Posten	Newspaper article	Thomas Høy Davidsen	2017-0324
Intet kan stoppe en gal mand	Jyllands-Posten	Newspaper article	Jesper Kongstad Michael Thykier	2017-0324
Muslimer i Marine Le Pens højborg er bekymrede	Jyllands-Posten	Newspaper article	Marie Louise Albers	2017-0222
Spillerforeningen vil jage midler til homofobi-kampagne	Jyllands-Posten	Newspaper article	Emma Gelbjerg-Hansen	2017-0222
I det udskældte bederum sætter muslimer sig i massagestolen	Jyllands-Posten	Newspaper article	Anders Leonhard	2017-0222
De islandske "viking raiders" har fået over 75 års fængsel - men nu bobler økonomien igen	Jyllands-Posten	Newspaper article	Lone Andersen	2017-0222
Oliepriserne truer den økonomiske vækst	Jyllands-Posten	Newspaper article	Keld Louie Pedersen	2017-0929
INDBLIK: P-vagter ska fier penge i kassen - og hjælper trafikanterne	Jyllands-Posten	Newspaper article	Lars From Klaus Dohm	2017-0929
Udfordringen er at lave et så stort show, hvor folk stadig bliver berørt	Jyllands-Posten	Newspaper article	Nana Elving Hansen	2017-0820
Atomkrise kan blusse op igen mandag	Jyllands-Posten	Newspaper article	Jørgen Ullerup	2017-0820

Ligamålmænd raser mod DHF: Vi frygter for vores helbred	Jyllands-Posten	Newspaper article	Kenneth Thygesen	2017-0820
Bombarder hovedkvarteret!	Jyllands-Posten	Newspaper article	Niels Lillelund	2017-0820
Den hullede historie om Kaptajn Nemo i Køge Bugt	Jyllands-Posten	Newspaper article	Morten Vestergaard	2017-0820
INDBLIK: Vi har ikke set så kraftig en manifestation siden Black Army	Jyllands-Posten	Newspaper article	Jacob Haislund	2017-0624
Karriere bygget på bæredygtighed	Jyllands-Posten	Newspaper article	Simon Kudal	2017-0624
Vi har nok været lidt for selvfede	Jyllands-Posten	Newspaper article	René Deichgræber	2017-0624
El-kongen sendte firmabilen til rens med lastbil. Nu er han fyret for dårlig dømmekraft	Jyllands-Posten	Newspaper article	Jakob Martini Kasper Brøndgaard	2017-0624
Kulturelt krydstogt	Jyllands-Posten	Newspaper article	Thomas Linder Kamure Thomsen	2017-0624
De fandt drømmejobbet	Jyllands-Posten	Newspaper article	Simon Kudal	2017-0520
INDBLIK: Arbejdsgiverne vil udskyde pensionen, men murerne frygter at være slidt op til den tid	Jyllands-Posten	Newspaper article	Anders Leonhard Marchen Neel Gjertsen	2017-0520
Flere sygemeldinger i udflyttede styrelser	Jyllands-Posten	Newspaper article	Simon Kudal	2017-0520
Højstemt barok i Bøhmen	Jyllands-Posten	Newspaper article	Rasmus L. Ottesen	2017-0520
Her har ingen mellemledere...	Jyllands-Posten	Newspaper article	Anders Mørkbak Bruun	2017-0520
PÅ TVÆRS: Så sku' den skat være skrinlagt	Jyllands-Posten	Newspaper article	Lone Andersen	2017-0520
Jeg blev fyret, men er ikke færdig	Jyllands-Posten	Newspaper article	Ove Klausen	2017-0520
Klar kommunikation fra lederen	Jyllands-Posten	Newspaper article	Simon Kudal	2017-0520
Kreditværdighed får et hak i tuden	Jyllands-Posten	Newspaper article	Thomas Høy Davidsen	2017-0525
Det skal ikke være en ren mormor cabaret	Berlingske	Newspaper article	Søren Frank	2017-0608

Et splittet Storbritannien går til valg i dag	Berlingske	Newspaper article	Uffe Taudal	2017-0608
Politifolk blev inviteret på safari af Karsten Ree: Det har ikke en skid med bestikkelse at gøre. Vi er jo bare venner	Berlingske	Newspaper article	David Rue Honoré	2017-0608
Mange i ejendomsbranchen er uforberedt på beskyttelse af persondata	Berlingske	Newspaper article	Sten Thorup Kristensen	2017-0607
Børneforskningsprojekt: Det er spild af penge	Berlingske	Newspaper article	Nathalie Ostrynski	2017-0807
Islamisk Stats selvmordsbrigade på vej til EU	Berlingske	Newspaper article	Allan Sørensen	2017-0807
De orange løvinder er Europas bedste	Berlingske	Newspaper article	Jens Anton Bjørnager Jensen	2017-0807
Test	Berlingske	Newspaper article	Lise Penter	2017-0807
Tiden med ufaglært børnepasning rinder ud	Berlingske	Newspaper article	Ane Halsboe-Jørgensen	2017-0807
Erfarne folk uden EU-pas søges: Politikere, hjælp os!	Berlingske	Newspaper article	Klaus Nyengaard	2017-1004
P3 får ny profil: Ironien skal væk, kanalen skal være klogere	Berlingske	Newspaper article	Sarah Iben Almbjerg	2017-1004
Så har statsministeren åbnet Folketinget: Her er fem centrale punkter fra talen	Berlingske	Newspaper article	Jens Beck Nielsen	2017-1004
DTU: Udenlandske studerende er en gevinst for Danmark	Berlingske	Newspaper article	Jens Ejning og Christoffer Schyth Kjær	2017-1004
Efter blodbadet og fundet af 42 våben: Giftigt opgør med Trump om våbenloven	Berlingske	Newspaper article	Michael Bjerre	2017-1004
Succesfuld BIG-topchef: Jeg har været lige så usikker som alle andre unge piger	Berlingske	Newspaper article	Ditte Vinterberg Weng	2017-1004
Systemet har svigtet: Kommunalvalget handler om vores børn, brudte aftaler og ligestilling	Berlingske	Newspaper article	Julia Lahme	2017-1004
Hvis du bliver gravid, tror branchen, at det varer ti år	Berlingske	Newspaper article	Sarah Iben Almbjerg	2017-0904
Her rynker vi ikke på næsen ad musicals	Berlingske	Newspaper article	Trine Munk-Petersen	2017-0904
Studerende stresses: 7 i snit kræver 55 timers studier om ugen	Berlingske	Newspaper article	Jens Ejning	2017-0904



Blondinen som fortalte sandheden	Berlingske	Newspaper article	Poul Høi	2017-0904
Hvordan undgår jeg at blive dansk?	Berlingske	Newspaper article	Knut Skjærven	2017-0302
Jeg er nødt til at føle mig som nybegynder	Berlingske	Newspaper article	Louise Kidde Sauntved	2017-0302
Sidste nat med kliken fra noma	Berlingske	Newspaper article	Søren Frank Madredaktør	2017-0302
Staten lander Nordsøaftale med Mærsk: Her er detaljerne	Berlingske	Newspaper article	Lasse Friis Jesper Kongskov og Peter Suppli Benson	2017-0302
Advokat tav om private handler med byggematadors milliardfond	Berlingske	Newspaper article	Birgitte Erhardtsen og Anders Sebastian Kauffeldt	2017-0302
Mød manden, du gerne vil have til at investere i din virksomhed	Berlingske	Newspaper article	Niels P. York	2017-0302
DRs seneste selvmål	Berlingske	Newspaper article	Thomas Larsen	2017-0806
Alle talte ned til EM-finalen til det sølvbryllup, jeg var til i går	Berlingske	Newspaper article	Jens Anton Bjørnager og Sofie Mathiassen	2017-0806
Vi børnefamilier tager jo heller ikke i Fårup Sommerland, før vi har betalt huslejen, vel?	Berlingske	Newspaper article	Kasper Kildegaard	2017-0806
Malenes minder	Berlingske	Newspaper article	Maise Njor	2017-0806
Han var hele verdens kærlige far - men var han også sexkrænker?	Berlingske	Newspaper article	Michael Bjerre	2017-0608
Københavns uønskeede flaskesamlere	Berlingske	Newspaper article	Pernille Dreyer	2017-0607
Er valget allerede torsdag? Åh nej	Berlingske	Newspaper article	Uffe Taudal	2017-0607
Nu tager libyerne også gummibåden til Europa	Berlingske	Newspaper article	Ole Damkjær	2017-0607
Kronik: Tiden er løbet fra FNs Flygtningekonvention	Berlingske	Newspaper article	Poul Christian Matthiessen	2017-0905
Alting smager bedre efter et lille eventyr	Berlingske	Newspaper article	Sarah Skarum	2017-0430

En stråmand taler ud: Jeg var direktør i 250 selskaber	Berlingske	Newspaper article	Simon Bendtsen, Michael Lund og Eva Jung	2017-0430
Udenlandske kvinder bliver i voldelige ægteskaber - frygter at blive sendt hjem	Berlingske	Newspaper article	Mikkel Walentin og Malthe Sommerand	2017-0730
I 2013: Han, hun og håbet	Berlingske	Newspaper article	Christina Hilstrøm	2017-0507
Feminismen var bedre i gamle dage	Berlingske	Newspaper article	Majbritt Maria Nielsen	2017-0306
HPVvaccinen er effektiv. Den forebygger livmoderhalskræft. Og den er lige så sikker som andre vacciner	Berlingske	Newspaper article	Søren Brostrøm	2017-0509
Jeg blev meget hurtigt castet som den helt store syndebuk	Berlingske	Newspaper article	Nathalie Ostrynski	2017-0805
Manden der gerne ville være kong Henrik	Berlingske	Newspaper article	Nathalie Ostrynski	2017-0805
Man bliver nervøs, når man hører stiletterne komme...	Berlingske	Newspaper article	Maise Njor	2017-0802
Ligestilling i kongehuset: Kong Margrethe og dronning Henrik	Berlingske	Newspaper article	Hans Christian Bjerg	2017-0808
Costa del Sharia	Berlingske	Newspaper article	Allan Sørensen	2017-0808
Indvandringen fører til brutale overgreb	Berlingske	Newspaper article	Kasper Støvring	2017-0830
God fiktion har altid ret	Berlingske	Newspaper article	Søren Kassebeer	2017-0830
Debat: Hvad ved mænd om kvinders frygt?	Berlingske	Newspaper article	Micha Fuglede	2017-0909
Den uslukkkelige tørst efter fart	Berlingske	Newspaper article	Ulrik Andersen	2017-0207
NEKROLOG: Zygmunt Bauman var et intellektuelt fyrtårn - og et ydmygt menneske	Kristeligt Dagblad	Newspaper article	Björg Tulinius	2017-0111
"Jeg kunne godt tænke mig at få en familie"	Kristeligt Dagblad	Newspaper article	Susanne Utzon	2017-0518
I London er tro en integreret del af hverdagslivet	Kristeligt Dagblad	Newspaper article	Dorte J. Thorsen	2017-0515
Cubanerne holder vejret	Kristeligt Dagblad	Newspaper article	Anne Lea Landsted	2017-0515

Singapores skyggesider	Kristeligt Dagblad	Newspaper article	Sofie Buch Hoyer	2017-0410
Krimtatareernes leder: Rusland prøver at presse os til at blive russere - eller forsvinde fra Krim	Kristeligt Dagblad	Newspaper article	Ota Tiefenböck	2017-0317
Verden oplever færre selvmord	Kristeligt Dagblad	Newspaper article	Maja Funch	2017-0317
Forsidehenvisning: Uden gården og grisene er livsværket og identiteten væk	Kristeligt Dagblad	Newspaper article	Eva Emborg Bejder	2017-0317
Jeg var angst for at svigte min familie	Kristeligt Dagblad	Newspaper article	Eva Emborg Bejder	2017-0317
Som vågekone er man trænet i nærværet	Kristeligt Dagblad	Newspaper article	Stephanie Hollender	2017-0317
Inderst inde ønsker jeg, at min stedfar adopterer mig	Kristeligt Dagblad	Newspaper article	Kære brevkasse	2017-0317
Adoptivforældre skal rumme deres børns drømme om den biologiske familie	Kristeligt Dagblad	Newspaper article	Britta Søndergaard	2017-0714
"Folkekirken giver mine krimier klangbund"	Kristeligt Dagblad	Newspaper article	Daniel Øhrstrøm	2017-0216
Tvangsfri psykiatri giver nye etiske dilemmaer	Kristeligt Dagblad	Newspaper article	Maja Funch	2017-0316
Ventetid skubber asylansøgere væk fra integration	Kristeligt Dagblad	Newspaper article	Simon Skou og Christian Klein	2017-0314
Forsidehenvisning: Både røde og blå politikere har brug for borgernes tillid til Skat	Kristeligt Dagblad	Newspaper article	Henrik Hoffmann-Hansen	2017-0614
Japans kejserfamilie risikerer at uddø	Kristeligt Dagblad	Newspaper article	Asger Røjle	2017-0614
Debat: Aspergerautist og folketingskandidat: De kognitivt handicappede blev glemt på perronen	Kristeligt Dagblad	Newspaper article	Niels Christiansen	2017-0614
Sommerferie med et hjertesuk	Kristeligt Dagblad	Newspaper article	Kære brevkasse	2017-0818
Nu vil hele verden lære at hygge sig	Kristeligt Dagblad	Newspaper article	Tine Maria Winther	2017-0818
Vreden mod USA lever – på mølædt museum i Teheran	Politiken.dk	Newspaper article	Anders Jerichow	2017-0519
Professor i miljømedicin: Drop amning efter tre-fire måneder	Politiken.dk	Newspaper article	Lars Igum Rasmussen	2015-0821
KL-topfolk: Vi advarede Fogh og Co. om skattekaos	Politiken.dk	Newspaper article	Kristian Klarskov og Anders Bæksgaard	2017-0518

Terrorsigtet Kundby-pige troede, at hun skrev med Islamisk Stats leder	Politiken.dk	Newspaper article	Kristian Corfixen	2017-0419
Vreden mod USA lever – på mølædt museum i Teheran	Politiken.dk	Newspaper article	Anders Jerichow	2017-0519
Professor i miljømedicin: Drop amning efter tre-fire måneder	Politiken.dk	Newspaper article	Lars Igum Rasmussen	2015-0821
KL-topfolk: Vi advarede Fogh og Co. om skattekaos	Politiken.dk	Newspaper article	Kristian Klarskov og Anders Bæksgaard	2017-0518
Terrorsigtet Kundby-pige troede, at hun skrev med Islamisk Stats leder	Politiken.dk	Newspaper article	Kristian Corfixen	2017-0419
I gymnastik var eleverne for langsomme til at klæde om, så Keld måtte love dem flødeboller, hvis de kunne klare omklædningen på 8 minutter	Politiken.dk	Newspaper article	Mette Dalgaard	2017-0516
Udstilling i Den Danske Pavillon er nogle steder lige så dyb som en Disneyfilm	Politiken.dk	Newspaper article	Mathias Kryger	2017-0518
Tag en pause med Peter Høeg	Politiken.dk	Newspaper article	Carsten Andersen	2015-1012
Mustafa al-Saadi: »Jeg fik at vide, at vi skulle på bilferie med min onkel i Europa i et par uger, og det ville jeg gerne. Men det blev til 11 år i Bagdad«	Politiken.dk	Newspaper article	Christian E. Holm og Elisabeth Yskes	2017-0412
Sportsblik: Brian Holms mund er lukket med syv højtalere	Politiken.dk	Newspaper article	Rasmus Bech	2017-0518
Når 4.c. dyrker grønt får de mere smag for mad	Politiken.dk	Newspaper article	Mikkel Bækgaard	2017-0516
Dansk for vidensarbejdere - Del 1	CIP, University of Copenhagen	Teaching material	Karen-Margrete Frederiksen	2017
Dansk for vidensarbejdere - Del 2	CIP, University of Copenhagen	Teaching material	Karen-Margrete Frederiksen	2017
Dansk for vidensarbejdere - Del 3	CIP, University of Copenhagen	Teaching material	Karen-Margrete Frederiksen	2017
Dansk for vidensarbejdere - Del 4	CIP, University of	Teaching material	Karen-Margrete Frederiksen	2017

	Copenhagen			
Dansk for vidensarbejdere - Del 5	CIP, University of Copenhagen	Teaching material	Karen-Margrete Frederiksen	2017
Puk Elgård voksede op i et misbrugshjem: "Jeg vidste aldrig, hvad jeg kom hjem til, og hvilken tilstand mine forældre var i"	NYT - www.alt.dk	Weekly magazine	Michala Rosendah	2017- 0516
Præst Kathrine Lilleør: "Du bør gøre dig nogle nogle tanker og ønsker om din egen død. Det vil hjælpe dine efterladte"	NYT - www.alt.dk	Weekly magazine	Marie-Louise Truelsen	2017- 0517
Derfor bruger vi Facebook	NYT - www.alt.dk	Weekly magazine	Mi Skjold Bri	2013- 0826
Sarah Grünwald har aldrig været bange for at miste en ekskæreste: "Jeg har altid følt, at jeg var mere værd end dem"	NYT - www.alt.dk	Weekly magazine	Cille Lewinsky	2017- 0523
Det var sådan et meget smukt øjeblik, hvor jeg vidste, at hende ville jeg blive kæreste med	NYT - www.alt.dk	Weekly magazine	Marie-Louise Truelsen	2017- 0410
Få styr på din økonomi i 20'erne, 30'erne og 40'erne	NYT - www.alt.dk	Weekly magazine	Stinne Kaasgaard	2017- 0522
Bliv genfødt i Alperne	NYT - www.alt.dk	Weekly magazine	Parastou Booyash	2017- 0315
10 ting du ikke må gå glip af i Tokyo	NYT - www.alt.dk	Weekly magazine	Christina Zemanova	2017- 0216
Familieferie til Sri Lanka: Bountystrande, frodig natur og fornyet energi	NYT - www.alt.dk	Weekly magazine	Wendy Plovmand	2016- 1215
Når folk siger, at det er min pligt som kvinde at få børn, siger jeg: Vrøvl, der er mennesker nok i verden	NYT - www.alt.dk	Weekly magazine	Marie Varming	2017- 2205
Kærligheden har fået Esben Smed til at tilbringe "usandsynligt meget tid på DSB"	NYT - www.alt.dk	Weekly magazine	Ditte-Marie Ascanius	2017- 0327
Krimiforfatter elsebeth egholm om aarhus	NYT - www.alt.dk	Weekly magazine	Jonas Langvad Nilsson	2017- 0518
Hvis kvinder så lidt mindre på telefonen og lidt mere på deres partner, ville der være færre skilsmisser	NYT - www.alt.dk	Weekly magazine	Marie-Louise Truelsen	2017- 0519
Om livet på første klasse: "Hvorfor har hende i minkpelsen sådan nogle triste øjne?"	NYT - www.alt.dk	Weekly magazine	Annette Lykken Sørensen	2017- 0426

En gang sagde jeg nej til et job, og det irriterer mig stadig, at jeg ikke ved, hvad jeg gik glip af	NYT - www.alt.dk	Weekly magazine	Marie Varming	2017- 0410
Lea Korsgaard sagde sit trygge job op: "Frygt er den værste karrierevejleder, du kan få"	NYT - www.alt.dk	Weekly magazine	Simone Brandt Hald	2017- 0203
Sanni Wulff: "Træningen holder mig på plads mentalt og giver mig en fornemmelse af fremdrift"	Sundhed - www.alt.dk	Weekly magazine	Sanni Wulff Vangsø og Lene Roe Rasmussen	2017- 0518
Maria kæmpede sig tilbage efter en svær hovedskade: "Jeg græd kun, når jeg var alene"	Sundhed - www.alt.dk	Weekly magazine	Annette Lykken Sørensen	2017- 0517
Martin Krasnik	NYT - www.alt.dk	Weekly magazine	Stinne Kaasgaard	2017- 0519
Alle syntes, vi var det perfekte par. Men vi blev skilt efter blot to år	NYT - www.alt.dk	Weekly magazine	Marie Varming	2017- 0324

## Appendix D. Co-author Declaration and Confirmation

### Co-author Declaration and Confirmation

Faculty of Humanities  
University of Copenhagen  
2015



---

#### **Describing the independent research contributions of the candidate author**

This declaration should describe the independent research contributions of both the candidate and each of the co- authors for each paper constituting the thesis. The descriptions follow the recommendation from The International Committee of Medical Journal Editors (the "[Vancouver Declaration](#)") See the four criteria:

Attribution of authorship should in general be based on criteria a-d adopted from the Vancouver guidelines<sup>1</sup>, and all individuals who meet these criteria should be recognised as authors:

- a) Substantial contributions to the conception or design of the work, or the acquisition, analysis, or interpretation of data for the work, *and*
- b) drafting the work or revising it critically for important intellectual content, *and*
- c) final approval of the version to be published, *and*
- d) agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

In addition to being accountable for the parts of the work he or she has done, an author should be able to identify which co-authors are responsible for other specific parts of the work.

For each article the declaration should be completed (capital letters if handwritten) and (electronic) signed by the candidate and the co-author(s). Use additional form(s) if necessary. The last page should include all authors' signatures to ensure that you have looked through the declarations, and find the descriptions in accordance with your view of the co-operation that has taken place.

---

**Paper no.:** The article is used in a revised form in the thesis (monograph). This will be clearly stated in the thesis.

**Title:** General and academic high frequency vocabulary in Danish

**Candidate:** Anne Sofie Jakobsen

**Authors:** Anne Sofie Jakobsen, Averil Coxhead, & Birgit Henriksen

**The contribution of the candidate:**

1.) Conception and design of study

2.) Data collection and analysis (corpus compilation) (development of word lists for vocabulary load programme; interpretation of results)

3.) Drafting and writing of article plus critical revision and final draft

**Candidate (capital letters): ANNE SOFIE JAKOBSEN**

---

**Co-author's contribution:**

1.) Conception and design of study and interpretation of results

2.) Drafting and writing of article plus critical revision and final draft

**Name (capital letters): AVERIL COXHEAD**

---

**Co-author's contribution:**

1.) Critical revision of article with minor contributions to the interpretation of results

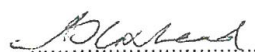
**Name (capital letters): BIRGIT HENRIKSEN**

---



I have looked through the declaration from the other co-authors, and find the descriptions of their contribution in accordance with my view of the cooperation that has taken place

  
.....  
Signature of candidate

  
.....  
Signature of co-author

  
.....  
Signature of co-author

## Appendix E. 402 academic lemmas

Lemma	Translation	Part of speech
af	of	Preposition
afgørende	decisive	Adjective
afhængig	dependent	Adjective
afslutning	end, ending, termination, conclusion	Noun
afslutte	end, finish, terminate, conclude	Verb
afsnit	section	Noun
afstand	distance	Noun
aktiv	active	Adjective
aktivitet	activity	Noun
aktuel	current	Adjective
alternativ (adj)	alternative	Adjective
alternativ (n)	alternative	Noun
an	to	Adverb
analyse	analysis	Noun
angå	concern	Verb
anledning	occasion	Noun
anse	regard	Verb
anvende	usable	Adjective
argument	argument	Noun
arkitekt	architect	Noun
art	species	Noun
baggrund	background	Noun
barn	child	Noun
basere	base	Verb
befinde	find	Verb
befolkning	population	Noun
begge	both	Pronoun
begreb	concept	Noun
begrænse	limit	Verb
behandle	treat	Verb
behov	need	Noun
bemærke	notice	Verb
benytte	use	Verb
beskrive	describe	Verb
beskæftigelse	occupation	Noun
bestå	exist	Verb
betegne	denote	Verb
betragte	regard	Verb
betydelig	significant	Adjective
betydning	meaning	Noun
bevare	preserve	Verb

bevæge	move	Verb
bevægelse	movement	Noun
bidrag	contribution	Noun
bidrage	contribute	Verb
binde	bind	Verb
blomst	flower	Noun
bolig	residence, house, accommodation	Noun
bort	away	Adverb
bred	broad	Adjective
by	town, city	Noun
bygge	build	Verb
byggeri	building, construction	Noun
bygning	building	Noun
central	central	Adjective
centrum	centre	Noun
danne	create	Verb
dans	dance	Noun
dels	partly	Conjunction
deltagelse	participation	Noun
deltager	participant	Noun
denne	this	Pronoun
derimod	however	Adverb
dermed	thus	Adverb
derved	thereby	Adverb
desto	the	Adverb or conjunction
desuden	moreover	Adverb
dialog	dialogue	Noun
dominere	dominate	Verb
drage	draw	Verb
dyrke	grow, cultivate	Verb
død	death	Noun
effektiv	efficient	Adjective
egentlig	actual	Adjective
ejendom	property	Noun
eksempelvis	as an example	Adverb
eksisterende	existing	Adjective
element	element	Noun
eller	or	Conjunction
engelsk	English	Noun
enhed	unit	Noun
enkelt	simply	Adjective
ens	identical	Adjective
enten	either	Conjunction
erfaring	experience	Noun

etablere	establish	Verb
etnisk	ethnic	Adjective
europæisk	European	Adjective
eventuel	possible	Adjective
evne	ability	Noun
faktor	factor	Noun
fokus	focus	Noun
fokusere	focus	Verb
forandring	change, alteration	Noun
forbinde	connect	Verb
forbindelse	connection	Noun
forblive	stay	Verb
fore	line	Verb
forekomme	occur	Verb
foretage	undertake	Verb
forfatter	author	Noun
forhold	condition	Noun
forlag	publisher	Noun
form	form	Noun
formål	purpose	Noun
forskel	difference	Noun
forskellig	different	Adjective
forsker	scientist, researcher	Noun
forskning	research	Noun
forstand	intellect, meaning	Noun
forståelse	understanding	Noun
forsøg	experiment	Noun
forud	ahead	Adverb
foruden	besides	Adverb, preposition or conjunction
forudsætning	basis, assumption	Noun
forventning	expectation	Noun
forælder	parent	Noun
fremgå	appear	Verb
fremhæve	emphasise	Verb
fremme	further	Verb
fremmed	foreign, alien	Adjective
fremmest	foremost	Adjective
fremstille	represent	Verb
fremstå	appear	Verb
fremtidig	future	Adjective
frugt	fruit	Noun
funktion	function	Noun
fysisk	physical	Adjective
fælles	common	Adjective

fænomen	phenomenon	Noun
føde (v)	give birth	Verb
følge (n)	consequence, sequence	Noun
følge (v)	follow	Verb
generation	generation	Noun
generel	general	Adjective
gennem	through	Preposition
gennemføre	complete	Verb
gift	married	Adjective
global	global	Adjective
grad	extent, degree	Noun
grundlag	basis	Noun
grundlæggende	fundamental, basic	Adjective
gruppe	group	Noun
handling	action	Noun
hav	sea, ocean	Noun
havn	harbour	Noun
henholdsvis	respectively	Adverb
hensyn	consideration	Noun
henvisning	eference, referral	Noun
heraf	hereof	Adverb
herefter	henceforth, hereafter	Adverb
hertil	for this purpose, here	Adverb
herunder	such as, including	Adverb
historisk	historical, historic	Adjective
hjælp	help	Noun
hvilken	which	Pronoun
hvoraf	of which, whereof, of whom	Adverb
hvorefter	whereafter	Adverb
hvorvidt	whehter	Adverb
høj	high	Adjective
højde	height	Noun
højre	right	Adjective
håndtere	handle	Verb
identitet	identity	Noun
idet	as	Conjunction
imellem	between	Adverb
imidlertid	however	Adverb
indbygger	inhabitant, citizen	Noun
inddrage	implicate	Verb
indeholde	contain	Verb
indenfor	within	Adverb or preposition
indflydelse	influence	Noun
indgå	enter	Verb

individuel	individual	Adjective
indlede	begin	verb
indre	interior, internal	Adjective
indrette	arrange, organize	Verb
indtage	take in	Verb
inspiration	inspiration	Noun
inspirere	inspire	Verb
institut	department	Noun
institution	institution	Noun
involvere	involve	Verb
is	ice	Noun
især	particular	Adverb
jord	earth, soil, ground, land	Noun
karakter	character, grade	Noun
kendskab	knowledge	Noun
kilo	kilo	Noun
klasse	class	Noun
knytte	bind	Verb
kombinere	combine	Verb
kompetence	competence, skill	Noun
koncentrere	concentrate	Verb
konkludere	conclude	Verb
konklusion	conclusion	Noun
konkret	concrete	Adjective
konstant	constant	Adjective or noun
kontrol	control	Noun
kontrollere	control, monitor, check	Verb
kraft	force	Noun
kraftig	powerful	Adjective
krop	body	Noun
kulturel	cultural	Adjective
kvalitet	quality	Noun
kvinde	woman	Noun
køkken	kitchen	Noun
landskab	landscape	Noun
leg	playing, game	Noun
levende	living, alive	Adjective
ligeledes	as well	Adverb
lille	little	Adjective
litteratur	literature	Noun
liv	life	Noun
lokal	local	Adjective
løbende	running, continuous	Adjective
maleri	painting	Noun

markant	marked	Adjective
markere	mark	Verb
materiale	material	Noun
medføre	entail	Verb
mellem	between	Preposition
menneskelig	human	Adjective
mens	while	Conjunction
mester	master	Noun
metode	method	Noun
middel	means	Noun
midte	centre, middle	Noun
miljø	environment, setting	Noun
model	model	Noun
moderne	modern	Adjective
modstand	opposition, resistance	Noun
modsætning	contrast	Noun
mulig	possible	Adjective
mål	goal	Noun
national	national	Adjective
natur	nature	Noun
naturlig	natural	Adjective
negativ	negative	Adjective
niveau	level	Noun
nordisk	Nordic	Adjective
nævne	mention	Verb
nødvendig	necessary	Adjective
nødvendigvis	necessarily	Adverb
offentlig	public	Adjective
ofte	often	Adverb
omfang	extent	Noun
omfatte	include	Verb
omfattende	extensive, comprehensive	Adjective
omgivelse	surroundings	Noun
område	field, area	Noun
omstændighed	circumstance	Noun
opbygge	construct	Verb
opfatte	perceive	Verb
opfattelse	understanding	Noun
opfylde	satisfy, fulfil	Verb
ophold	stay	Noun
oplevelse	experience	Noun
opmærksomhed	attention	Noun
opnå	obtain	Verb
oprindelig	original	Adjective

opstå	arise	Verb
organisere	organise	Verb
park	park	Noun
part	part, portion, share	Noun
pege	point	Verb
periode	period	Noun
person	person	Noun
perspektiv	perspective	Noun
placere	place	Verb
placering	placement	Noun
point	point	Noun
position	position	Noun
positiv	positive	Adjective
praksis	practice	Noun
praktisk	practical	Adjective
primær	primary	Adjective
princip	principle	Noun
proces	process	Noun
projekt	project	Noun
præge	mark	Verb
påvirke	influence	Verb
radikal	radical	Adjective
ramme (n)	frame	Noun
regel	rule	Noun
relation	relation	Noun
relativ	relative	Adjective
relevant	relevant	Adjective
repræsentere	represent	Verb
ressource	ressource	Noun
resultat	result	Noun
retning	direction	Noun
rig	rich	Adjective
ring	ring	Noun
rod	root	Noun
rolle	role	Noun
rum	room, space	Noun
rumme	hold	Verb
række (n)	row	Noun
samfund	society	Noun
samle	unite assemble, join, connect	Verb
sammenhæng	context	Noun
sandsynligvis	probably	Adjective
seksuel	sexual	Adjective
selve	actual	Adjective



selvom	even if	Conjunction
sjælden	rare	Adjective
skabe	create	Verb
skjule	hide	Verb
slutning	end	Noun
smag	taste, flavour	Noun
snarere	if anything, rather, sooner	Adverb
social	social	Adjective
spille (v)	play	Verb
stamme	originate	Verb
statistik	statistics	Noun
statslig	national, state	Adjective
status	status	Noun
sten	stone	Noun
stige	rise	Verb
studere	study	Verb
studie	study	Noun
styrke	strength, strengthen	Noun, Verb
størrelse	size	Noun
sundhed	health	Noun
supplere	supplement	Verb
svag	weak	Adjective
synes	think	Verb
system	system	Noun
særlig	special	Adjective
sæt	set	Noun
således	thus	Adverb
såvel	as well as	Conjunction
takt	pace, rate	Noun
tale (n)	speech	Noun
tegn	sign	Noun
teknik	technique	Noun
teknologi	technology	Noun
tema	theme	Noun
tendens	tendency	Noun
teori	theory	Noun
test	test	Noun
the	the	Article
tilfælde	case	Noun
tilhøre	belong to	Verb
tilstrækkelig	sufficient	Adjective
tilsvarende	corresponding	Adjective
tiltag	initiative	Noun
tradition	tradition	Noun

traditionel	traditional	Adjective
trods (n)	defiance	Noun
træk	feature	Noun
træne	train	Verb
tværs	cross	Adjective
tyde	interpret	Verb
tydelig	distinct	Adjective
type	type	Noun
typisk	typical	Adjective
tysk	German	Adjective
udarbejde	prepare	Verb
udbrede	spread	Verb
udføre	carry out	Verb
udgangspunkt	starting point	Noun
udgøre	constitute	Verb
udtryk	expression	Noun
udtrykke	express	Verb
udvide	extend	Verb
udvikle	develop	Verb
udvikling	development	Noun
udvælge	select	Verb
ukendt	unknown	Adjective
umiddelbar	immediate	Adjective
undersøge	investigate	Verb
undersøgelse	investigation, examination	Noun
undervisning	teaching	Noun
unge	adolescents	Noun
universitet	university	Noun
usikkerhed	uncertainty, insecurity	Noun
vanskelig	difficult	Adjective
ved	at, by, near	Preposition
vestlig	western	Adjective
vid	wide	Adjective
viden	knowledge	Noun
videnskabelig	scientific	Adjective
vinter	winter	Noun
virkning	effect	Noun
vis	certain, way	Adjective, Noun
vise	show	Verb
vold	violence	Noun
vurdering	assessment	Noun
væg	wall	Noun
vægt	weight	Noun
vækst	growth, increase	Noun

værdi	value	Noun
væsentlig	essential	Adjective
yderlig	extreme	Adjective
ændring	change, alteration	Noun
ø	island	Noun
ønske	desire, wish	Noun
årig	yearly, annual	Adjective
årsag	cause	Noun
årti	decade	Noun

## Appendix F. The DAWL

The DAWL lemmas are ranked according to frequency with the most frequent lemma first.

Words in italics also occur among the 2,000 most frequently used lemmas in Danish (Det Danske Sprog- og Litteraturselskab, n.d.)

Words in bold overlap with 402 academic lemmas identified in Study 1, Chapter 5. As such, they also occur among the 2,000 most frequently used lemmas in Danish (Det Danske Sprog- og Litteraturselskab, n.d.).

Words followed by an asterisk are phrasal elements as shown in Study 3, Chapter 7.

DAWL lemma	Translation	Part of speech
<i>den</i>	it	Article or pronoun
<b>af</b>	of	Preposition
<i>som</i>	which	Pronoun
<i>med</i>	with	Preposition
<i>om</i>	about	Preposition, adverb or conjunction
<b>denne</b>	this	Pronoun
<b>eller</b>	or	Conjunction
<i>sig</i>	oneself	Pronoun
<i>anden</i>	other	Pronoun
<i>men</i>	but	Conjunction
<b>mellem</b>	between	Preposition
<i>dansk</i>	Danish	Adjective
<i>mere</i>	more	Adjective
<b>forhold</b>	condition	Noun
<i>to</i>	two	Numeral
<i>deres</i>	their	Pronoun
<b>forskellig</b>	different	Adjective
<i>end</i>	than	Adverb or conjunction
<b>lille</b>	small	Adjective
<b>vise</b>	show	Verb
<b>således</b>	thus	Adverb
<b>hvilken</b>	which	Pronoun
<i>først</i>	first	Adverb
<i>derfor</i>	therefore	Adverb
<b>høj</b>	high	Adjective
<i>del</i>	part	Noun
<i>hvordan</i>	how	Adverb
<b>form</b>	form	Noun
<i>både</i>	both	Conjunction
<i>måde</i>	way	Noun

<i>samme</i>	same	Adjective
<i>dog</i>	however	Adverb
<i>fx</i>	for example	Abbreviation
<b>analyse</b>	analysis	Noun
<b>udvikling</b>	development	Noun
<i>flere</i>	several	Adjective
<b>gennem</b>	through	Preposition
<i>sådan</i>	such	Adjective
<i>spørgsmål</i>	question	Noun
<b>grad*</b>	extent, degree	Noun
<b>følge</b>	follow	Verb
<b>mens, medens</b>	while	Conjunction
<b>betydning</b>	meaning	Noun
<b>enkelt</b>	simply	Adjective
<i>tidlig</i>	early	Adjective
<b>eksempel</b>	example	Noun
<b>særlig</b>	particular	Adjective
<i>egen</i>	own	Adjective
<b>forskel</b>	difference	Noun
<b>sammenhæng</b>	context	Noun
<b>forbindelse</b>	connection	Noun
<b>skabe</b>	create	Verb
<b>mulig</b>	possible	Adjective
<i>altså</i>	therefore	Adverb
<b>række</b>	row	Noun
<i>samtidig</i>	simultaneous	Adjective or adverb
<i>vigtig</i>	important	Adjective
<b>resultat</b>	result	Noun
<i>sen</i>	late	Adjective
<b>ofte</b>	often	Adverb
<b>dermed</b>	thus	Adverb
<i>tre</i>	three	Numeral
<b>beskrive</b>	describe	Verb
<b>praksis</b>	practice	Noun
<i>problem</i>	problem	Noun
<b>generel</b>	general	Adjective
<b>type</b>	type	Noun
<b>tilfælde</b>	case	Noun
<i>gælde</i>	be valid	Noun
<i>forstå</i>	understand	Verb
<b>værdi</b>	value	Noun
<b>udtryk</b>	expression	Noun
<b>baggrund</b>	background	Noun
<b>offentlig</b>	public	Adjective

<b>især</b>	particular	Adverb
<b>udvikle</b>	develop	Verb
<i>betyde</i>	mean	Verb
<b>imidlertid</b>	however	Adverb
findes	be	Verb
<b>fokus</b>	focus	Noun
<b>forståelse</b>	understanding	Noun
<i>idet</i>	as	Conjunction
<b>undersøge</b>	investigate	Verb
<b>udgangspunkt</b>	starting point	Noun
<b>væsentlig</b>	essential	Adjective
<b>synes</b>	think	Verb
<i>blot</i>	only	Adverb, adjective or conjunction
<b>basere</b>	base	Verb
<b>central</b>	central	Adjective
<b>dels</b>	partly	Conjunction
<i>dvs.</i>	i.e.	Abbreviation
<b>nævne</b>	mention	Verb
<b>relativ</b>	relative	Adjective
<b>indgå</b>	enter	Verb
bl.a.	among other	Abbreviation
<i>øvrig</i>	besides	Adjective
<i>situation</i>	situation	Noun
<b>metode</b>	method	Noun
<b>primær</b>	primary	Adjective
<b>udgøre</b>	constitute	Verb
<b>rolle</b>	role	Noun
<i>nær</i>	near	Adjective, preposition or adverb
<b>konkret</b>	concrete	Adjective
<i>ifølge</i>	according to	Preposition
<b>litteratur</b>	literature	Noun
<b>formål</b>	purpose	Noun
<i>krav</i>	demand	Noun
<b>national</b>	national	Adjective
<i>bestemt</i>	certainly	Adjective
én	one	Numeral
<b>vis</b>	certain	Adjective
<b>model</b>	model	Noun
<i>stærk</i>	strong	Adjective
<i>kalde</i>	call	Verb
<b>erfaring</b>	experience	Noun
<b>niveau</b>	level	Noun
<i>direkte</i>	direct	Adjective
<i>ligesom</i>	like	Adverb

<b>behov</b>	need	Noun
<b>positiv</b>	positive	Adjective
<i>jf.</i>	cf.	Abbreviation
<b>perspektiv</b>	perspective	Noun
<b>system</b>	system	Noun
<b>proces</b>	process	Noun
<b>oplevelse</b>	experience	Noun
<b>påvirke</b>	influence	Verb
<b>bestå</b>	exist	Verb
<b>begge</b>	both	Pronoun
<b>pege</b>	point	Verb
<i>svare</i>	answer	Verb
<b>universitet</b>	university	Noun
<b>mål</b>	goal	Noun
<b>indeholde</b>	contain	Verb
<i>tal</i>	number	Noun
<b>aktivitet</b>	activity	Noun
<b>foretage</b>	undertake	Verb
<i>hinanden</i>	each other	Pronoun
<i>interesse</i>	interest	Noun
<b>relevant</b>	relevant	Adjective
<b>nødvendig</b>	necessary	Adjective
<i>ene</i>	alone	Noun, pronoun or adjective
<i>international</i>	international	Adjective
<i>føre</i>	carry	Noun
<i>indhold</i>	content	Noun
<b>såvel</b>	as well as	Conjunction
<b>knytte</b>	bind	Verb
<b>vurdering</b>	assessment	Noun
<b>aktiv</b>	active	Adjective
<b>funktion</b>	function	Noun
<i>eksistere</i>	exist	Verb
<i>diskussion</i>	discussion	Noun
<i>fortsætte</i>	continue	Verb
<i>beskrivelse</i>	description	Noun
<b>derimod</b>	however	Adverb
<i>alene</i>	alone	Adjective
<b>bidrage</b>	contribute	Verb
<i>kategori</i>	category	Noun
<i>fald*</i>	fall	Noun
<b>opstå</b>	arise	Verb
<b>regel</b>	rule	Noun
<i>mangle</i>	lack	Verb
<b>befolkning</b>	population	Noun

<b>fælles</b>	common	Adjective
<b>typisk</b>	typical	Adjective
<i>interview</i>	interview	Noun
<b>omfang</b>	extent	Noun
<b>tendens</b>	tendency	Noun
specifik	specific	Adjective
<b>begrænse</b>	limit	Verb
<b>opfattelse</b>	understanding	Noun
<b>traditionel</b>	traditional	Adjective
definere	define	verb
<b>betragte</b>	regard	Verb
<b>bred</b>	broad	Adjective
<b>danne</b>	create	Verb
<i>endelig</i>	finally	Adjective or adverb
<i>diskutere</i>	discuss	Verb
<i>forklare</i>	explain	Verb
<b>grundlag</b>	basis	Noun
<b>etablere</b>	establish	Verb
<b>tydelig</b>	distinct	Adjective
<b>bevægelse</b>	movement	Noun
overordnet	superior, general	Adjective
<b>fokusere</b>	focus	Verb
<b>forsøg</b>	experiment	Noun
<b>umiddelbar</b>	immediate	Adjective
<i>ganske</i>	quite	Adjective
tilgang	approach	Noun
<b>tilsvarende</b>	corresponding	Adjective
<b>fremgå</b>	appear	Verb
<b>behandle</b>	treat	Verb
<b>afgørende</b>	decisive	Adjective
<b>opfatte</b>	perceive	Verb
<i>kritisk</i>	critical	Adjective
<i>konsekvens</i>	consequence, consistency	Noun
<i>skylde</i>	owe	Verb
<b>selve</b>	actual	Adjective
<b>afsnit</b>	section	Noun
<i>efterfølge</i>	succeed	Verb
<b>benytte</b>	use	Verb
<b>yderlig</b>	extreme	Adjective
<b>enten</b>	either	Conjunction
<b>hensyn*</b>	consideration	Noun
<b>årsag</b>	cause	Noun
analysere	analyse	Verb
<b>trods*</b>	despite	Noun or preposition



<b>eksempelvis</b>	as an example	Adverb
<b>ligeledes</b>	as well	Adverb
<b>udtrykke</b>	express	Verb
<b>praktisk</b>	practical	Adjective
<b>forekomme</b>	occur	Verb
<i>kilde</i>	source	Noun
struktur	structure	Noun
<b>fremhæve</b>	emphasise	Verb
<i>interessant</i>	interesting	Adjective
<b>henholdsvis</b>	respectively	Adverb
<i>grænse</i>	limit	Noun or verb
<b>videnskabelig</b>	scientific	Adjective
<b>natur</b>	nature	Noun
<b>repræsentere</b>	represent	Verb
<b>retning</b>	direction	Noun
<i>foregå</i>	take place	Verb
aspekt	aspect	Noun
<b>individuel</b>	individual	Adjective
<b>princip</b>	principle	Noun
endvidere	moreover	Adverb
<b>aktuel</b>	current	Adjective
<b>konklusion</b>	conclusion	Noun
<b>naturlig</b>	natural	Adjective
<i>såkaldt</i>	so-called	Adjective
<b>vanskelig</b>	difficult	Adjective
<i>udfordring</i>	challenge	Noun
<b>modsætning</b>	contrast	Noun
<b>hvorvidt</b>	whether	Adverb
<b>inddrage</b>	implicate	Verb
<b>ressource, resurse</b>	ressource	Noun
<i>leder</i>	leader	Noun
<b>negativ</b>	negativ	Adjective
<i>holdning</i>	attitude	Noun
<b>sjælden</b>	rare	Adjective
<i>omtale</i>	comment	Noun
indebære	imply	Verb
<i>understrege</i>	emphasise	Verb
<i>klassisk</i>	classic	Adjective
potentiel	potential	Adjective
<b>markant</b>	marked	Adjective
<b>vægt</b>	weight	Noun
antage	assume	Verb
<b>forudsætning</b>	basis	Noun
karakterisere	characterise	Verb

genstand	object	Noun
adskille	separate	Verb
<b>evne</b>	ability	Noun
<i>henvise</i>	refer	Verb
<b>afhængig</b>	dependent	Adjective
<i>rette</i>	right	Verb or noun
<b>institut</b>	department	Noun
tredje	third	Numeral
<b>rumme</b>	hold	Verb
<b>indflydelse</b>	influence	Noun
<i>overveje</i>	consider	Verb
<i>strategi</i>	strategy	Noun
<i>forklaring</i>	explanation	Noun
<i>stand*</i>	condition	Noun
<b>eventuel</b>	possible	Adjective
<b>placere</b>	place	Verb
henblik*	regard	Noun
<b>nødvendigvis</b>	necessarily	Adverb
<b>status</b>	status	Noun
<b>udvælge</b>	select	Verb
<b>heraf</b>	hereof	Adverb
reference	reference	Noun
<i>hverken</i>	either	Adverb
belyse	illustrate	verb
<b>oprindelig</b>	original	Adjective
definition	definition	Noun
relatere	relate	Verb
<b>betegne</b>	denote	Verb
<i>medvirke</i>	contribute	Verb
<b>følge</b>	sequence	Noun
<i>omvende</i>	convert	Verb
<i>optræde</i>	appear	Verb
<b>udbrede</b>	spread	Verb
<i>dreje</i>	be about sth	Verb
<b>tilstrækkelig</b>	sufficient	Adjective
<b>dominere</b>	dominate	Verb
<i>punkt</i>	point	Noun
<b>svag</b>	weak	Adjective
<i>beslutning</i>	decision	Noun
afspejle	reflect	Verb
<b>angå</b>	concern	Verb
<i>bestemme</i>	decide	Verb
problemstilling	problem	Noun
ovenfor	above	Adverb

<b>opmærksomhed</b>	attention	Noun
tilgængelig	accessible	Adjective
<i>vor</i>	our	Pronoun
<b>anse</b>	regard	Verb
hermed	herewith	Adverb
<b>fremmest*</b>	foremost	Adjective
<i>fastholde</i>	maintain	Verb
såsom	such as	Conjunction
understøtte	support	Verb
<b>supplere</b>	supplement	Verb
<b>udvide</b>	extend	Verb
<b>tyde</b>	interpret	Verb
illustrere	illustrate	Verb
overensstemmelse	accordance	Noun
hvorledes	how	Adverb
<b>omfattende</b>	extensive, comprehensive	Adjective
<b>involvere</b>	involve	Verb
<b>konkludere</b>	conclude	Verb
<i>påpege</i>	indicate	Verb
<i>foreslå</i>	suggest	Verb
overvejelse	consideration	Noun
interaktion	interaction	Noun
overfor	opposite	Adverb
<i>begyndelse</i>	beginning	Noun
forudsætte	assume	Verb
kriterium	criterion	Noun
<b>vid</b>	wide	Noun or adjective
fase	phase	Noun
<b>derved</b>	thereby	Adverb
reel	real	Adjective
<i>formentlig</i>	supposed	Adjective
<b>tema</b>	theme	Noun
skelne	distinguish	Verb
oftest	most often	Adverb
<i>vidt*</i>	far	Adverb
<i>pågældende</i>	in question	Adjective
egenskab	attribute	Noun
given	given	Adjective
<b>styrke</b>	strength	Noun
<b>bevæge</b>	move	Verb
<b>forbinde</b>	connect	Verb
<i>vedkommende</i>	concerned	Noun
betegnelse	designation	Noun
<i>udelukkende</i>	solely	Adjective

<i>beskæftige</i>	employ	Verb
<b>tværs*</b>	cross	Adjective
konstruere	construct	Verb
<i>tilsyneladende</i>	apparent	Adjective
<b>tegn</b>	sign	Noun
<b>fremstille</b>	represent	Verb
<b>anledning</b>	occasion	Noun
<b>placering</b>	placement	Noun
<b>slutning</b>	end	Noun
<i>bekræfte</i>	confirm	verb
<b>tiltag</b>	initiative	Noun
<b>fremtidig</b>	future	Adjective
<b>tilhøre</b>	belong to	Verb
integrere	integrate	Verb
<i>intern</i>	internal	Adjective
redskab	tool	Noun
tilstedeværelse	presence	Noun
<b>vis</b>	manner	Noun
<b>binde</b>	bind	Verb
<i>konflikt</i>	conflict	Noun
<b>organisere</b>	organise	Verb
<b>bidrag</b>	contribution	Noun
<b>hertil</b>	for this purpose, here	Adverb
lighed	similarity	Noun
<i>acceptere</i>	accept	Verb
publicere	publish	Verb
<b>årti</b>	decade	Noun
delvis	partially	Adjective
orientere	orientate	Verb
<i>mangel</i>	lack	Noun
<i>osv.</i>	etc.	Abbreviation
<i>simpel</i>	simple	Adjective
<i>konstatere</i>	ascertain	Verb
<i>omgang</i>	turn	Noun
<i>afgøre</i>	determine	Verb
observation	observation	Noun
entydig	unambiguous	Adjective
problematisk	problematic	Adjective
betinge	determine	Verb
indsigt	insight	Noun
<b>kendskab</b>	knowledge	Noun
helhed	whole	Noun
stede*	present	Noun
tilpasse	adapt	Verb

modsat	opposite	Adjective
basis	basis	Noun
<i>officiel</i>	official	Adjective
tolke	interpret	Verb
anføre	state	Verb
betragtning	consideration	Noun
<i>indføre</i>	introduce	Verb
samspil	interplay	Noun
<i>dokumentere</i>	document	Verb
<i>initiativ</i>	initiative	Noun
antype	indicate	Verb
nedenfor	below	Adverb
etc.	etc.	Abbreviation
afgrænse	delineate	Verb
<b>styrke</b>	strengthen	Verb
afhænge	depend	verb
hensigt	intention	Noun
<i>rund</i>	round	Adjective
<i>spor</i>	track	Noun
ovenstående	the above	Adjective
gradvis	gradual	Adjective
<b>desto</b>	the	Adverb or conjunction
<i>ej</i>	no	Adverb
opretholde	sustain	Verb
<b>levende</b>	alive	Adjective
formel	formula	Adjective
indirekte	indirect	Adjective
omhandle	concern	Verb
formulering	formulation	Noun
karakteristisk	characteristical	Adjective
<b>studere</b>	study	Verb
relevans	relevance	Noun
introducere	introduce	Verb
<i>udfordre</i>	challenge	Verb
<i>dertil</i>	besides	Adverb
begrænsning	limitation	Noun
operere	operate	Verb
organisering	organising	Noun
præference	preference	Noun
formode	suppose	Verb
<i>udpege</i>	designate	Verb
hvorved	at which	Adverb
<i>gentage</i>	repeat	Verb
hvor	where	Adverb

muliggøre	make possible, permit	Verb
intention	intention	Noun
mønster	pattern	Noun
sidstnævnte	the latter	Adjective
<i>indtryk</i>	impression	Noun
<i>sigte</i>	aim	Verb
<b>forblive</b>	stay	Verb
<b>sandsynligvis</b>	probably	Adjective
<i>erstatte</i>	replace	Verb
oplagt	obviously	Adjective
<b>omgivelse</b>	surroundings	Noun
tillige	as well	Adverb
<b>indtage</b>	take in	Verb
<b>stamme</b>	originate	Verb
undtagelse	exception	Noun
<b>opbygge</b>	construct	Verb
tolkning	interpretation	Noun
<i>tilfældig</i>	accidental	Adjective
kerne	core	Noun
rationel	rational	Adjective
afsæt	starting point	Noun
gensidig	mutual	Adjective
sammenhængende	coherent	Adjective
m.fl.	and others	Abbreviation
<i>samlige</i>	all	Adjective
velkendt	well-known	Adjective
introduktion	introduction	Noun
<i>ringe</i>	bad	Adjective
<i>planlægge</i>	plan	Verb
tilskrive	attribute to	Verb
detaljeret	detailed	Adjective
afdække	uncover	Verb
hensigtsmæssig	appropriate	Adjective
indbyrdes	mutual, reciprocal	Adjective
<i>dobbelt</i>	double	Adjective
lig	like	Adjective
tillægge	ascribe to	Verb
forskningsprojekt	research project	Noun
<i>tillid</i>	trust	Noun
<i>hidtil</i>	so far	Adverb
udformning	version	Noun
isolere	isolate	Verb
metodisk	methodical	Adjective
berøre	affect	Verb

tilknytte	attach	Verb
<i>anlægge</i>	establish	Verb
hvorimod	whereas	Adverb
<i>udelukke</i>	exclude	Verb
<i>efterlade</i>	leave behind	Verb
dernæst	next	Adverb
fundamental	fundamental	Adjective
<i>sigt*</i>	sight	Noun
bemærkelsesværdig	remarkable	Adjective
<i>fornemmelse</i>	sense	Noun
mangfoldighed	diversity	Noun
konventionel	conventional	Adjective
<b>ukendt</b>	unknown	Adjective
skift	change	Noun
<b>forud</b>	ahead	Adverb
opdele	divide up	Verb
<b>koncentrere</b>	concentrate	Verb
neutral	neutral	Adjective
<i>overleve</i>	survive	Verb
repræsentativ	representative	Adjective
univers	universe	Noun
nærliggende	nearby	Adjective
enighed	agreement	Noun
kompliceret	complicated	Adjective
uklar	indistinct	Adjective
uddybe	clarify	Verb
<b>foruden</b>	besides	Adverb, preposition or conjunction
udvise	display	Verb
dynamik	dynamics	Noun
overgang	passage	Noun
passiv	passive	Noun
sammenfatte	summarise	Verb
<i>erkende</i>	acknowledge	Verb
iværksætte	implement	Verb
orientering	orientation	Noun
udfylde	fill up	Verb
fremkomme	appear	Verb
koble	link	Verb
prioritering	prioritization	Noun
<i>tiltrække</i>	attract	Verb
<i>nogenlunde</i>	tolerable	Adjective
betydningsfuld	important	Adjective
flyde	flow	Verb
underlægge	place under	Verb

<i>krise</i>	crisis	Noun
forstærke	strengthen	Verb
gennemgående	common	Adjective
skitsere	outline	Verb
afgrænsning	delineation	Noun
kategorisere	categorise	Verb
beskeden	modest	Adjective
<i>daværende</i>	then	Adjective
færdighed	skill	Noun
omgive	surround	Verb
redegøre	give an account of	Verb
forudgående	preceding	Adjective
tilsammen	altogether	Adverb
gengive	render	Verb
<i>fastslå</i>	demonstrate, establish	Verb
fjerde	fourth	Numeral
parallel	parallel	Adjective
efterspørge	demand	Verb
heri	herein	Adverb
hvormed	with what	Adverb
hvorfra	where	Adverb
sammensætte	compound, compile	Verb
skærpe	sharpen	Verb
spore	monitor	Verb
udforske	explore	Verb
accept	accept	Noun
tankegang	mentality	Noun
regi	framework	Noun
udforme	frame	Verb
benævne	designate	Verb
formel	formal	Noun
genfinde	recover	Verb
konkurrere	compete	Adjective
undlade	omit	Verb
påbegynde	start	Verb
styring	administration	Noun
frembringe	produce	Verb
omverden	surrounding world	Noun
genkende	recognise	Verb
underliggende	underlying	Adjective
værdifuld	valuable	Adjective
fremlægge	present	Verb
<i>grunde</i>	base	Verb
klassificere	classify	Verb



hvile	rest	Verb
sammenholde	relate	Verb
<i>forvejen*</i>	ahead	Noun
generalisere	generalise	Verb
generere	generate	Verb
<i>omtale</i>	comment on	Verb
opløse	dissolve	Verb
svække	weaken	Verb
fremtrædende	salient	Adjective
<i>bortset</i>	apart from	Adverb
nuanceret	varied	Adjective
tilfredsstille	satisfy	Verb
ydermere	further	Adverb
opsummere	sum up	Verb
differentiere	differentiate	Verb
vigtighed	importance	Noun
forstyrre	disrupt	Verb
deraf	thereof	Adverb
indblik	insight	Noun
essentiell	essential	Adjective
præg	character	Noun
tydeliggøre	elucidate	Verb
underbygge	substantiate	Verb
indlysende	obvious, evident	Adjective
overens*	similar	Adverb
respektive	respectively	Adjective
identisk	identical	Adjective
brugbar	useful	Adjective
anonym	anonymous	Adjective
arv	inheritance	Noun
herfor	for this	Adverb
parallel	parallel	Noun
bedømme	assess	Verb
indvandring	immigration	Noun
endog	even	Adverb
udlægge	construe	Verb
almindelighed*	generality	Noun
bevirke	cause	Verb
udefra	outside	Adverb
uendelig	infinite	Adjective
<i>afløse</i>	relieve, replace	Verb
forskydning	displacement	Noun
fundament	foundation	Noun
udførelse	execution	Noun

hidtidig	hitherto	Adjective
modificere	modify	Verb
disposition	outline	Noun
opdagelse	discovery	Noun
tiltage	increase	Verb
<b>række</b>	reach	Verb
selvsagt	obvious	Adjective
aflede	derive	Verb
imødekomme	oblige	Verb
nytte	use	Noun
<b>stamme</b>	tribe	Noun
produktiv	productive	Adjective
troværdighed	reliability	Noun
udredning	explanation	Noun
utvivlsom	undoubtedly	Adjective
vifte	fan	Noun
besiddelse	possession	Noun
målestok	scale	Noun
substans	substance	Noun
særskilt	separate	Adjective
anvendelig	usable	Adjective
betragtelig	considerable	Adjective
bearbejdning	processing	Noun
bekostning*	cost	Noun
forståelig	comprehensible	Adjective
forvalte	manage	Verb
vægte	weight	Verb
bedømmelse	assessment	Noun
tilstræbe	aim to	Verb
herpå	subsequently	Adverb
klarhed	clarity	Noun
forelæsning	lecture	Noun
foreskrive	prescribe	Verb
gyldighed	validity	Noun
udforskning	exploration	Noun
videreføre	continue	Verb
igangværende	in progress	Adjective
revidere	revise	Verb
adressere	address	Verb
obligatorisk	mandatory	Adjective
tilblivelse	birth	Noun
udsnit	sample	Noun
videreudvikle	develop further	Verb
anslå	estimate	Verb

bredde	width	Noun
undergå	undergo	Verb
nævneværdig	mentionable	Adjective
fortrinsvis	preferential	Adjective
modsvare	correspond to	Verb
tilbagevendende	returning	Adjective
intensivere	intensify	Verb
stadighed*	steadiness	Noun
tolerance	tolerance	Noun
balancere	balance	Verb
indgående	thoroughly	Adjective
mente*	number carried	Verb
vedr.	pertaining to	Abbreviation
hastig	hurried	Adjective
overfladisk	superficial	Adjective
fastholdelse	insistence	Noun
cirkel	circel	Noun
tilbøjelig	disposed	Adjective
tilnærmelsesvis	approximate	Adjective
værdsætte	appreciate	Verb
påkrævet	required	Adjective
vanskeliggøre	complicate	Verb
forventelig	expected	Adjective
impuls	impulse	Noun
nedskrive	write down	Verb
slås	fight	Verb
forveksle	confound	Verb
konsistent	consistant	Adjective
opretholdelse	maintenance	Noun
rekonstruere	reconstruct	Verb
kulminere	culminate	Verb
nøgleord	keyword	Noun
fordelagtig	advantageous	Adjective
langtfra	far from it	Adverb
sideløbende	parallel	Adjective
alle	all	Pronoun
andetsteds	elsewhere	Adverb
ønskelig	desirable	Adjective
formodning	presumption	Noun
dagligdags	everyday	Adjective
heterogenitet	heterogeneity	Noun
varetage	attend to, manage	Verb
berige	enrich	Verb
fortsættelse	continuation	Noun

symmetrisk	symmetrical	Adjective
veksle	change	Verb
afklaring	clarification	Noun
konstatering	ascertainment	Noun
<i>vige</i>	retreat	Verb
eksakt	exact	Adjective
kompensere	compensate	Verb
overdrive	exaggerate	Verb
bibeholde	maintain	Verb
gennemgribende	thorough	Adjective
veletableret	well-established	Adjective
visualisere	visualise	Verb
afsøge	search	Verb
befordre	promote	Verb
perspektivere	put into perspective	Verb
systematisere	systematise	Verb
ubetydelig	insignificant	Adjective
spilleregul	rule of the game	Noun
uklarhed	indistinctness	Noun
sjette	sixth	Numeral
lign.	similar	Abbreviation
nytte	be of use	Verb
hvile	rest	Noun
indledningsvis	by way of introduction	Adverb
negligere	neglect	Verb
omdiskuteret	controversial	Adjective
sædvanligvis	usually	Adjective
håndgribelig	tangible	Adjective
modifikation	modification	Noun
<i>munde*</i>	result in	Verb
Sande	admit	Verb
selvfølgelig	matter of course	Noun
tilskynde	prompt	Verb
handlekraft	resourcefulness	Noun
<i>sigte</i>	aim	Noun
kontur	contour	Noun
spinkel	slight	Adjective
vidtgående	far-reaching	Adjective
bevågenhed	attention	Noun
forefinde	find	Verb
binær	binary	Adjective
hvorunder	under which	Adverb
spekter, spektrum	spectrum	Noun
svinge	swing	Verb

indskrænkning	reduction, restriction	Noun
optegnelse	note	Noun
tilsigte	aim at	Verb
uforudset	unforeseen	Adjective
enkeltvis	individually	Adjective
indblanding	intervention	Noun
mainstream	mainstream	Noun
nuancering	making sth varied	Noun
virkeliggøre	realise	Verb
omskrivning	revision	Noun
uegnet	unfit	Adjective
afrunding	conclusion	Noun
cirkulation	circulation	Noun
<b>evne</b>	be able to	Verb
granskning	scrutiny	Noun
redefinere	redefine	Verb
tilkendegivelse	expression	Noun
underordne	subordinate	Verb
omskrive	rewrite	Verb
tvungen	compulsory	Adjective
hobe*	heap	Noun
hovedårsag	main cause	Noun
indicium	indication	Noun
ministeriel	ministerial	Adjective
scenario, scenarie	scenario	Noun
påbegyndelse	commencement	Noun
koordinat	coordinate	Noun
viderebringe	send forward	Verb
føromtalt	above-mentioned	Adjective
hovedelement	main element	Noun
nordamerikansk	north american	Adjective
frembyde	offer	Verb
gennemskære	intersect	Verb
værdigrundlag	fundamental values	Noun
brugbarhed	usefulness	Noun
sammentænke	synthesise	Verb

## Appendix G. The S-DAWL

The S-DAWL is listed according to morphological relatedness with the DAWL items listed first followed by the added derivations and then compounds. Words in bold are DAWL words. Underlined words are added derivations, and words in bold are compounds.

Word group	S-DAWL	Translation	D value	POS
1	<b>accept</b>	accept	0.80	Noun
1	<b>acceptere</b>	accept	0.80	Verb
1	<u>acceptabel</u>	acceptable	0.70	Adjective
1	<u>uacceptabel</u>	unacceptable	0.70	Adjective
2	<b>adressere</b>	address	0.80	Verb
3	<b>adskille</b>	separate	0.80	Verb
4	<b>af</b>	of	0.80	Preposition
5	<b>afdække</b>	uncover	0.80	Verb
5	<u>afdækning</u>	uncover	0.70	Noun
6	<b>afgrænse</b>	delineate	0.80	Verb
6	<b>afgrænsning</b>	delineation	0.80	Noun
7	<b>afgøre</b>	determine	0.80	Verb
7	<b>afgørende</b>	decisive	0.80	Adjective
8	<b>afhænge</b>	depend	0.80	Verb
8	<b>afhængig</b>	dependent	0.80	Adjective
8	<u>uafhængig</u>	independent	0.70	Adjective
8	<u>afhængighed</u>	dependence	0.70	Noun
8	<u>uafhængighed</u>	independence	0.60	Noun
8	<i>afhængighedsforhold</i>	dependency, state of dependence	0.70	Noun
9	<u>afklare</u>	clarify	0.70	Verb
9	<b>afklaring</b>	clarification	0.80	Noun
9	<u>uafklaret</u>	undetermined	0.60	Adjective
10	<b>aflede</b>	derive	0.80	Verb
11	<b>afløse</b>	relieve, replace	0.80	Verb
12	<b>afrunding</b>	conclusion	0.80	Noun
13	<b>afsnit</b>	section	0.80	Noun
14	<b>afspejle</b>	reflect	0.80	Verb
15	<b>afsæt</b>	starting point	0.80	Noun
16	<b>afsøge</b>	search	0.80	Verb
17	<b>aktiv</b>	active	0.80	Adjective
17	<b>aktivitet</b>	activity	0.80	Noun
17	<u>aktivere</u>	activate	0.70	Verb
17	<u>aktivering</u>	activation	0.70	Noun
17	<u>aktivisme</u>	activism	0.60	Noun
17	<u>aktivistisk</u>	activist, activistic	0.60	Adjective
18	<b>aktuel</b>	current	0.80	Adjective

18	<u>aktualitet</u>	topicality, current interest	0.60	Noun
19	<b>alene</b>	alone	0.80	Adjective
20	<b>alle</b>	all	0.80	Pronoun
21	<b>almindelighed</b>	generality	0.80	Noun
21	<u>almindeligvis</u>	generally	0.70	Adverb
22	<b>altså</b>	therefore	0.80	Adverb
23	<b>analyse</b>	analysis	0.80	Noun
23	<b>analysere</b>	analyse	0.80	Verb
23	<u>analytisk</u>	analytic, analytical	0.60	Adjective
23	<i>analysemetode</i>	method of analysis	0.70	Noun
23	<i>analysemodel</i>	analysis model	0.60	Noun
23	<i>analyseniveau</i>	level of analysis	0.60	Noun
23	<i>analysestrategi</i>	analysis strategy	0.60	Noun
23	<i>analysetilgang</i>	analytical approach	0.60	Noun
23	<i>diskursanalyse</i>	discourse analysis	0.60	Noun
24	<b>anden</b>	other	0.80	Pronoun
24	<b>andetsteds</b>	elsewhere	0.80	Adverb
25	<b>angå</b>	concern	0.80	Verb
26	<b>anledning</b>	occasion	0.80	Noun
27	<b>anlægge</b>	establish	0.80	Verb
28	<b>anonym</b>	anonymous	0.80	Adjective
28	<u>anonymisere</u>	anonymise	0.70	Verb
28	<u>anonymitet</u>	anonymity	0.60	Noun
29	<b>anse</b>	regard	0.80	Verb
30	<b>anslå</b>	estimate	0.80	Verb
31	<b>antage</b>	assume	0.80	Verb
31	<u>antagelig</u>	presumably (ADV), permissible (ADJ)	0.60	Adverb, Adjective
31	<u>antagelse</u>	assumption	0.70	Noun
32	<b>antydde</b>	indicate	0.80	Verb
32	<u>antydning</u>	hint, suggestion, implication	0.70	Noun
33	<b>anvendelig</b>	usable	0.80	Adjective
33	<u>anvendelighed</u>	applicability, usability, usefulness	0.70	Noun
33	<u>anvendelse</u>	application, use, employment	0.70	Noun
34	<b>arv</b>	inheritance	0.80	Noun
34	<u>arvelig</u>	hereditary	0.60	Adjective
35	<b>aspekt</b>	aspect	0.80	Noun
36	<b>baggrund</b>	background	0.80	Noun
36	<i>baggrundsinformation</i>	background information	0.60	Noun
36	<i>baggrundsviden</i>	background knowledge	0.60	Noun
37	<b>balancere</b>	balance	0.80	Verb
37	<i>balancegang</i>	balancing act	0.60	Noun
38	<b>basere</b>	base	0.80	Verb
38	<b>basis</b>	basis	0.80	Noun
38	<u>basal</u>	basic, fundamental	0.70	Adjective

39	<b>bearbejdning</b>	processing	0.80	Noun
39	<u>bearbejdelse</u>	processing	0.60	Noun
40	<b>bedømme</b>	assess	0.80	Verb
40	<b>bedømmelse</b>	assessment	0.80	Noun
40	<u>dømme</u>	judge	0.70	Verb
41	<b>befolkning</b>	population	0.80	Noun
41	<u>befolke</u>	populate	0.70	Verb
41	<i>befolkningsgruppe</i>	population base, section of the population	0.60	Noun
41	<i>befolkningssammensætning</i>	composition of population, demographic structure	0.60	Noun
41	<i>landbefolkning</i>	rural population	0.70	Noun
42	<b>befordre</b>	promote	0.80	Verb
43	<b>begge</b>	both	0.80	Pronoun
44	<b>begrænse</b>	limit	0.80	Verb
44	<b>begrænsning</b>	limitation	0.80	Noun
45	<b>begyndelse</b>	beginning	0.80	Noun
45	<b>påbegynde</b>	start	0.80	Verb
45	<b>påbegyndelse</b>	commencement	0.80	Noun
46	<b>behandle</b>	treat	0.80	Verb
47	<b>behov</b>	need	0.80	Noun
48	<b>bekostning</b>	cost	0.80	Noun
50	<b>bekræfte</b>	confirm	0.80	Verb
50	<u>bekræftelse</u>	confirmation	0.60	Noun
51	<b>belyse</b>	illustrate	0.80	Verb
52	<b>bemærkelsesværdig</b>	remarkable	0.80	Adjective
53	<b>nytte</b>	use	0.80	Noun
53	<b>nytte</b>	be of use	0.80	Verb
53	<b>benytte</b>	use	0.80	Verb
53	<i>nytteværdi</i>	usefulness, utility value	0.70	Noun
53	<u>udnytte</u>	utilize, exploit	0.70	Verb
54	<b>benævne</b>	designate	0.80	Verb
54	<b>nævne</b>	mention	0.80	Verb
54	<b>nævneværdig</b>	mentionable	0.80	Adjective
54	<b>sidstnævnte</b>	the latter	0.80	Adjective
54	<u>benævnelse</u>	designation	0.60	Noun
54	<u>førnævnt</u>	before-mentioned	0.60	Adjective
54	<i>førstnævnte</i>	first mentioned	0.70	Adjective
55	<b>berige</b>	enrich	0.80	Verb
56	<b>berøre</b>	affect	0.80	Verb
56	<u>berøring</u>	touch, contact	0.60	Noun
56	<u>uberørt</u>	untouched	0.70	Adjective
57	<b>besiddelse</b>	possession	0.80	Noun
58	<b>beskeden</b>	modest	0.80	Adjective
59	<b>beskrive</b>	describe	0.80	Verb
59	<b>beskrivelse</b>	description	0.80	Noun



59	<u>ubeskreven</u>	undescribed	0.60	Adjective
60	<b>beskæftige</b>	employ	0.80	Verb
60	<u>beskæftigelse</u>	employment	0.60	Noun
61	<b>beslutning</b>	decision	0.80	Noun
61	<i>beslutningsproces</i>	decision process	0.70	Noun
61	<i>beslutningstager</i>	decision-maker	0.70	Noun
62	<b>bestemme</b>	decide	0.80	Verb
62	<b>bestemt</b>	certainly	0.80	Adjective
62	<u>ubestemt</u>	undetermined	0.60	Adjective
62	<u>forudbestemt</u>	predetermined	0.60	Adjective
62	<i>selvbestemmelse</i>	self-determination, autonomy	0.60	Noun
63	<b>bestå</b>	exist	0.80	Verb
64	<b>betegne</b>	denote	0.80	Verb
64	<b>betegnelse</b>	designation	0.80	Noun
64	<b>tegn</b>	sign	0.80	Noun
64	<b>optegnelse</b>	note	0.80	Noun
64	<u>aftegne</u>	draw	0.70	Verb
64	<u>optegne</u>	record	0.60	Verb
65	<b>betinge</b>	determine	0.80	Verb
65	<u>betingelse</u>	condition, term	0.60	Noun
66	<b>betragte</b>	regard	0.80	Verb
66	<b>betragtelig</b>	considerable	0.80	Adjective
66	<b>betragtning</b>	consideration	0.80	Noun
67	<b>betyde</b>	mean	0.80	Verb
67	<b>betydning</b>	meaning	0.80	Noun
67	<b>betydningsfuld</b>	important	0.80	Adjective
67	<b>ubetydelig</b>	insignificant	0.80	Adjective
67	<u>betydelig</u>	significant	0.70	Adjective
67	<i>betydningsdannelse</i>	meaning-making, signification	0.60	Noun
68	<b>bevirke</b>	cause	0.80	Verb
68	<b>medvirke</b>	contribute	0.80	Verb
68	<b>påvirke</b>	influence	0.80	Verb
68	<u>indvirkning</u>	impact	0.70	Noun
68	<u>påvirkelig</u>	impressionable, susceptible to influence	0.60	Adjective
68	<u>påvirkning</u>	influence	0.60	Noun
68	<u>upåvirket</u>	unaffected	0.70	Adjective
69	<b>bevæge</b>	move	0.80	Verb
69	<b>bevægelse</b>	movement	0.80	Noun
69	<u>bevægelighed</u>	mobility, movability	0.60	Noun
70	<b>bevågenhed</b>	attention	0.80	Noun
71	<b>bibeholde</b>	maintain	0.80	Verb
72	<b>bidrag</b>	contribution	0.80	Noun
72	<b>bidrage</b>	contribute	0.80	Verb
73	<b>binde</b>	bind	0.80	Verb

73	<u>bind</u>	volume	0.60	Noun
73	<u>bindeled</u>	connecting link	0.70	Noun
73	<u>binding</u>	bond, tie	0.70	Noun
73	<b>forbinde</b>	connect	0.80	Verb
73	<b>forbindelse</b>	connection	0.80	Noun
73	<u>forbunden</u>	connected	0.70	Adjective
74	<b>binær</b>	binary	0.80	Adjective
75	<b>bl.a.</b>	among other	0.80	Abbreviation
76	<b>blot</b>	only	0.80	Adverb
77	<b>bortset</b>	apart from	0.80	Verb
78	<b>bred</b>	broad	0.80	Adjective
78	<b>bredde</b>	width	0.80	Noun
79	<b>brugbar</b>	useful	0.80	Adjective
79	<b>brugbarhed</b>	usefulness	0.80	Noun
80	<b>både</b>	both	0.80	Konjunction
81	<b>central</b>	central	0.80	Adjective
82	<b>cirkel</b>	circle	0.80	Noun
82	<b>cirkulation</b>	circulation	0.80	Noun
82	<u>cirkulere</u>	circulate	0.70	Verb
83	<b>dagligdags</b>	everyday	0.80	Adjective
84	<b>danne</b>	create	0.80	Verb
85	<b>dansk</b>	Danish	0.80	Adjective
85	<i>dansksproget</i>	Danish-speaking	0.60	Adjective
86	<b>daværende</b>	then	0.80	Adjective
87	<b>definere</b>	define	0.80	Verb
87	<b>definition</b>	definition	0.80	Noun
87	<b>redefinere</b>	redefine	0.80	Verb
87	<u>omdefinering</u>	redefining	0.60	Noun
87	<u>veldefineret</u>	well-defined	0.70	Adjective
88	<b>del</b>	part	0.80	Noun
88	<b>dels</b>	partly	0.80	Konjunction
88	<b>delvis</b>	partially	0.80	Adjective
89	<b>den</b>	it	0.80	Pronoun
89	<b>denne</b>	this	0.80	Pronoun
90	<b>deres</b>	their	0.80	Pronoun
91	<b>deraf</b>	thereof	0.80	Adverb
91	<b>derfor</b>	therefore	0.80	Adverb
91	<b>derimod</b>	however	0.80	Adverb
91	<b>dermed</b>	thus	0.80	Adverb
91	<b>dernæst</b>	next	0.80	Adverb
91	<b>dertil</b>	besides	0.80	Adverb
91	<b>derved</b>	thereby	0.80	Adverb
92	<b>desto</b>	the	0.80	Adverb
93	<b>detaljeret</b>	detailed	0.80	Adjective

93	<i>detaljerigdom</i>	wealth of detail	0.60	Noun
93	<i>detaljeringsgrad</i>	level of detail	0.60	Noun
94	<b>differentiere</b>	differentiate	0.80	Verb
94	<u>differentiering</u>	differentiation	0.60	Noun
95	<b>direkte</b>	direct	0.80	Adjective
95	<b>indirekte</b>	indirect	0.80	Adjective
96	<b>diskussion</b>	discussion	0.80	Noun
96	<b>diskutere</b>	discuss	0.80	Verb
96	<b>omdiskuteret</b>	controversial	0.80	Adjective
96	<u>indiskutabel</u>	indisputable	0.70	Adjective
97	<b>disposition</b>	outline	0.80	Noun
98	<b>dobbelt</b>	double	0.80	Adjective
98	<u>dobbelthed</u>	doubleness, duality, duplicity	0.60	Noun
98	<i>dobbeltrølle</i>	dual role	0.60	Noun
98	<u>fordobling</u>	doubling	0.60	Noun
99	<b>dog</b>	however	0.80	Adverb
100	<b>dokumentere</b>	document	0.80	Verb
100	<u>dokument</u>	document	0.60	Noun
100	<u>dokumentation</u>	documentation	0.70	Noun
101	<u>dominans</u>	dominance	0.70	Noun
102	<b>dominere</b>	dominate	0.80	Verb
103	<b>dreje</b>	be about sth	0.80	Verb
103	<u>drejning</u>	turn	0.60	Noun
104	<b>dvs.</b>	i.e.	0.80	Abbreviation
105	<b>dynamik</b>	dynamics	0.80	Noun
105	<u>dynamisk</u>	dynamic	0.60	Adjective
106	<b>efterlade</b>	leave behind	0.80	Verb
107	<b>efterspørge</b>	demand	0.80	Verb
108	<b>egen</b>	own	0.80	Adjective
109	<b>egenskab</b>	attribute	0.80	Noun
110	<b>ej</b>	no	0.80	Adverb
111	<b>eksakt</b>	exact	0.80	Adjective
112	<b>eksempel</b>	example	0.80	Noun
112	<b>eksempelvis</b>	as an example	0.80	Adverb
112	<u>eksemplificere</u>	exemplify	0.70	Verb
113	<b>eksistere</b>	exist	0.80	Verb
113	<u>eksistens</u>	existence	0.70	Noun
114	<b>eller</b>	or	0.80	Konjunction
114	<b>én</b>	one	0.80	Pronoun
115	<b>end</b>	than	0.80	Conjunction/Adverb
116	<b>endelig</b>	finally	0.80	Adverb/Adjective
116	<b>uendelig</b>	infinite	0.80	Adjective
117	<b>endog</b>	even	0.80	Adverb
118	<b>endvidere</b>	moreover	0.80	Adverb

119	<b>ene</b>	alone	0.80	Adjective
120	<b>enighed</b>	agreement	0.80	Noun
120	<u>uenig</u>	disagreeing	0.70	Adjective
120	<u>uenighed</u>	disagreement	0.60	Noun
121	<b>enkelt</b>	simply	0.80	Adjective
121	<b>enkeltvis</b>	individually	0.80	Adverb
121	<i>enkeltindivid</i>	single individual	0.70	Noun
121	<i>enkeltstående</i>	isolated	0.70	Adjective
121	<u>forenkling</u>	simplification	0.60	Noun
122	<b>enten</b>	either	0.80	Konjunction
123	<b>entydig</b>	unambiguous	0.80	Adjective
124	<b>erfaring</b>	experience	0.80	Noun
124	<u>erfare</u>	learn, experience, discover	0.70	Verb
124	<i>arbejdserfaring</i>	professional experience	0.60	Noun
125	<b>erkende</b>	acknowledge	0.80	Verb
125	<u>erkendelse</u>	recognition	0.70	Noun
125	<i>erkendelsesteori</i>	cognition theory	0.60	Noun
126	<b>erstatte</b>	replace	0.80	Verb
127	<b>essentiell</b>	essential	0.80	Adjective
127	<u>essens</u>	essence	0.70	Noun
128	<b>etablere</b>	establish	0.80	Verb
128	<b>veletableret</b>	well-established	0.80	Adjective
128	<u>etablering</u>	establishment, construction	0.70	Noun
129	<b>etc.</b>	etc.	0.80	Abbreviation
130	<b>eventuel</b>	possible	0.80	Adjective
130	evt. (forkortelse)	potential	0.70	Abbreviation
131	<b>evne</b>	ability	0.80	Noun
131	<b>evne</b>	be able to	0.80	Verb
132	<b>fald</b>	fall	0.80	Noun
133	<b>fase</b>	phase	0.80	Noun
134	<b>fastholde</b>	maintain	0.80	Verb
134	<b>fastholdelse</b>	insistence	0.80	Noun
134	<b>fastslå</b>	demonstrate, establish	0.80	Verb
135	<b>findes</b>	be	0.80	Verb
136	<b>fjerde</b>	fourth	0.80	Numeral
137	<b>flere</b>	several	0.80	Adjective
137	<u>flerhed</u>	plurality	0.60	Noun
138	<b>flyde</b>	flow	0.80	Verb
139	<b>fokus</b>	focus	0.80	Noun
139	<b>fokusere</b>	focus	0.80	Verb
139	<u>fokusering</u>	focusing	0.70	Noun
139	<i>fokusområde</i>	area of focus	0.60	Noun
139	<i>hovedfokus</i>	main focus	0.60	Noun
140	<b>forblive</b>	stay	0.80	Verb

141	<b>fordelagtig</b>	advantageous	0.80	Adjective
142	<b>forefinde</b>	find	0.80	Verb
143	<b>foregå</b>	take place	0.80	Verb
143	<u>foregående</u>	previous	0.70	Adjective
144	<b>forekomme</b>	occur	0.80	Verb
145	<b>forelæsning</b>	lecture	0.80	Noun
145	<i>tiltrædelsesforelæsning</i>	inaugural lecture	0.60	Noun
146	<b>foreskrive</b>	prescribe	0.80	Verb
147	<b>foreslå</b>	suggest	0.80	Verb
148	<b>foretage</b>	undertake	0.80	Verb
149	<b>forhold</b>	condition	0.80	Noun
149	<u>forholde</u>	relate, withhold,	0.70	Verb
149	<u>forholdsvis</u>	proportionate, proportional, comparative, relative	0.70	Adjective
150	<b>forklare</b>	explain	0.80	Verb
150	<b>forklaring</b>	explanation	0.80	Noun
150	<i>forklaringskraft</i>	explanatory power	0.60	Noun
151	<b>form</b>	form	0.80	Noun
151	<u>omforme</u>	convert	0.60	Verb
151	<u>omformning</u>	conversion	0.60	Noun
152	<b>formel</b>	formula	0.80	Adjective
152	<u>formalisere</u>	formalise	0.70	Verb
152	<u>uformel</u>	informal	0.70	Adjective
153	<b>formel</b>	formal	0.80	Noun
154	<b>formentlig</b>	supposed	0.80	Adjective
155	<b>formode</b>	suppose	0.80	Verb
155	<b>formodning</b>	presumption	0.80	Noun
156	<b>formulering</b>	formulation	0.80	Noun
156	<u>formulere</u>	formulate	0.70	Verb
156	<u>omformulere</u>	reformulate	0.60	Verb
157	<b>formål</b>	purpose	0.80	Noun
157	<i>hovedformål</i>	primary purpose	0.60	Noun
158	<b>fornemmelse</b>	sense	0.80	Noun
159	<b>forskel</b>	difference	0.80	Noun
159	<b>forskellig</b>	different	0.80	Adjective
159	<u>forskelligartet</u>	different, varied, diverse, heterogeneous	0.70	Adjective
159	<u>forskellighed</u>	difference, dissimilarity, heterogeneity, diversity, variety	0.70	Noun
160	<b>forskningsprojekt</b>	research project	0.80	Noun
160	<u>forsker</u>	researcher, scientist	0.70	Noun
160	<u>forskning</u>	research	0.70	Noun
160	<u>forskningsmæssig</u>	research-wise	0.60	Adjective
160	<i>forskningsindsats</i>	research effort	0.70	Noun
160	<i>forskningsinterview</i>	research interview	0.60	Noun
160	<i>forskningslitteratur</i>	research literature, scholarly literature	0.60	Noun

160	<i>forskningsmiljø</i>	research environment	0.60	Noun
160	<i>forskningsoversigt</i>	research review	0.60	Noun
160	<i>forskningspraksis</i>	research practice	0.60	Noun
160	<i>grundforskning</i>	basic research	0.70	Noun
160	<i>seniorforsker</i>	senior researcher, senior research associate	0.70	Noun
161	<b>forskydning</b>	displacement	0.80	Noun
161	<u>forskyde</u>	displace, dislocate	0.70	Verb
162	<b>forstyrre</b>	disrupt	0.80	Verb
163	<b>forstærke</b>	strengthen	0.80	Verb
164	<b>forstå</b>	understand	0.80	Verb
164	<b>forståelig</b>	comprehensible	0.80	Adjective
164	<b>forståelse</b>	understanding	0.80	Noun
164	<u>forståelsesramme</u>	frame for understanding	0.60	Noun
164	<u>forforståelse</u>	preunderstanding	0.70	Noun
164	<u>indforstået</u>	informed	0.60	Adjective
164	<u>misforståelse</u>	misunderstanding	0.70	Noun
164	<u>uforståelig</u>	incomprehensible	0.60	Adjective
164	<u>underforstå</u>	imply	0.60	Verb
164	<i>selvforståelse</i>	self-knowledge	0.70	Noun
164	<i>verdensforståelse</i>	world understanding	0.60	Noun
165	<b>forsøg</b>	experiment	0.80	Noun
166	<b>fortrinsvis</b>	preferential	0.80	Adverb
167	<b>fortsætte</b>	continue	0.80	Verb
167	<b>fortsættelse</b>	continuation	0.80	Noun
168	<b>forud</b>	ahead	0.80	Adverb
168	<b>foruden</b>	besides	0.80	Preposition
168	<b>forudgående</b>	preceding	0.80	Adjective
169	<b>forudsætning</b>	basis	0.80	Noun
169	<b>forudsætte</b>	assume	0.80	Verb
170	<b>forvalte</b>	manage	0.80	Verb
170	<u>forvalter</u>	administrator, manager, overseer	0.60	Noun
170	<u>forvaltning</u>	administration, management	0.60	Noun
170	<u>forvaltningsmæssig</u>	administrative, management	0.60	Adjective
171	<b>forvejen</b>	ahead	0.80	Noun
172	<b>forveksle</b>	confound	0.80	Verb
172	<b>veksle</b>	change	0.80	Verb
172	<u>forveksling</u>	confusion, mix-up, mistake	0.60	Noun
172	<u>veksel</u>	exchange	0.60	Noun
172	<i>vekselvirkning</i>	interplay, interaction	0.60	Noun
172	<u>veksler</u>	switchboard	0.70	Noun
172	<u>udveksle</u>	exchange	0.70	Verb
172	<u>udveksling</u>	exchange	0.70	Noun
173	<b>forventelig</b>	expected	0.80	Adjective
173	<u>forvente</u>	expect, anticipate	0.70	Verb

174	<b>frembringe</b>	produce	0.80	Verb
174	<u>frembringelse</u>	production, generation	0.60	Noun
175	<b>frembyde</b>	offer	0.80	Verb
175	<b>fremgå</b>	appear	0.80	Verb
176	<b>fremhæve</b>	emphasise	0.80	Verb
177	<b>fremkomme</b>	appear	0.80	Verb
177	<u>fremkomst</u>	appearance, emergence	0.70	Noun
178	<b>fremlægge</b>	present	0.80	Verb
178	<u>fremlæggelse</u>	presentation	0.60	Noun
179	<b>fremmest</b>	foremost	0.80	Adjective
179	<u>fremme</u>	ahead	0.70	Adverb
179	<u>fremme</u>	promotion, encouragement, advancement	0.70	Noun
179	<u>fremme</u>	promote, further, advance	0.70	Verb
180	<b>fremstille</b>	represent	0.80	Verb
180	<u>fremstilling</u>	description, account	0.70	Noun
181	<b>fremtidig</b>	future	0.80	Adjective
182	<b>fremtrædende</b>	salient	0.80	Adjective
182	<u>fremtræde</u>	appear, emerge	0.70	Verb
183	<b>fundament</b>	foundation	0.80	Noun
183	<b>fundamental</b>	fundamental	0.80	Adjective
184	<b>funktion</b>	function	0.80	Noun
184	<u>funktionel</u>	functional	0.70	Adjective
184	<i>funktionsmåde</i>	functioning, mode of operation	0.60	Noun
184	<i>servicefunktion</i>	service function	0.60	Noun
185	<b>fx</b>	for example	0.80	Abbreviation
186	<b>fælles</b>	common	0.80	Adjective
186	<i>fællestræk</i>	common feature	0.70	Noun
187	<b>færdighed</b>	skill	0.80	Noun
187	<i>læsefærdighed</i>	reading skill	0.60	Noun
188	<b>følge</b>	follow	0.80	Noun
188	<b>følge</b>	sequence	0.80	Verb
188	<b>efterfølge</b>	succeed	0.80	Verb
188	<u>følgelig</u>	consequently	0.70	Adverb
188	<i>følgevirkning</i>	consequence, effect, outcome	0.60	Noun
189	<b>føre</b>	carry	0.80	Verb
189	<b>anføre</b>	state	0.80	Verb
189	<b>indføre</b>	introduce	0.80	Verb
189	<u>indføring</u>	introduction	0.60	Noun
189	<u>indførsel</u>	import	0.60	Noun
189	<u>medføre</u>	entail	0.70	Verb
189	<b>videreføre</b>	continue	0.80	Verb
189	<u>videreførelse</u>	continuation	0.70	Noun
190	<b>føromtalt</b>	above-mentioned	0.80	Adjective
191	<b>først</b>	first	0.80	Adverb

192	<b>ganske</b>	quite	0.80	Adjective
193	<b>generalisere</b>	generalise	0.80	Verb
193	<u>generaliserbarhed</u>	generalizability	0.60	Noun
193	<u>generalisering</u>	generalization	0.60	Noun
193	<b>generel</b>	general	0.80	Adjective
194	<b>generere</b>	generate	0.80	Verb
195	<b>genfinde</b>	recover	0.80	Verb
196	<b>gengive</b>	render	0.80	Verb
196	<u>gengivelse</u>	reproduction	0.60	Noun
197	<b>genkende</b>	recognise	0.80	Verb
197	<u>genkendelig</u>	recognizable, identifiable	0.70	Adjective
198	<b>gennem</b>	through	0.80	Preposition
199	<b>gennemgribende</b>	thorough	0.80	Adjective
200	<b>gennemgående</b>	common	0.80	Adjective
200	<u>gennemgå</u>	undergo	0.70	Verb
201	<b>gennemskære</b>	intersect	0.80	Verb
202	<b>gensidig</b>	mutual	0.80	Adjective
203	<b>genstand</b>	object	0.80	Noun
203	<i>genstandsfelt</i>	domain	0.60	Noun
204	<b>gentage</b>	repeat	0.80	Verb
205	<b>given</b>	given	0.80	Adjective
205	<u>givetvis</u>	certainly	0.70	Adverb
206	<b>grad</b>	extent, degree	0.80	Noun
206	<i>gradsforskel</i>	varying degree	0.60	Noun
206	<u>graduere</u>	graduate, grade	0.60	Verb
206	<b>gradvis</b>	gradual	0.80	Adverb
207	<b>granskning</b>	scrutiny	0.80	Noun
208	<b>grunde</b>	base	0.80	Verb
208	<u>begrunde</u>	motivate	0.70	Verb
208	<u>begrundelse</u>	motivation	0.70	Noun
208	<u>ubegrundet</u>	unfounded	0.60	Adjective
208	<u>velbegrundet</u>	well-founded	0.70	Adjective
209	<b>grundlag</b>	basis	0.80	Noun
209	<u>grundliggende</u>	basic, fundamental, underlying	0.60	Adjective
209	<u>grundlægge</u>	found, establish	0.70	Verb
209	<u>grundlæggelse</u>	foundation, establishment	0.60	Noun
209	<u>grundlæggende</u>	fundamental, basic, underlying	0.70	Adjective
209	<u>tilgrundliggende</u>	underlying	0.70	Adjective
210	<b>grænse</b>	limit	0.80	Noun
210	<i>grænseområde</i>	margin, periphery	0.60	Noun
210	<u>tilgrænsende</u>	adjacent	0.60	Adjective
211	<b>gyldighed</b>	validity	0.80	Noun
212	<b>gælde</b>	be valid	0.80	Verb
212	<b>pågældende</b>	in question	0.80	Adjective



213	<b>handlekraft</b>	resourcefulness	0.80	Noun
213	<u>handle</u>	act	0.70	Verb
213	<u>handling</u>	action, act	0.70	Noun
213	<i>handlemønster</i>	response pattern	0.60	Noun
213	<i>handlemåde</i>	course of action, behaviour	0.60	Noun
213	<i>handlingsanvisning</i>	action plan	0.60	Noun
214	<b>hastig</b>	hurried	0.80	Adjective
215	<b>helhed</b>	whole	0.80	Noun
216	<b>henblik</b>	regard	0.80	Noun
217	<b>henholdsvis</b>	respectively	0.80	Adverb
217	<u>henholde</u>	take one's stand on	0.70	Verb
217	hhv.	respectively	0.60	Abbreviation
218	<b>hensigt</b>	intention	0.80	Noun
218	<b>hensigtsmæssig</b>	appropriate	0.80	Adjective
218	<i>hensigtserklæring</i>	declaration of intent	0.60	Noun
218	<u>uhensigtsmæssig</u>	inappropriate	0.70	Adjective
219	<b>hensyn</b>	consideration	0.80	Noun
219	<i>hensyntagen</i>	consideration	0.70	Noun
220	<b>henvise</b>	refer	0.80	Verb
220	<u>henvisning</u>	reference, referral	0.70	Noun
221	<b>heraf</b>	hereof	0.80	Adverb
222	<b>herfor</b>	for this	0.80	Adverb
223	<b>heri</b>	herein	0.80	Adverb
224	<b>hermed</b>	herewith	0.80	Adverb
225	<b>herpå</b>	subsequently	0.80	Adverb
226	<b>hertil</b>	for this purpose, here	0.80	Adverb
227	<b>heterogenitet</b>	heterogeneity	0.80	Noun
228	<b>hithtidig</b>	hitherto	0.80	Adjective
229	<b>hidtil</b>	so far	0.80	Adverb
230	<b>hinanden</b>	each other	0.80	Pronoun
231	<b>hobe</b>	heap	0.80	Verb
231	<u>ophobe</u>	accumulate	0.60	Verb
231	<u>ophobning</u>	accumulation	0.60	Noun
232	<b>holdning</b>	attitude	0.80	Noun
232	<i>grundholdning</i>	basic position	0.60	Noun
233	<b>hovedelement</b>	main element	0.80	Noun
234	<b>hverken</b>	either	0.80	Konjunction
235	<b>hvile</b>	rest	0.80	Noun
235	<b>hvile</b>	rest	0.80	Verb
236	<b>hvilken</b>	which	0.80	Pronoun
237	<b>hvordan</b>	how	0.80	Adverb
238	<b>hvorfra</b>	where	0.80	Adverb
239	<b>hvor</b>	where	0.80	Adverb
240	<b>hvorimod</b>	whereas	0.80	Adverb

241	<b>hvorledes</b>	how	0.80	Adverb
242	<b>hvormed</b>	with what	0.80	Adverb
243	<b>hvorunder</b>	under which	0.80	Adverb
244	<b>hvorved</b>	at which	0.80	Adverb
245	<b>hvorvidt</b>	whether	0.80	Adverb
246	<b>høj</b>	high	0.80	Adjective
247	<b>håndgribelig</b>	tangible	0.80	Adjective
248	<b>identisk</b>	identical	0.80	Adjective
249	<b>idet</b>	as	0.80	Konjunction
250	<b>ifølge</b>	according to	0.80	Preposition
251	<b>igangværende</b>	in progress	0.80	Adjective
252	<b>illustrere</b>	illustrate	0.80	Verb
252	<u>illustration</u>	illustration	0.70	Noun
253	<b>imidlertid</b>	however	0.80	Adverb
253	<b>impuls</b>	impulse	0.80	Noun
254	<b>imødekomme</b>	oblige	0.80	Verb
255	<b>indblanding</b>	intervention	0.80	Noun
256	<b>indblik</b>	insight	0.80	Noun
257	<b>indbyrdes</b>	mutual, reciprocal	0.80	Adjective
258	<b>inddrage</b>	implicate	0.80	Verb
259	<b>indebære</b>	imply	0.80	Verb
260	<b>indeholde</b>	contain	0.80	Verb
261	<b>indflydelse</b>	influence	0.80	Noun
261	<u>indflydelsesrig</u>	influential	0.60	Adjective
262	<b>indgå</b>	enter	0.80	Verb
262	<b>indgående</b>	thoroughly	0.80	Adjective
263	<b>indhold</b>	content	0.80	Noun
263	<u>indholdsmæssig</u>	content-wise, in terms of contents, in relation to content	0.60	Adjective
264	<b>indicium</b>	indication	0.80	Noun
264	<u>indicere</u>	indicate	0.60	Verb
264	<u>indikator</u>	indicator	0.60	Noun
265	<b>indviduel</b>	individual	0.80	Adjective
265	<u>indvid</u>	individual	0.60	Noun
265	<u>individualitet</u>	individuality	0.60	Noun
267	<b>indledningsvis</b>	by way of introduction	0.80	Adverb
268	<b>indlysende</b>	obvious, evident	0.80	Adjective
269	<b>indsigt</b>	insight	0.80	Noun
270	<b>indskrænkning</b>	reduction, restriction	0.80	Noun
270	<u>indskrænke</u>	limit, restrict, reduce	0.70	Verb
271	<b>indtage</b>	take in	0.80	Verb
271	<u>indtog</u>	entry	0.70	Noun
272	<b>indtryk</b>	impression	0.80	Noun
272	<i>hovedindtryk</i>	main impression	0.60	Noun
273	<b>indvandring</b>	immigration	0.80	Noun

273	<u>indvandre</u>	immigrate	0.70	Verb
273	<u>indvandrer</u>	immigrant	0.60	Noun
275	<b>initiativ</b>	initiative	0.80	Noun
275	<u>initiere</u>	initiate	0.70	Verb
276	<b>institut</b>	department	0.80	Noun
276	<u>institutionalisere</u>	institutionalize	0.60	Verb
276	<u>institutionalisering</u>	institutionalization	0.60	Noun
276	<u>institutionel</u>	institutional	0.60	Adjective
277	<b>integrere</b>	integrate	0.80	Verb
277	<u>integration</u>	integration	0.60	Noun
278	<b>intensivere</b>	intensify	0.80	Verb
278	<u>intensivering</u>	intensification	0.60	Noun
278	<u>intensitet</u>	intensity	0.60	Noun
278	<u>intensiv</u>	intensive	0.60	Adjective
279	<b>intention</b>	intention	0.80	Noun
280	<b>interaktion</b>	interaction	0.80	Noun
280	<u>interagere</u>	interact	0.70	Verb
280	<u>interagerende</u>	interacting	0.60	Adjective
281	<b>interessant</b>	interesting	0.80	Adjective
281	<b>interesse</b>	interest	0.80	Noun
281	<u>interessent</u>	stakeholder	0.70	Noun
281	<i>interessefelt</i>	field of interest, area of interest	0.70	Noun
281	<i>interesseorganisation</i>	interest group, interest organisation	0.70	Noun
281	<u>uinteressant</u>	uninteresting	0.70	Adjective
282	<b>intern</b>	internal	0.80	Adjective
282	<u>internalisere</u>	internalize	0.60	Verb
283	<b>international</b>	international	0.80	Adjective
284	<b>interview</b>	interview	0.80	Noun
284	<u>interviewer</u>	interviewer	0.70	Noun
284	<i>interviewguide</i>	interview guide	0.60	Noun
284	<i>interviewundersøgelse</i>	interview study	0.70	Noun
285	<b>introducere</b>	introduce	0.80	Verb
285	<b>introduktion</b>	introduction	0.80	Noun
286	<b>involvere</b>	involve	0.80	Verb
286	<u>involvering</u>	involvement	0.60	Noun
287	<b>isolere</b>	isolate	0.80	Verb
288	<b>især</b>	particular	0.80	Adverb
289	<b>iværksætte</b>	implement	0.80	Verb
290	jf	compare	0.60	Abbreviation
290	<b>jf.</b>	cf.	0.80	Abbreviation
290	jvf	compare	0.60	Abbreviation
291	<b>kalde</b>	call	0.80	Verb
291	<u>fremkalde</u>	induce	0.60	Verb
291	<u>påkalde</u>	invoke	0.60	Verb

292	<b>karakterisere</b>	characterise	0.80	Verb
292	<b>karakteristisk</b>	characteristical	0.80	Adjective
292	<u>karakter</u>	character, nature, grade	0.70	Noun
292	<u>karakteristik</u>	characteristic	0.70	Noun
292	<u>karakteristikon</u>	characteristic	0.70	Noun
293	<b>kategori</b>	category	0.80	Noun
293	<b>kategorisere</b>	categorise	0.80	Verb
293	<u>kategorisering</u>	categorization	0.70	Noun
293	<u>kategorisk</u>	categoric, categorical	0.70	Adjective
294	<b>kendskab</b>	knowledge	0.80	Noun
294	<i>kendetegn</i>	distinctive feature	0.70	Noun
295	<b>kerne</b>	core	0.80	Noun
295	<i>kernebegreb</i>	key concept	0.60	Noun
295	<i>kerneværdi</i>	core value	0.70	Noun
296	<b>kilde</b>	source	0.80	Noun
297	<b>klarhed</b>	clarity	0.80	Noun
297	<i>klarlægge</i>	clarify, explain	0.60	Verb
297	<b>uklar</b>	indistinct	0.80	Adjective
297	<b>uklarhed</b>	indistinctness	0.80	Noun
298	<b>klassificere</b>	classify	0.80	Verb
299	<b>klassisk</b>	classic	0.80	Adjective
300	<b>knytte</b>	bind	0.80	Verb
300	<b>tilknytte</b>	attach	0.80	Verb
300	<u>tilknytning</u>	attachment, association	0.60	Noun
301	<b>koble</b>	link	0.80	Verb
301	<u>kobling</u>	linking, linkage	0.70	Noun
301	<u>sammenkoble</u>	link	0.70	Verb
302	<b>kompensere</b>	compensate	0.80	Verb
303	<b>kompliceret</b>	complicated	0.80	Adjective
303	<u>ukompliceret</u>	uncomplicated	0.60	Adjective
304	<b>koncentrere</b>	concentrate	0.80	Verb
305	<b>konflikt</b>	conflict	0.80	Noun
306	<b>konkludere</b>	conclude	0.80	Verb
306	<b>konklusion</b>	conclusion	0.80	Noun
307	<b>konkret</b>	concrete	0.80	Adjective
307	<u>konkretisere</u>	concretize, clarify, flesh out	0.60	Verb
307	<u>konkretisering</u>	clarification, concretization	0.60	Noun
308	<b>konkurrere</b>	compete	0.80	Verb
309	<b>konsekvens</b>	consequence, consistency	0.80	Noun
309	<u>konsekvent</u>	consistent	0.60	Adjective
310	<b>konsistent</b>	consistant	0.80	Adjective
311	<b>konstatere</b>	ascertain	0.80	Verb
311	<b>konstatering</b>	ascertainment	0.80	Noun
312	<b>konstruere</b>	construct	0.80	Verb

312	<b>rekonstruere</b>	reconstruct	0.80	Verb
312	<u>konstruktion</u>	construction, design, structure	0.60	Noun
312	<u>konstruktiv</u>	constructive	0.70	Adjective
312	<u>rekonstruktion</u>	reconstruction	0.70	Noun
313	<b>kontur</b>	contour	0.80	Noun
313	<u>konvention</u>	convention	0.60	Noun
313	<b>konventionel</b>	conventional	0.80	Adjective
314	<b>koordinat</b>	coordinate	0.80	Noun
314	<u>koordinering</u>	co-ordination	0.60	Noun
315	<b>krav</b>	demand	0.80	Noun
316	<b>krise</b>	crisis	0.80	Noun
317	<b>kriterium</b>	criterion	0.80	Noun
318	<b>kritisk</b>	critical	0.80	Adjective
318	<u>kritik</u>	criticism, critique	0.70	Noun
318	<u>kritiker</u>	critic	0.60	Noun
318	<u>kritisere</u>	criticize	0.70	Verb
318	<i>selvkritik</i>	self-criticism	0.60	Noun
318	<i>samfundskritik</i>	social criticism	0.60	Noun
319	<b>kulminere</b>	culminate	0.80	Verb
319	<u>kulmination</u>	culmination	0.60	Noun
320	<b>langtfra</b>	far from it	0.80	Adverb
321	<b>leder</b>	leader	0.80	Noun
321	<u>ledelsesmæssig</u>	managerial, in terms of management	0.60	Adjective
321	<i>ledetråd</i>	guiding principle	0.60	Noun
322	<b>levende</b>	alive	0.80	Adjective
322	<i>levedygtig</i>	viable, sustainable	0.60	Adjective
322	<i>levetilstand</i>	living conditions	0.70	Noun
322	<i>levevis</i>	way of life, way of living	0.70	Noun
323	<b>lig</b>	like	0.80	Adjective
323	<b>lighed</b>	similarity	0.80	Noun
323	<b>lign.</b>	similar	0.80	Abbreviation
323	<u>ligelig</u>	equal	0.70	Adjective
323	<u>ulig</u>	unlike	0.70	Adjective
323	<i>ligeværdighed</i>	equal opportunities, equal status	0.70	Noun
324	<b>ligeledes</b>	as well	0.80	Adverb
325	<b>ligesom</b>	like	0.80	CONounJ/Adverb
326	<b>lille</b>	small	0.80	Adjective
327	<b>litteratur</b>	literature	0.80	Noun
327	<i>litteraturliste</i>	bibliography, reading list, list of references	0.70	Noun
328	<b>m.fl.</b>	and others	0.80	Abbreviation
329	<b>mainstream</b>	mainstream	0.80	Noun
330	<b>mangel</b>	lack	0.80	Noun
330	<b>mangle</b>	lack	0.80	Verb
330	<b>mangfoldighed</b>	diversity	0.80	Noun

330	<u>mangelfuld</u>	faulty, flawed, defective, insufficient	0.70	Adjective
330	<u>mangfoldig</u>	manifold , multiple , multitudinous	0.70	Adjective
331	<b>markant</b>	marked	0.80	Adjective
331	<u>markere</u>	mark, indicate	0.70	Verb
331	<u>markering</u>	marking, indication	0.60	Noun
331	<u>markør</u>	cursor, pointer, marker	0.60	Noun
332	<b>med</b>	with	0.80	Preposition
333	<u>medens (alternative spelling)</u>	while	0.70	Konjunction
334	<b>mellem</b>	between	0.80	Preposition
335	<b>men</b>	but	0.80	Konjunction
336	<b>mens</b>	while	0.80	Konjunction
337	<b>mente</b>	number carried	0.80	Noun
338	<b>mere</b>	more	0.80	Adjective
339	<b>metode</b>	method	0.80	Noun
339	<b>metodisk</b>	methodical	0.80	Adjective
339	<u>metodik</u>	method, methodology	0.70	Noun
339	<u>metodologi</u>	methodology	0.60	Noun
339	<u>metodologisk</u>	methodological	0.60	Adjective
340	<b>ministeriel</b>	ministerial	0.80	Adjective
341	<b>model</b>	model	0.80	Noun
341	<u>modellere</u>	model	0.60	Verb
341	<u>modellering</u>	modelling	0.60	Noun
342	<b>modificere</b>	modify	0.80	Verb
342	<b>modifikation</b>	modification	0.80	Noun
343	<b>modsat</b>	opposite	0.80	Adjective
343	<i>modsatrettet</i>	opposing	0.70	Adjective
344	<b>modsætning</b>	contrast	0.80	Noun
344	<u>modsætningsfyldt</u>	contradictory, incompatible, contrasting	0.60	Adjective
344	<u>modsætte</u>	oppose	0.70	Verb
345	<b>mulig</b>	possible	0.80	Adjective
345	<b>muliggøre</b>	make possible, permit	0.80	Verb
345	<i>mulighedsrum</i>	room of opportunity	0.60	Noun
345	<u>umuliggøre</u>	render impossible	0.60	Verb
346	<b>munde</b>	result in	0.80	Verb
347	<b>mønster</b>	pattern	0.80	Noun
348	<b>måde</b>	way	0.80	Noun
349	<b>mål</b>	goal	0.80	Noun
349	<b>målestok</b>	scale	0.80	Noun
349	<u>målbar</u>	measurable	0.70	Adjective
349	<u>måle</u>	measure	0.60	Verb
349	<u>målelig</u>	measurable	0.60	Adjective
349	<i>målgruppe</i>	target audience	0.60	Noun
349	<i>målrette</i>	target	0.70	Verb
349	<i>målsætning</i>	objective, target	0.60	Noun

350	<b>national</b>	national	0.80	Adjective
350	<u>nation</u>	nation	0.60	Noun
350	<u>nationalistisk</u>	nationalistic	0.60	Adjective
350	<i>nationalstat</i>	nation state	0.60	Noun
351	<b>natur</b>	nature	0.80	Noun
351	<b>naturlig</b>	natural	0.80	Adjective
351	<u>naturalistisk</u>	naturalistic	0.60	Adjective
351	<i>naturforhold</i>	nature, natural conditions	0.60	Noun
351	<i>naturkraft</i>	natural force, force of nature	0.60	Noun
352	<b>nedenfor</b>	below	0.80	Adverb
353	<b>nedskrive</b>	write down	0.80	Verb
354	<b>negativ</b>	negativ	0.80	Adjective
354	<u>negation</u>	negation	0.60	Noun
355	<b>negligere</b>	neglect	0.80	Verb
356	<b>neutral</b>	neutral	0.80	Adjective
356	<u>neutralisere</u>	neutralize	0.60	Verb
356	<u>neutralitet</u>	neutrality	0.70	Noun
357	<b>niveau</b>	level	0.80	Noun
358	<b>nogenlunde</b>	tolerable	0.80	Adjective
359	<b>nordamerikansk</b>	north american	0.80	Adjective
360	<b>nuanceret</b>	varied	0.80	Adjective
360	<b>nuancering</b>	making sth varied	0.80	Noun
360	<u>nuancere</u>	vary	0.70	Verb
360	<u>unuanceret</u>	undifferentiated, unnuanced	0.60	Adjective
361	<b>nær</b>	near	0.80	Adjective/Preposition/Adverb
361	<b>nærliggende</b>	nearby	0.80	Adjective
361	<u>nærværende</u>	present	0.70	Adjective
362	<b>nødvendig</b>	necessary	0.80	Adjective
362	<b>nødvendigvis</b>	necessarily	0.80	Adverb
362	<u>nødvendiggøre</u>	necessitate	0.70	Verb
362	<u>nødvendighed</u>	necessity	0.70	Noun
363	<b>nøgleord</b>	keyword	0.80	Noun
364	<b>obligatorisk</b>	mandatory	0.80	Adjective
364	<b>observation</b>	observation	0.80	Noun
365	<b>offentlig</b>	public	0.80	Adjective
365	<u>offentliggøre</u>	publish	0.60	Verb
365	<u>offentliggørelse</u>	publication	0.70	Noun
366	<b>officiel</b>	official	0.80	Adjective
367	<b>ofte</b>	often	0.80	Adverb
367	<b>oftest</b>	most often	0.80	Adverb
368	<b>om</b>	about	0.80	Preposition/Conjunction
369	<b>omfang</b>	extent	0.80	Noun
369	<u>omfangsrig</u>	extensive	0.60	Adjective

370	<b>omfattende</b>	extensive, comprehensive	0.80	Adjective
370	<b>sammenfatte</b>	summarise	0.80	Verb
370	<b>opfatte</b>	perceive	0.80	Verb
370	<b>opfattelse</b>	understanding	0.80	Noun
370	<u>omfatte</u>	include	0.70	Verb
370	<u>sammenfatning</u>	summary	0.60	Noun
370	<i>selvopfattelse</i>	self-perception, self-image	0.60	Noun
371	<b>omgang</b>	turn	0.80	Noun
372	<b>omgive</b>	surround	0.80	Verb
372	<b>omgivelse</b>	surroundings	0.80	Noun
373	<b>omhandle</b>	concern	0.80	Verb
374	<b>omskrive</b>	rewrite	0.80	Verb
374	<b>omskrivning</b>	revision	0.80	Noun
375	<b>omtale</b>	comment	0.80	Noun
375	<b>omtale</b>	comment on	0.80	Verb
376	<b>omvende</b>	convert	0.80	Verb
377	<b>omverden</b>	surrounding world	0.80	Noun
378	<b>opbygge</b>	construct	0.80	Verb
378	<u>opbygning</u>	structure	0.70	Noun
379	<b>opdagelse</b>	discovery	0.80	Noun
380	<b>opdele</b>	divide up	0.80	Verb
381	<b>operere</b>	operate	0.80	Verb
381	<u>operationel</u>	operational	0.70	Adjective
381	<u>operativ</u>	operational	0.60	Adjective
382	<b>oplagt</b>	obviously	0.80	Adjective
385	<u>opleve</u>	experience	0.70	Verb
385	<b>oplevelse</b>	experience	0.80	Noun
385	<i>sanseoplevelse</i>	sensory experience	0.60	Noun
386	<b>opløse</b>	dissolve	0.80	Verb
386	<u>opløsning</u>	dissolution	0.70	Noun
387	<b>opmærksomhed</b>	attention	0.80	Noun
388	<b>opretholde</b>	sustain	0.80	Verb
388	<b>opretholdelse</b>	maintenance	0.80	Noun
389	<b>oprindelig</b>	original	0.80	Adjective
389	<u>oprindelse</u>	origin, source	0.70	Noun
390	<b>opstå</b>	arise	0.80	Verb
391	<b>opsummere</b>	sum up	0.80	Verb
391	<u>opsummering</u>	summary	0.70	Noun
391	<u>summarisk</u>	summary	0.70	Adjective
391	<u>summere</u>	summarize	0.70	Verb
392	<b>optræde</b>	appear	0.80	Verb
393	<b>organisere</b>	organise	0.80	Verb
393	<b>organisering</b>	organising	0.80	Noun
393	<u>organisation</u>	organisation	0.60	Noun



393	<u>organisatorisk</u>	organizational	0.70	Adjective
393	<u>omorganisering</u>	reorganisation	0.60	Noun
393	<u>reorganisere</u>	reorganize	0.60	Verb
393	<u>reorganisering</u>	reorganization	0.70	Noun
393	<u>organisme</u>	organism	0.70	Noun
393	<i>organisationsstruktur</i>	organisational structure	0.60	Noun
394	<b>orientere</b>	orientate	0.80	Verb
394	<b>orientering</b>	orientation	0.80	Noun
394	<i>markedsorienteret</i>	market-oriented	0.70	Adjective
395	<b>osv.</b>	etc.	0.80	Abbreviation
396	<b>ovenfor</b>	above	0.80	Adverb
397	<b>ovenstående</b>	the above	0.80	Adjective
398	<b>overdrive</b>	exaggerate	0.80	Verb
399	<b>overens</b>	similar	0.80	Adjective
399	<b>overensstemmelse</b>	accordance	0.80	Noun
399	<u>overensstemmende</u>	concordant, corresponding	0.60	Adjective
399	<u>uoverensstemmelse</u>	discrepancy	0.60	Noun
400	<b>overfladisk</b>	superficial	0.80	Adjective
401	<b>overfor</b>	opposite	0.80	Adverb
402	<b>overgang</b>	passage	0.80	Noun
403	<b>overleve</b>	survive	0.80	Verb
403	<u>overlevelse</u>	survival	0.70	Noun
404	<b>overordnet</b>	superior, general	0.80	Adjective
404	<u>overordentlig</u>	extraordinary	0.70	Adjective
405	<b>overveje</b>	consider	0.80	Verb
405	<b>overvejelse</b>	consideration	0.80	Noun
406	<u>overvejende</u>	predominant	0.70	Adjective
406	<b>parallel</b>	parallel	0.80	Adjective
406	<b>parallel</b>	parallel	0.80	Noun
407	<b>passiv</b>	passive	0.80	Adjective
408	<b>pege</b>	point	0.80	Verb
408	<b>påpege</b>	indicate	0.80	Verb
408	<b>udpege</b>	designate	0.80	Verb
409	<b>perspektiv</b>	perspective	0.80	Noun
409	<b>perspektivere</b>	put into perspective	0.80	Verb
409	<u>perspektivering</u>	perspectivation	0.70	Noun
409	<i>tidsperspektiv</i>	time perspective	0.70	Noun
410	<b>placere</b>	place	0.80	Verb
410	<b>placering</b>	placement	0.80	Noun
411	<b>planlægge</b>	plan	0.80	Verb
411	<u>planlægning</u>	planning	0.60	Noun
412	<b>positiv</b>	positive	0.80	Adjective
412	<u>positivistisk</u>	positivistic	0.60	Adjective
413	<b>potentiel</b>	potential	0.80	Adjective

413	<u>potentiale</u>	potential	0.70	Noun
414	<b>praksis</b>	practice	0.80	Noun
414	<u>praksisnær</u>	practice-oriented, practice-based	0.60	Adjective
415	<b>praktisk</b>	practical	0.80	Adjective
415	<u>praktiker</u>	practician	0.60	Noun
416	<b>primær</b>	primary	0.80	Adjective
417	<b>princip</b>	principle	0.80	Noun
417	<u>principiel</u>	in principle	0.70	Adjective
417	<i>grundprincip</i>	basic principle	0.60	Noun
418	<b>prioritering</b>	prioritization	0.80	Noun
418	<u>prioritere</u>	prioritize	0.70	Verb
419	<b>problem</b>	problem	0.80	Noun
419	<b>problematisk</b>	problematic	0.80	Adjective
419	<b>problemstilling</b>	problem	0.80	Noun
419	<u>problematik</u>	problem	0.60	Noun
419	<u>problematisering</u>	problematization	0.60	Noun
419	<u>problemløs</u>	unproblematic	0.60	Adjective
419	<u>uproblematisk</u>	unproblematic	0.70	Adjective
419	<i>samfundsproblem</i>	social problem	0.60	Noun
420	<b>proces</b>	process	0.80	Noun
420	<u>procedure</u>	procedure	0.70	Noun
420	<i>arbejdsproces</i>	work process	0.60	Noun
420	<i>omstillingsproces</i>	readjustment process	0.60	Noun
421	<b>produktiv</b>	productive	0.80	Adjective
421	<u>producere</u>	produce	0.70	Verb
421	<u>produktivitet</u>	productivity	0.60	Noun
421	<u>biprodukt</u>	by-product	0.60	Noun
421	<u>reproducere</u>	reproduce	0.60	Verb
421	<u>reproduktion</u>	reproduction	0.70	Noun
421	<i>produktionsbetingelse</i>	production conditions	0.60	Noun
421	<i>produktionsforhold</i>	conditions of production, production environment	0.60	Noun
421	<i>produktudvikling</i>	product development	0.60	Noun
422	<b>præference</b>	preference	0.80	Noun
422	<i>smagspræference</i>	taste preference	0.60	Noun
423	<b>præg</b>	character	0.80	Noun
423	<u>præge</u>	mark	0.70	Verb
424	<b>publicere</b>	publish	0.80	Verb
425	<b>punkt</b>	point	0.80	Noun
425	<b>udgangspunkt</b>	starting point	0.80	Noun
425	<i>omdrejningspunkt</i>	central point, focal point, pivotal point	0.60	Noun
425	<i>startpunkt</i>	starting point, point of departure	0.60	Noun
426	<b>påkrævet</b>	required	0.80	Adjective
427	<u>rationale</u>	rationale	0.60	Noun
427	<b>rationel</b>	rational	0.80	Adjective

428	<b>reel</b>	real	0.80	Adjective
428	<u>realisere</u>	realize	0.70	Verb
428	<u>realisering</u>	realization	0.60	Noun
428	<u>realistisk</u>	realistic	0.70	Adjective
428	<u>realitet</u>	reality, fact	0.70	Noun
429	<b>redegøre</b>	give an account of	0.80	Verb
430	<b>redskab</b>	tool	0.80	Noun
431	<b>reference</b>	reference	0.80	Noun
431	<i>referencepunkt</i>	reference point, point of reference	0.60	Noun
431	<i>referenceramme</i>	frame of reference	0.70	Noun
431	<u>referere</u>	refer, give an account of	0.70	Verb
432	<b>regel</b>	rule	0.80	Noun
432	<i>regelsæt</i>	code of practice, protocol, regulatory framework	0.60	Noun
433	<b>regi</b>	framework	0.80	Noun
434	<b>relatere</b>	relate	0.80	Verb
434	<b>relativ</b>	relative	0.80	Adjective
434	<u>relation</u>	relation	0.70	Noun
434	<u>relationel</u>	relational	0.70	Adjective
435	<b>relevans</b>	relevance	0.80	Noun
435	<b>relevant</b>	relevant	0.80	Adjective
436	<b>repræsentativ</b>	representative	0.80	Adjective
436	<b>repræsentere</b>	represent	0.80	Verb
436	<u>repræsentant</u>	representative	0.60	Noun
436	<u>repræsentation</u>	representation	0.60	Noun
436	<u>repræsentativitet</u>	representativity	0.60	Noun
436	<u>overrepræsentation</u>	overrepresentation	0.60	Noun
437	<b>respektive</b>	respectively	0.80	Adjective
438	<b>ressource, resurse</b>	ressource	0.80	Noun
438	<i>ressourcekrævende</i>	require resources	0.60	Adjective
438	<i>Ressourcemæssig, resursekrævende</i>	in terms of resources	0.70	Adjective
438	<u>resursemæssig</u>	in terms of resources	0.70	Adjective
439	<b>resultat</b>	result	0.80	Noun
439	<u>resultere</u>	result	0.70	Verb
439	<i>hovedresultat</i>	main findings, key findings	0.60	Noun
439	<i>slutresultat</i>	final result	0.70	Noun
440	<b>retning</b>	direction	0.80	Noun
441	<b>rette</b>	right	0.80	Verb
441	<i>rettesnor</i>	guiding principle, benchmark	0.70	Noun
442	<b>revidere</b>	revise	0.80	Verb
443	<b>ringe</b>	bad	0.80	Adjective
444	<b>rolle</b>	role	0.80	Noun
444	<i>rollefordeling</i>	role assignment, distribution of roles, division of roles	0.70	Noun
444	<i>nøglerolle</i>	key role	0.60	Noun

445	<b>rumme</b>	hold	0.80	Verb
445	<u>rum</u>	room, space	0.70	Noun
445	<u>rumlig</u>	spatial	0.70	Adjective
445	<i>tomrum</i>	gap, void, vacuum	0.60	Noun
446	<b>rund</b>	round	0.80	Adjective
447	<b>række</b>	row	0.80	Noun
447	<b>række</b>	reach	0.80	Verb
447	<i>rækkefølge</i>	order	0.70	Noun
447	<u>vidtrækkende</u>	far-reaching	0.70	Adjective
448	<b>samme</b>	same	0.80	Adjective
449	<b>sammenholde</b>	relate	0.80	Verb
450	<b>sammenhæng</b>	context	0.80	Noun
450	<b>sammenhængende</b>	coherent	0.80	Adjective
450	<i>livssammenhæng</i>	current life situation, situation of life	0.60	Noun
451	<b>sammensætte</b>	compound, compile	0.80	Verb
451	<u>sammensat</u>	compound , composite	0.70	Adjective
451	<u>sammensætning</u>	composition	0.60	Noun
452	<b>sammentænke</b>	synthesise	0.80	Verb
453	<b>samspil</b>	interplay	0.80	Noun
454	<b>samtlige</b>	all	0.80	Adjective
454	<u>samt</u>	plus	0.70	Konjunction
455	<b>samtidig</b>	simultaneous	0.80	Adjective
455	<u>samtid</u>	contemporary, one's age	0.60	Noun
456	<b>sande</b>	admit	0.80	Verb
457	<b>sandsynligvis</b>	probably	0.80	Adverb
457	<u>sandsynliggøre</u>	render probable	0.60	Verb
457	<u>usandsynlig</u>	unlikely, improbable	0.70	Adjective
458	<b>scenario, scenarie</b>	scenario	0.80	Noun
459	<b>selve</b>	actual	0.80	Adjective
460	<b>selvfølgelig</b>	matter of course	0.80	Noun
461	<b>selvsagt</b>	obvious	0.80	Adjective
462	<b>sen</b>	late	0.80	Adjective
463	<b>sideløbende</b>	parallel	0.80	Adjective
464	<b>sig</b>	oneself	0.80	Pronoun
465	<b>sigt</b>	sight	0.80	Noun
465	<b>sigte</b>	aim	0.80	Noun
465	<b>sigte</b>	aim	0.80	Verb
465	<b>tilsigte</b>	aim at	0.80	Verb
466	<b>simpel</b>	simple	0.80	Adjective
467	<b>situation</b>	situation	0.80	Noun
467	<u>situationel</u>	situational	0.60	Adjective
467	<i>situationsfornemmelse</i>	sense of occasion, situational judgement	0.60	Noun
467	<i>livssituation</i>	current life situation, situation of life	0.60	Noun
468	<b>sjette</b>	sixth	0.80	NounUM

469	<b>sjælden</b>	rare	0.80	Adjective
469	<u>sjældenhed</u>	rarity	0.70	Noun
470	<b>skabe</b>	create	0.80	Verb
470	<b>skelne</b>	distinguish	0.80	Verb
470	<u>skel</u>	boundary	0.70	Noun
471	<b>skift</b>	change	0.80	Noun
472	<b>skitsere</b>	outline	0.80	Verb
472	<u>skitse</u>	outline, draft	0.70	Noun
473	<b>skyld</b>	owe	0.80	Verb
474	<b>skærpe</b>	sharpen	0.80	Verb
474	<b>slutning</b>	end	0.80	Noun
474	<u>sluttelig</u>	eventually, ultimately, finally, in conclusion	0.70	Adverb
474	<u>afslutning</u>	end, termination, completion, conclusion	0.70	Noun
474	<u>afslutningsvis</u>	finally, in conclusion, in closing	0.60	Adjective
474	<u>afslutte</u>	end, finish, conclude, terminate, finalize	0.70	Verb
475	<b>slås</b>	fight	0.80	Verb
476	<b>som</b>	which	0.80	Pronoun
477	<b>specifik</b>	specific	0.80	Adjective
477	<u>specificere</u>	specify	0.70	Verb
477	<u>uspecificeret</u>	unspecified	0.60	Adjective
477	<i>kønsspecifik</i>	gender specific	0.60	Adjective
478	<b>spekter, spektrum</b>	spectrum	0.80	Noun
478	<u>spektrum (alternativ stavemåde)</u>	spectrum	0.60	Noun
479	<b>spillereg</b>	rule of the game	0.80	Noun
479	<i>spillerum</i>	scope, leeway	0.60	Noun
480	<b>spinkel</b>	slight	0.80	Adjective
481	<b>spor</b>	track	0.80	Adjective
481	<b>spore</b>	monitor	0.80	Verb
481	<u>opspore</u>	track down	0.70	Verb
482	<b>spørgsmål</b>	question	0.80	Noun
482	<i>spørgsmålstegn</i>	question mark	0.60	Noun
483	<b>stadighed</b>	steadiness	0.80	Noun
484	<b>stamme</b>	origin	0.80	Noun
484	<b>stamme</b>	tribe	0.80	Verb
485	<b>stand</b>	condition	0.80	Noun
485	<u>tilstand</u>	condition	0.60	Noun
486	<b>status</b>	status	0.80	Noun
487	<b>stede</b>	present	0.80	Noun
488	<b>strategi</b>	strategy	0.80	Noun
488	<u>strategisk</u>	strategic	0.70	Adjective
489	<b>struktur</b>	structure	0.80	Noun
489	<u>strukturel</u>	structural	0.60	Adjective
489	<u>strukturere</u>	structure	0.70	Verb

489	<u>struktureret</u>	structured	0.60	Adjective
489	<u>infrastruktur</u>	infrastructure	0.60	Noun
489	<u>semistruktureret</u>	semi-structured	0.60	Adjective
490	<b>studere</b>	study	0.80	Verb
490	<u>studie</u>	study	0.60	Noun
490	<i>feltstudium</i>	field research	0.70	Noun
491	<b>styring</b>	administration	0.80	Noun
492	<b>styrke</b>	strength	0.80	Noun
492	<b>styrke</b>	strengthen	0.80	Verb
492	<u>styrkelse</u>	strengthening	0.70	Noun
493	<b>stærk</b>	strong	0.80	Adjective
494	<b>substans</b>	substance	0.80	Noun
494	<u>substantiel</u>	substantial	0.70	Adjective
495	<b>supplere</b>	supplement	0.80	Verb
495	<u>supplement</u>	supplement	0.70	Noun
495	<u>supplering</u>	supplementation	0.60	Noun
496	<b>svag</b>	weak	0.80	Adjective
496	<u>svaghed</u>	weakness	0.70	Noun
497	<b>svare</b>	answer	0.80	Verb
497	<b>modsvare</b>	correspond to	0.80	Verb
497	<u>modsvar</u>	response, retaliation	0.60	Noun
497	<i>svarmulighed</i>	answer, choice	0.60	Noun
497	<i>svarprocent</i>	response rate	0.60	Noun
497	<u>besvare</u>	respond	0.70	Verb
497	<u>besvarelse</u>	reply, solution	0.70	Noun
498	<b>svinge</b>	swing	0.80	Verb
498	<u>svingning</u>	swing, oscillation	0.60	Noun
499	<b>svække</b>	weaken	0.80	Verb
499	<u>svækkelse</u>	weakening	0.60	Noun
500	<b>symmetrisk</b>	symmetrical	0.80	Adjective
501	<b>synes</b>	think	0.80	Verb
502	<b>system</b>	system	0.80	Noun
502	<b>systematisere</b>	systematise	0.80	Verb
502	<u>systematik</u>	systematism	0.60	Noun
502	<u>systematisk</u>	systematic	0.70	Adjective
502	<u>usystematisk</u>	unsystematic	0.70	Adjective
503	<i>sædvane</i>	custom, standard practice	0.70	Noun
503	<b>sædvanligvis</b>	usually	0.80	Adverb
503	<u>usædvanlig</u>	unusual, uncommon, exceptional	0.70	Adjective
504	<b>særlig</b>	particular	0.80	Adjective
505	<b>særskilt</b>	separate	0.80	Adjective
506	<b>sådan</b>	such	0.80	Adjective
507	<b>såkaldt</b>	so-called	0.80	Adjective
508	<b>således</b>	thus	0.80	Konjunction

509	<b>såsom</b>	such as	0.80	Konjunction
510	<b>såvel</b>	as well as	0.80	Konjunction
511	<b>tal</b>	number	0.80	Noun
512	<b>tankegang</b>	mentality	0.80	Noun
513	<b>tema</b>	theme	0.80	Noun
513	<i>temanummer</i>	special issue, special feature issue	0.60	Noun
513	<u>tematik</u>	theme	0.60	Noun
514	<b>tendens</b>	tendency	0.80	Noun
514	<u>tendere</u>	tend	0.60	Verb
515	<b>tidlig</b>	early	0.80	Adjective
516	<b>tilbagevendende</b>	returning	0.80	Adjective
516	<b>tilblivelse</b>	birth	0.80	Noun
516	<b>tilbøjelig</b>	disposed	0.80	Adjective
516	tilbøjelighed	inclination, tendency, propensity	0.60	Noun
516	<b>tilfredsstille</b>	satisfy	0.80	Verb
516	<u>tilfredsstillelse</u>	satisfaction	0.60	Noun
516	<u>utilfredsstillende</u>	unsatisfactory	0.70	Adjective
517	<b>tilfælde</b>	case	0.80	Noun
517	<b>tilfældig</b>	accidental	0.80	Adjective
517	<u>tilfældighed</u>	chance, coincidence	0.70	Noun
518	<b>tilgang</b>	approach	0.80	Noun
519	<b>tilgængelig</b>	accessible	0.80	Adjective
520	<b>tilhøre</b>	belong to	0.80	Verb
520	<i>dertilhørende</i>	related	0.60	Adjective
521	<b>tilkendegivelse</b>	expression	0.80	Noun
521	<u>tilkendegive</u>	indicate, express	0.60	Verb
522	<b>tillid</b>	trust	0.80	Noun
522	<u>tillidsfuld</u>	trusting	0.60	Adjective
523	<b>tillige</b>	as well	0.80	Adverb
524	<b>tillægge</b>	ascribe to	0.80	Verb
524	<u>tillæg</u>	addition	0.60	Noun
525	<b>tilnærmelsesvis</b>	approximate	0.80	Adverb
525	<u>tilnærme</u>	approximate	0.60	Verb
526	<b>tilpasse</b>	adapt	0.80	Verb
526	<u>tilpasning</u>	adjustment	0.70	Noun
527	<b>tilsammen</b>	altogether	0.80	Adverb
528	<b>tilskrive</b>	attribute to	0.80	Verb
529	<b>tilskynde</b>	prompt	0.80	Verb
530	<b>tilstedeværelse</b>	presence	0.80	Noun
530	<u>tilstedeværende</u>	present, existing	0.70	Adjective
531	<b>tilstræbe</b>	aim to	0.80	Verb
531	<u>tilstræbt</u>	targeted, intended	0.60	Adjective
532	<b>tilstrækkelig</b>	sufficient	0.80	Adjective
533	<b>tilsvarende</b>	corresponding	0.80	Adjective

534	<b>tilsyneladende</b>	apparent	0.80	Adjective
535	<b>tiltag</b>	initiative	0.80	Noun
535	<b>tiltage</b>	increase	0.80	Verb
536	<b>tiltrække</b>	attract	0.80	Verb
536	<u>tiltrækning</u>	attraction	0.70	Noun
537	<b>to</b>	two	0.80	NounUM
537	<i>todelt</i>	split, divided, divided into two parts, bipartite	0.70	Adjective
538	<b>tolerance</b>	tolerance	0.80	Noun
538	<u>tolerere</u>	tolerate	0.60	Verb
539	<b>tolke</b>	interpret	0.80	Verb
539	<b>tolkning</b>	interpretation	0.80	Noun
539	<u>fortolke</u>	interpret	0.70	Verb
539	<u>fortolkning</u>	interpretation	0.70	Noun
539	<i>fortolkningsmulighed</i>	interpretation possibility	0.70	Noun
539	<u>genfortolke</u>	reinterpret	0.60	Verb
539	<u>nyfortolkning</u>	reinterpretation	0.60	Noun
540	<b>traditionel</b>	traditional	0.80	Adjective
540	<u>tradition</u>	tradition	0.70	Noun
540	<i>traditionsbestemt</i>	traditional	0.60	Adjective
541	<b>tre</b>	three	0.80	NounUM
541	<b>tredje</b>	third	0.80	NounUM
541	<i>tredeling</i>	tripartition, trisection, trichotomy	0.60	Noun
541	<i>tredelt</i>	tripartite, three-pronged	0.60	Adjective
541	<i>tredjedel</i>	third	0.60	Noun
541	<i>tredoble</i>	triple	0.60	Verb
542	<b>trods</b>	despite	0.80	Noun/Preposition
543	<b>troværdighed</b>	reliability	0.80	Noun
544	<b>tvungen</b>	compulsory	0.80	Adjective
545	<b>tværs</b>	cross	0.80	Adverb
545	<u>tvær</u>	cross	0.70	Adjective
546	<b>tyde</b>	interpret	0.80	Verb
546	<b>tydelig</b>	distinct	0.80	Adjective
546	<b>tydeliggøre</b>	elucidate	0.80	Verb
546	<u>tydeliggørelse</u>	clarification	0.60	Noun
546	<u>tydelighed</u>	clarity	0.60	Noun
546	<u>tydeligvis</u>	evidently	0.70	Adverb
546	<u>utydelig</u>	vague	0.70	Adjective
546	<i>flertydig</i>	ambiguous	0.70	Adjective
546	<i>flertydighed</i>	ambiguity	0.60	Noun
547	<b>type</b>	type	0.80	Noun
547	<b>typisk</b>	typical	0.80	Adjective
547	<u>typologi</u>	typology	0.60	Noun
547	<i>hovedtype</i>	principle type, basic type	0.70	Noun
547	<i>tidstypisk</i>	characteristic of the period/time	0.60	Adjective



548	<b>udbrede</b>	spread	0.80	Verb
548	<b><u>udbredelse</u></b>	spread, distribution	0.70	Noun
549	<b>uddybe</b>	clarify	0.80	Verb
549	<b><u>uddybning</u></b>	clarification, elaboration	0.70	Noun
550	<b>udefra</b>	outside	0.80	Adverb
551	<b>udelukke</b>	exclude	0.80	Verb
511	<b>udelukkende</b>	solely	0.80	Adjective
551	<b><u>udelukkelse</u></b>	exclusion	0.60	Noun
552	<b>udfordre</b>	challenge	0.80	Verb
552	<b>udfordring</b>	challenge	0.80	Noun
552	<b><u>udfordrer</u></b>	challenger	0.70	Noun
553	<b>udforme</b>	frame	0.80	Verb
553	<b>udformning</b>	version	0.80	Noun
554	<b>udforske</b>	explore	0.80	Verb
554	<b>udforskning</b>	exploration	0.80	Noun
555	<b>udfylde</b>	fill up	0.80	Verb
556	<b>udførelse</b>	execution	0.80	Noun
556	<b><u>udføre</u></b>	carry out, conduct	0.70	Verb
557	<b>udgøre</b>	constitute	0.80	Verb
558	<b>udlægge</b>	construe	0.80	Verb
558	<b><u>udlægning</u></b>	interpretation	0.60	Noun
559	<b>udredning</b>	explanation	0.80	Noun
560	<b>udsnit</b>	sample	0.80	Noun
561	<b>udtryk</b>	expression	0.80	Noun
561	<b>udtrykke</b>	express	0.80	Verb
562	<b>udvide</b>	extend	0.80	Verb
563	<b>udvikle</b>	develop	0.80	Verb
563	<b>udvikling</b>	development	0.80	Noun
563	<b>videreudvikle</b>	develop further	0.80	Verb
563	<b><u>uudviklet</u></b>	undeveloped	0.60	Adjective
563	<b><u>veludviklet</u></b>	well-developed	0.70	Adjective
563	<i>udviklingsarbejde</i>	development work, development effort	0.70	Noun
563	<i>udviklingshistorie</i>	history of development, evolutionary history	0.70	Noun
563	<i>udviklingsmulighed</i>	development opportunity, potential for development	0.60	Noun
563	<i>udviklingsorientere</i>	development orient	0.70	Verb
563	<i>udviklingsproces</i>	developmental proces	0.60	Noun
563	<i>udviklingstendens</i>	development tendency	0.60	Noun
563	<i>udviklingstrin</i>	developmental stage	0.60	Noun
563	<i>kompetenceudvikling</i>	professional development, career development	0.60	Noun
563	<b><u>videreudvikling</u></b>	further development	0.70	Noun
564	<b>udvælge</b>	select	0.80	Verb
565	<b>uegnet</b>	unfit	0.80	Adjective
566	<b>uforudset</b>	unforeseen	0.80	Adjective

567	<b>ukendt</b>	unknown	0.80	Adjective
567	<b>velkendt</b>	well-known	0.80	Adjective
568	<b>umiddelbar</b>	immediate	0.80	Adjective
569	<b>underbygge</b>	substantiate	0.80	Verb
570	<b>undergå</b>	undergo	0.80	Verb
571	<b>underliggende</b>	underlying	0.80	Adjective
571	<b>underlægge</b>	place under	0.80	Verb
572	<b>underordne</b>	subordinate	0.80	Verb
572	<u>underordnet</u>	secondary, subordinate	0.60	Adjective
573	<b>understrege</b>	emphasise	0.80	Verb
574	<b>understøtte</b>	support	0.80	Verb
575	<b>undersøge</b>	investigate	0.80	Verb
575	<u>undersøgelse</u>	examination, investigation, study	0.70	Noun
575	<i>undersøgelsesfelt</i>	study area, subject area of investigation , focus of investigation	0.60	Noun
575	<i>undersøgelsestidspunkt</i>	time of investigation, time of study	0.60	Noun
576	<b>undlade</b>	omit	0.80	Verb
577	<b>undtagelse</b>	exception	0.80	Noun
577	<u>undtagelsesvis</u>	unusually	0.60	Adverb
578	<b>univers</b>	universe	0.80	Noun
578	<u>universalitet</u>	universality	0.60	Noun
578	<u>universel</u>	universal	0.70	Adjective
579	<b>universitet</b>	university	0.80	Noun
579	<i>universitetsforlag</i>	university press, university publishing company	0.60	Noun
579	<i>universitetsniveau</i>	university level	0.60	Noun
579	<i>universitetsuddannelse</i>	university degree, university education	0.60	Noun
580	<b>utvivlsom</b>	undoubtedly	0.80	Adjective
581	<b>vanskelig</b>	difficult	0.80	Adjective
581	<b>vanskeliggøre</b>	complicate	0.80	Verb
582	<b>varetage</b>	attend to, manage	0.80	Verb
582	<u>varetagelse</u>	safeguarding, management, handling	0.60	Noun
583	<b>vedkommende</b>	concerned	0.80	Adjective
583	<b>vedr.</b>	pertaining to	0.80	Abbreviation
584	<b>vid</b>	wide	0.80	Adjective
584	<b>vidt</b>	far	0.80	Adverb
584	<b>vidtgående</b>	far-reaching	0.80	Adjective
585	<b>videnskabelig</b>	scientific	0.80	Adjective
585	<i>lægevidenskabelig</i>	medical, medical science	0.60	Adjective
585	<i>naturvidenskabelig</i>	natural science	0.70	Adjective
585	<i>populærvideenskabelig</i>	popular science	0.60	Adjective
585	<i>tværvideenskabelig</i>	interdisciplinary	0.70	Adjective
585	<u>videnskab</u>	science	0.70	Noun
585	<u>videnskabelighed</u>	science	0.60	Noun
585	<i>videnskabsmand</i>	scientist	0.60	Noun

585	<i>lægevidenskab</i>	medical science	0.60	Noun
585	<i>religionsvidenskab</i>	comparative religion	0.60	Noun
585	<i>vidensproduktion</i>	knowledge production	0.60	Noun
586	<b>viderebringe</b>	send forward	0.80	Verb
587	<b>vifte</b>	fan	0.80	Noun
588	<b>vige</b>	retreat	0.80	Verb
588	<u>afvige</u>	diverge, differ	0.70	Verb
588	<u>afvigelse</u>	diviation	0.60	Noun
588	<u>fravige</u>	deviate from	0.60	Verb
589	<b>vigtig</b>	important	0.80	Adjective
589	<b>vigtighed</b>	importance	0.80	Noun
590	<b>virkeliggøre</b>	realise	0.80	Verb
590	<u>virkelighed</u>	reality	0.70	Noun
591	<b>vis</b>	certain	0.80	Adjective
591	<u>uvis</u>	uncertain	0.70	Adjective
591	<u>vished</u>	certainty	0.60	Noun
592	<b>vis</b>	way	0.80	Noun
593	<b>vise</b>	show	0.80	Verb
593	<b>udvise</b>	display	0.80	Verb
593	<u>afvise</u>	reject	0.70	Verb
593	<u>afvisning</u>	rejection	0.60	Noun
593	<u>anvise</u>	shown, assign	0.60	Verb
593	<u>påvise</u>	demonstrate	0.60	Verb
593	<u>påvisning</u>	proof	0.60	Noun
593	<i>retvisende</i>	true, fair	0.60	Adjective
594	<b>visualisere</b>	visualise	0.80	Verb
595	<b>vor</b>	our	0.80	Pronoun
596	<b>vurdering</b>	assessment	0.80	Noun
596	<u>vurdere</u>	assess	0.60	Verb
596	<u>overvurdere</u>	overestimate	0.60	Verb
596	<u>undervurdere</u>	underestimate	0.70	Verb
596	<u>overvurdering</u>	overestimation	0.60	Noun
596	<u>revurdering</u>	reassessment	0.70	Noun
597	<b>vægt</b>	weight	0.80	Noun
597	<b>vægte</b>	weight	0.80	Verb
597	<u>vægtig</u>	weighty	0.60	Adjective
597	<u>vægtning</u>	weighting	0.70	Noun
597	<i>hovedvægt</i>	first priority	0.60	Noun
598	<b>værdi</b>	value	0.80	Noun
598	<b>værdifuld</b>	valuable	0.80	Adjective
598	<b>værdigrundlag</b>	fundamental values	0.80	Noun
598	<u>værdimæssig</u>	of value	0.60	Adjective
598	<i>værdikonflikt</i>	value conflict, conflict of values	0.60	Noun
598	<i>værdisætning</i>	valuation	0.60	Noun

598	<i>værdisætte</i>	value, estimate, assess the value of,	0.70	Verb
598	<i>signalværdi</i>	signal value, signalling value	0.70	Noun
599	<b>værdsætte</b>	appreciate	0.80	Verb
600	<b>væsentlig</b>	essential	0.80	Adjective
600	<u>væsentlighed</u>	essentiality	0.60	Noun
600	<u>uvæsentlig</u>	inessential	0.70	Adjective
601	<b>yderlig</b>	extreme	0.80	Adjective
601	<b>ydermere</b>	further	0.80	Adverb
601	<u>ydre</u>	external	0.70	Adjective
602	<b>ønskelig</b>	desirable	0.80	Adjective
602	<u>ønskværdig</u>	desirable	0.70	Adjective
603	<b>øvrig</b>	besides	0.80	Adjective
604	<b>årsag</b>	cause	0.80	Noun
604	<b>hovedårsag</b>	main cause	0.80	Noun
605	<b>årti</b>	decade	0.80	Noun
605	<i>tiår</i>	decade	0.60	Noun