Antonyms in Context : A Corpus-Based Semantic Analysis of Swedish Descriptive Adjectives

Willners, Caroline

2001

TRAVAUX DE L'INSTITUT DE LINGUISTIQUE DE LUND 40

Antonyms in Context

A corpus-based semantic analysis of Swedish descriptive adjectives

Caroline Willners

LUND
UNIVERSITY

ISSN 0347-2558
ISBN 91-974116-1-2

Printed in Sweden
Universitetstryckeriet, Lunds universitet
Lund 2001

CONTENTS

Acknowledgements

Life is a journey, and so is the process of writing a dissertation. I have come upon many obstacles on the way, but also many rewarding events. The people who have helped me are many.

My supervisors, Åke Viberg and Jan-Olof Svantesson have been indispensable. Åke has guided me in the field of lexical semantics and been of great help in finding a balance between lexical semantics and corpus-based methods. He has also developed my early work with Swedish adjectives in WordNet into a large project with the aim to cover the base vocabulary of Swedish, which I find very exciting and look forward to keep on working with. Jan-Olof, with his stoical calmness, has patiently answered an endless number of questions and been a great source of encouragement.

Bengt Sigurd, Barbara Gawronska and Mats Eeg-Olofsson are all responsible for my ending up in computational linguistics; they trained me as an undergraduate and inspired me to become a researcher. They have kept on supporting me and I am deeply grateful to Bengt and Mats for their thorough reading and commenting on the manuscript for this book.

Anders Holtsberg developed correct statistical methods and made sure I understood what they did. I owe particular thanks to Carita Paradis and Åsa Wengelin, for reading and commenting on parts of the book and for being great sources of encouragement and inspiration. I am indebted to "the SWordNet Girls": Ann Lindvall, Teresa Johansson, Kerstin Lindmark and Ingmarie Mellenius for many valuable discussions and comments on early manuscripts. Johan Segerbäck improved the readability of the book with great commitment and also provided many insightful comments and suggestions.

I have had many inspiring discussions with Birgitta Lastow and Johan Dahl, who have also been responsible for "computer support" in a wide sense. Olof Ekedahl helped me with the data from the BNC. Daniel Ridings kindly supplied material from the Parole corpus. Britt Nordbeck and Ingrid Mellqvist have helped me with a million practicalities.

Many people at the department have inspired me, helped me with practicalities, and been great partners in discussions (both at the seminars and around the coffee table): Niclas Burenhult, Anna Flyman, Elisabeth Zetterholm, Victoria Johansson, Sven Strömqvist, Eva Gårding, Gösta Bruce, Merle Horne, Gao Hong, Arthur Holmer, Karina Vamling, Lars-Åke Henningsson, Marcus

Unesson, Marianne Gullberg, Vivan Franzén, Paula Kuylenstierna, and many others.

I would also like to express my gratitude to my family for putting up with me and my piles of papers during the past years. My mother, Veronica, for her fantastic support; my sister, Josefine, for delivering an endless number of timely words, my favourite niece and nephew, Emma and Olle, for many laughters and insightful comments; and my brother in law, Per Anders, for taking such good care of them.

Per, my life companion and second spine, had endless patience with the diagrams in this book, with numerous computer problems, and with me.

My deepest thanks to all of you.

Parts of this work have been carried out within the projects Swedish WordNet (*Svenskt OrdNät*) and Crosslinguistic lexicology (*Tvärspråklig lexikologi*), both financially supported by the Swedish Council for Research in the Humanities and Social Sciences.

I dedicate this work to my grandfather, Harry Willners, who earned his hat but never got it, and to my father, Lars-Gösta Willners, who was a master of communication, with and without words. I regret that none of them got the opportunity to read this book.

Abbreviations

ABS	ABSTRACTION
ACT	HUMAN ACTION
ANI	ANIMAL
CREA	CREATURE
DEF	definite form
HUM	HUMAN
INDEF	indefinite form
MON	MONETARY REPRESENTATION
NEU	neuter
OBJ	OBJECT
ORG	ORGANISATION
PHEN	PHENOMENON
PL	plural
POSS	POSSESSION
PSY	PSYCHOLOGICAL FEATURE
SG	singular
SIT	SITUATION
UTR	non-neuter

Part I: Theoretical Background

1 Introduction

Researchers from various fields of study have wondered how things work inside the human brain. What is thought? What is emotion? What is language? These are extremely broad and general questions, and since I entered the field of linguistics my focus has narrowed bit by bit until I started asking questions such as: How do we know what different words mean? How do we select different words to use in different situations? How does input affect the acquisition of lexical relations?

The studies included in this book are attempts to shed light on a small portion of the last three questions. To further narrow down the field of study, I have chosen to focus on adjectives. I will discuss mainly the organisation of adjectives in the mental lexicon and explore what types of lexical relations between adjectives can be distinguished using corpus-based methods. The study is further limited to cover mainly descriptive and to some extent classifying adjectives.

Three separate studies are presented. The first study is an exploration of the mental lexicon, its organisation, and the acquisition of antonym relations between concepts.

The other studies focus on the semantic ranges of specific adjectives. Two case studies are presented, one describing and comparing the semantic ranges of two concepts in opposition: *full—tom* 'full'—'empty', and one comparing the semantic ranges of *stor* 'large' and a number of its synonyms.

The work has been developed within the framework of WordNet. The basic concepts and relations are the same, but new ideas have been implemented concerning semantic range. The lexical network has been further developed, and new methods for building it have been elaborated.

2 Some semantic definitions

"Semantically,
an adjective describes some important
but non-critical property of an object"
(Dixon 1977)

This chapter will introduce and define some semantic terms that will be used throughout the thesis. It will discuss the definitions and features of some semantic relations, especially antonymy.

2.1 Words and concepts

We know little about *how* words and their meaning(s) are represented in our minds, and this problem has occupied many great philosophers and linguists such as Kant and Jespersen. The approaches to meaning are many; a recent discussion by Frawley (1992) presents the following: meaning as reference, meaning as logical form, meaning as context and use, meaning as culture, and meaning as conceptual structure. All of these are serious attempts, but all have their drawbacks. I will follow here the WordNet view of how words and concepts are related. For more details on WordNet, see Chapter 3.

"Words express concepts" (Fellbaum 1998:8). A concept is represented by a set of synonyms that can be used to express that concept.

The mapping between word forms and concepts is of the many—many type – the members of a set of synonyms represent different ways to express the concept, while one word form can be related to several concepts (Miller et al. 1990:4).

2.2 Lexical fields

The concepts in our minds act on human needs and perceptual capabilities. Concepts are usually hard to describe and their boundaries are perceived differently by different individuals. But despite our differing opinions about what various concepts cover, we are able to communicate about them with one another. There is a consensus about roughly where the limits of a given concept

are drawn, what the core sense of that concept is, and to what other concepts it is related.

Colour terms are concepts that are obviously of the same kind and they stand in a natural relation to each other. It is a natural fact that mint green and olive green are more closely related than for example orange and blue (example from Dahllöf 1999). Many of us also have experienced differences of opinion on where the boundaries between colours go, sometimes ending up in ridiculous discussions about whether a curtain is brown or orange, etc.

The colour words all belong to the same *lexical field*. The key characteristic of a lexical field is that the meanings of the words within the field are interrelated. They cluster together to form fields of meaning which in turn cluster into larger fields.

2.2.1 Semantic scales

Many adjectives are gradable. Gradability means that there exists a *scale* on which the different comparison forms of an adjective correspond to different points or sections. For example, *strong* and *weak* are opposites on the scale of STRENGTH, but in a specific context, *stronger* may point to a place closer to *weak* than *strong* does. Usually there is a whole set of words that can be placed on the scale; on the scale of STRENGTH we also find for example *robust*, *powerful*, and *feeble*. The placing of the words on the scale is not fixed; it depends on the context. A certain person may be stronger than a weak person, but yet weaker than a third, feeble one, cf. Figure 1.

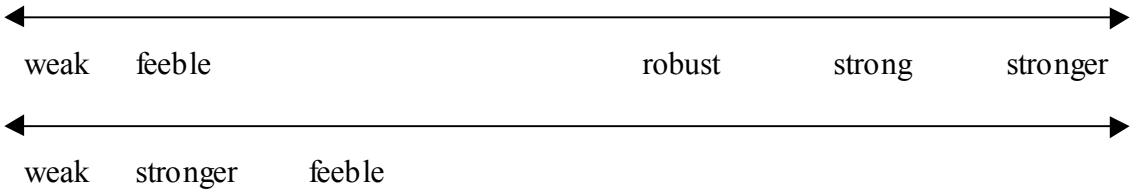


Figure 1: Two examples of some words on the scale of STRENGTH and how they can slide on the scale.

A scale is part of a lexical field. Scales are distinguished from ranks in that the lexemes on a scale are gradable (Lehrer 1974).

2.2.2 Semantic range

Words carry meaning, but the interpretation of a word depends on the context in which it appears. However, what types of words a certain word "keeps company with" (as Firth 1957 puts it) is determined by a finite set of semantic restrictions. This is the *semantic range* of the word. For an adjective, the semantic range is reflected by the set of nouns that it can modify. For example, a word like

himmelsvid ‘sky-wide = large’ modifies mainly ABSTRACTIONS, e.g. *himmelsvid skillnad* ‘big difference’, *himmelsvida avstånd* ‘large distances’, while its very frequent synonym *stor* can modify nouns from all different semantic categories.

2.2.3 Metaphors

Metaphors are general mappings across domains. Traditionally, the term ‘metaphor’ is used to refer to poetic metaphors, e.g. *death is the mother of beauty*, but metaphor is also a mechanism in everyday language (Lakoff 1992). Metaphors enable us to transfer structures from one domain to another and thus to look at a thing from another point of view or in a new costume. Metaphors are essential to our ability to think about abstract concepts such as time, change and causation.

In this book, metaphors will be used to explain how words undergo a semantic shift from a concrete meaning to an abstract but analogous meaning through metaphorical extension.

2.3 Semantic relations

2.3.1 Synonymy

Synonymy is the best-known semantic relation, and also a central relation in lexicography, but it is by no means easy to define. Lyons (1977:198) first defines synonyms as words sharing the same sense. Synonymy is then further restricted to obtain only if the words are substitutable for each other without affecting the descriptive meaning of the utterances (Lyons 1977:202). For two words to qualify as *absolute synonyms* they have to display complete identity of meaning and must be equinormal in all contexts. With these very severe requirements only few, if any, absolute synonyms exist (Cruse 2000:157, Miller & Fellbaum 1992:202). Miller & Fellbaum suggest a weakened, context-sensitive definition of synonymy: “two expressions are synonymous in a context *C* if the substitution of one for the other in *C* does not change the truth value.” This corresponds to Cruse’s (2000:158) *propositional synonyms*. His definition allows for differences in expressive meaning, stylistic level, and presupposed field of discourse.

From the WordNet view, synonymy is to be seen as a scale of similarity of meaning, where absolute synonymy is an end point from which the similarity of meaning gradually decreases (cf. Cruse 1986:265–291), cf Figure 2.

Figure 2: Scale of similarity of meaning.

Though it may be useful to think of synonymy as a scale of decreasing similarity we have no way of measuring the semantic distance between concepts on the scale and it is not adequate to say that synonyms are words whose meanings are relatively close, cf. Cruse (2000:159).

Within the EuroWordNet project[1] the following test sentences are used to determine synonymy in English (Climent et al. 1996):

> If it is (a/an) *X* then it is also (a/an) *Y*.
> If it is (a/an) *Y* then it is also (a/an) *X*.

The test requires both *X* and *Y* to be singular or plural nouns. Both test sentences have to be true. An example is:

> If it is a fiddle then it is also a violin.
> If it is a violin then it is also a fiddle.

Conceptual similarity between two words is enough; they do not have to correspond in morphosyntactic features, nor in register, style, or dialect. Using this liberal approach, words that may be quite different are grouped together according to the semantic kernel they have in common, e.g. *hoj* and *cykel* which both mean 'bicycle', though the former word is slang.

Synonyms, or what Cruse calls "propositional synonyms", are grouped in synonym sets, *synsets*, in WordNet. *Fiddle* and *violin* belong to the same synset, i.e. {fiddle, violin}, under the hypernym *bowed stringed instrument*.[2]

Near synonymy is another type of synonymy recognised by WordNet. According to Cruse (2000:159–160), near synonyms differ from propositional synonyms in that they may contrast with one another in certain contexts, e.g. *killed* and *murdered* in the example below.

> He was killed but I can assure you he was not murdered, madam.

Cruse (2000:160) lists the following minor differences that near synonyms may display:

- adjacent position on a scale of degree, e.g. *fog—mist*, *big—huge*;

[1] See Chapter 3.

[2] According to Princeton WordNet 1.6.

- certain adverbial specialisations of verbs, e.g. *amble—stroll*, *chuckle—giggle*;
- aspectual distinctions, e.g. *calm—placid*;
- difference of prototype centre, e.g. *brave* (prototypically physical)—*courageous* (prototypically psychological).

Near synonyms may also display a major distinction if it is backgrounded, e.g. *pretty—handsome*, which share the same propositional meaning 'good-looking' but differ in gender. *Pretty* is used about women, wheras *handsome* is used about men. In a context where the gender distinction is foregrounded, these words do not qualify as near synonyms, but in a context where it is backgrounded they do.

The EuroWordNet provides a link for near synonyms (Climent et al. 1996). It is possible to link closely related words that do not belong to the same synset with a NEAR_SYNONYM relation. For example, *garbage* and *trash* are co-ordinate sisters under the hypernym *waste* along with *pollutant*, *sewage*, etc. *Garbage* and *trash* are considered to be more closely related than the other co-ordinate sisters and are therefore linked with the NEAR_SYNONYM relation. Further criteria for the NEAR_SYNONYM relation are discussed by Climent et al. (1996).

2.3.2 Antonymy

The term *antonymy* was coined in the nineteenth century to describe oppositeness of meaning. A nice twitch to this is that the word itself was intended to be conceived as an antonym – of *synonymy*. However, antonymy does not really refer to the maximum degree of difference in meaning between two concepts. Rather, the words in an antonymous pair must be similar in all respects but one.

The term antonymy is used loosely about many different types of opposites. Egan (1968) distinguishes between seven different types of antonymy:

- *contradictory terms*: mutually exclusive terms with no intermediate possibilities, e.g. *perfect—imperfect*;
- *contrary terms*: true diametrical opposites but with possible values in between, e.g. *white—black*;
- *reverse terms*: words signalling opposition in the sense that they can reverse or undo an act, state, or quality, e.g. *constructive—deconstructive*;

- *contrasted terms*: words belonging to the same scale, but not designating the end points (cf. *indirect antonyms* in Chapter 4), e.g. *rich—destitute*;
- *incompatible* or *loosely contrasted terms*: words that stand in opposition but that do not "fully clash" because they do not share the same semantic range, e.g. *frank—hypocritical*;
- *relative terms*: word pairs indicating relations where one word cannot be used without implying the other, e.g. *parent—child*;
- *complementary terms*: reciprocal relations that imply each other, e.g. *question—answer*.

Egan further suggests the following definition of what an antonym is:

> "An antonym is a word so opposed in meaning to another word, its equal in breadth or range of application, that it negates or nullifies every single one of its implications."

She here touches upon an important feature of antonymy: the semantic ranges of the words involved in the antonymy relation must correspond. This feature tends to be overlooked in other definitions and characterisations of antonymy, of which some will be reviewed below. However, the importance of studying the semantic range of a word to learn its meaning will be shown later in the thesis.

Cruse (2000:167–172) distinguishes between *complementaries*, *antonyms*, *reversives*, and *converses*,[3] using the term *opposites* to encompass them all. Typical complementaries are *dead—alive*, *true—false*, *inside—outside*, *male—female*, and they are characterised by the reciprocal relation, i.e. in logic *f(x) entails and is entailed by not f(y)*.

Cruse (2000:169, 1986:204) follows Lyons in his definition of antonymy. He lists five different types of antonyms: *polar antonyms*, *equipollent antonyms*, *overlapping antonyms*, *reversives*, and *converses*. Only *polar antonyms* will be dealt with in this book.

Polar antonyms are fully gradable. They normally occur in comparative and superlative forms, which indicate degrees of some objective, unidimensional physical property. They are incompatibles, but not complementaries. The comparative forms of the word pair stand in a converse relationship. The comparative forms of both terms are impartial and one of the terms yields an impartial question in the frame *How X is it?* and an impartial nominalisation.

[3] For a discussion on converse terms in Swedish, see Sigurd 1976.

That the words must belong to the same scale implies that they must belong to the same semantic field, which is a criterion used by Lundbladh (1988) and Rusiecki (1985). According to them, antonyms (or *binary adjectives* as Rusiecki calls them) are:

- gradable;
- members of the same semantic field;
- incompatible; and
- at least semi-reciprocal.

Gradability, incompatibility and reciprocity will be further described below.

2.3.2.1 Gradability

Gradability is a fundamental semantic feature and a characteristic of descriptive adjectives. An adjective is gradable if it can be compared, either through morphological derivation, e.g. *happy—happier—happiest*, or by using degree modifiers such as *very*, e.g. *Emma is very happy*. Rusiecki (1985) gives the following phrase frames for identifying gradable adjectives:

Frame:	*Example:*
Aer (or: more A) than	*happier* than
as A as	as *happy* as
less A than	less *happy* than
the Aest (or: most A) of	the *happiest* of
very A	very *happy*

Many adjectives are gradable, but there are some semantic categories which are impossible to grade. Words that refer to a mathematical quality are not gradable, i.e. you cannot talk about to what extent a form is triangular, nor compare two triangular shapes and say that **A is more triangular than B*.

2.3.2.2 Incompatibility

Sisters or co-hyponyms such as *cat*, *dog*, *mouse*, and *rabbit* are traditionally said to be *incompatible*. They are incompatible since they contrast in a taxonomy. In the case of adjectives, two concepts are incompatible if there is nothing that can be modified by both concepts at the same time. Whereas it is easy to distinguish a cat from a dog, the border between gradable adjectives is relative and depends on the context.

Colour terms are a clear example of incompatibility. A car, for example, cannot be all *red* and all *black* at the same time, but it can be all *red* and all *new* at the same time.

2.3.2.3 Reciprocity

For an adjective pair where the words are gradable and incompatible there is a third crucial condition that the pair must meet to qualify as antonyms. They must be at least *semi-reciprocal*. Two gradable adjectives are *reciprocal* if they satisfy the following two entailment formulae:

NP_i is Aer than NP_j → NP_j is A'er than NP_i
NP_j is A'er than NP_i → NP_i is Aer than NP_j

NP_i and NP_j are noun phrases, and A and A' are gradable adjectives from the same lexical field, e.g. *short—long*.

If only one of the test sentences is satisfied, the adjective pair is *semi-reciprocal*, e.g. *economical—uneconomical*.

When I discuss antonymy in this book, the focus will be on polar antonymy. Word pairs designating the open ends of a scale are sometimes called true antonyms, but here they will be referred to as direct antonyms, the term used in the Princeton WordNet (Miller et al. 1990). Antonymy and the characteristics of antonymy will be further discussed in Chapter 4.

3 WordNet

WordNet is the main lexical framework used in this book. This chapter offers a short introduction to various wordnets and a more detailed description of how nouns are coded in the *Princeton WordNet*. The organisation of adjectives in WordNet is described in Chapter 4.

3.1 Introduction

WordNet is a lexical reference system designed to reflect the organisation of human memory as well as to be a useful on-line dictionary. The first WordNet was developed for English at Princeton University (Miller et al. 1990); presently there are WordNets under development for a number of languages: Swedish, Norwegian, Danish, Greek, Portuguese, Basque, Catalan, Romanian, Lithuanian, Russian, Bulgarian, and Slovene.[4]

WordNets for seven European languages (Dutch, Italian, Spanish, German, French, Czech, and Estonian) were developed between 1996 and 1999 in a multilingual database with wordnets of European languages: *The EuroWordNet Project*. The lexicons are structured in the same way as in the Princeton WordNet. In addition, the words are linked to an Inter-Lingual-Index, which makes it possible to use the dictionary for translation purposes.

In the middle of the 1990s, some Swedish adjectives were implemented in WordNet (Willners 1997) in an attempt to explore the possibility of building a Swedish version of WordNet. This pilot project developed into a large project, *Swedish WordNet* (Viberg 2000), which today involves several persons coding Swedish nouns, verbs, and adjectives in the Swedish WordNet (*SWordNet*) and connecting them to the Inter-Lingual-Index developed in the EuroWordNet project.

3.2 Nouns in WordNet

The studies in this book focus on descriptive adjectives but since adjectives are intimately connected with nouns, these cannot be completely disregarded. There is much work done on nouns, and much to say about them; however, I will

[4] According to the EuroWordNet homepage http://www.hum.uva.nl/~ewn/ (as of March 2001).

mention only how nouns are dealt with in the WordNet model and what semantic classes have been used in the empirical studies presented in this book.

The words in WordNet are linked by a number of hierarchical relations, e.g. *hyponymy* and *meronymy*. The basic semantic relation organising the nouns in the Princeton WordNet is hyponymy. Hyponymy is the lexical relation corresponding to the inclusion of one class in another. Nouns are organised in semantic hierarchies in such a way that a lexical inheritance system is created. For example: *canary* @→ *finch* @→ *passerine* @→ *bird* @→ *vertebrate* @→ *animal*. (The '@' indicates that the relation is hyponymic.) 'Animal' is at the top of one such hierarchy. Altogether there are twenty-five noun hierarchies stored in separate files.

Meronymy refers to relations of the part—whole type. The system offers the possibility of distinguishing three different types of meronymic features for nouns: *component—object* (e.g. *trunk—tree*), *member—collection* (e.g. *tree—forest*), and *stuff—object* (e.g. *aluminium—aeroplane*). Other features such as modification and predication are discussed in Miller et al. (1990) but are not implemented.

3.2.1 Semantic classes in the Princeton WordNet

An underlying principle in the Princeton WordNet is that all nouns are contained in one single hierarchy. For simplicity, a set of semantic primitives treated as unique beginners of each hierarchy is used. In the original WordNet a set of 25 unique beginners was used, but in WordNet 1.6 the set has been culled down to the following 9 top concepts:

1. Entity, something (anything having existence (living or nonliving)).
2. Psychological feature (a feature of the mental life of a living organism).
3. Abstraction (a general concept formed by extracting common features from specific examples).
4. State (the way something is with respect to its main attributes; "the current state of knowledge"; "his state of health"; "in a weak financial state").
5. Event (something that happens at a given place and time).
6. Act, human action, human activity (something that people do or cause to happen).
7. Group, grouping (any number of entities (members) considered as a unit).
8. Possession (anything owned or possessed).
9. Phenomenon (any state or process known through the senses rather than by intuition or reasoning).

The category POSSESSION is complex because all entities and many other types of words can be *owned* in one way or another, e.g. land, personal property, money, securities. In the following studies the term MONETARY REPRESENTATION

will be used for things such as money, securities, interest rate, and debts. Land and personal property will be coded as OBJECT.

One goal of this book is to introduce new relations in WordNet, more precisely to link the adjectives to the nouns they can modify. Therefore, the study must use the semantic categories used in the WordNet hierarchy. However, when visualising the distributions of the semantic categories in the empirical material, the semantic categories have been grouped according to the table below.

Semantic category			***Example***
Main tag	***Subgroups***		
LIVING ORGANISM	*ANIMAL*		*horse, fish, finch*
	HUMAN		*man, girl, Per, baker*
	CREATURE		*troll, elf, God*
OBJECT AND SUBSTANCE	ARTEFACT		*spoon, rope*
	NATURAL OBJECT		*stone*
	MATERIAL		*cloth, metal, leather, stone*
	PLANT		*tree, rose*
ABSTRACTION	ATTRIBUTE		*energy, flexibility, weight, volume*
	QUANTITY		*million, billion*
	RELATION		*kinship, grandmother*
	TIME		*year, month, second*
	SHAPE		*square, curve*
	COMMUNICATION		*smile, word, text*
PSYCHOLOGICAL FEATURE	COGNITION		*knowledge, sight, word, experience*
	FEELING		*longing, relief, rage, worry, happiness*
PHENOMENON	NATURAL PHENOMENON		*wind, stream, rain*
SITUATION	*EVENT*	*HUMAN ACTION*	*crime, simplification*
		OTHER	*fire, break-through, success*
	PROCESS		*growth, business trend*
	STATE		*lack, need, illness, silence*
GROUP	*GROUP*		*audience, opposition*
	ORGANISATION		*UN, college, football club*
MONETARY REPRESENTATION			*money, securities, debts*

Table 1: Semantic categories used to classify nouns in the Princeton WordNet. The italicised tags are used in the empirical studies concerning semantic range, see Chapter 8.

Words such as *book* and *newspaper* are inherently ambiguous. It is impossible to decide whether they are to be interpreted in the sense of an OBJECT or an ABSTRACTION without knowing the context. *En tung bok* 'a heavy book' can refer either to the weight of the book or to the fact that its content is hard to digest. In the EuroWordNet, this problem is avoided by introducing a special

semantic category for written material, TEXT. The TEXT category was not used in the present study, but it would be advisable to do so in future studies.

Some changes to the set of semantic categories will be made in Part III (Empirical Studies).

3.2.1.1 Differences from the Princeton WordNet

The suggested taxonomy differs from that of the Princeton WordNet on a couple of points, which will be accounted for below.

Fantasy creatures such as *trolls* and *leprechauns*[5] are, in one way, artefacts since they are "man-made", but they can also be viewed as abstract concepts. This two-foldedness is reflected in the fact that they are inconsistently categorised in the Princeton WordNet: *leprechaun* is found under the unique beginner ENTITY, while *troll* is classified as a PSYCHOLOGICAL FEATURE.

Trolls and leprechauns are more closely related and should be found in the same hierarchy. The EuroWordNet has introduced a unique beginner CREATURE to cover these words, and that category is used in this study as well. CREATURE is further incorporated in LIVING ORGANISMs in the model used here.

PLANTS are living organisms and are classified accordingly in the Princeton WordNet. However, they behave like OBJECTS in all occurrences found in the material studied. For that reason, PLANTS will here be grouped together with the OBJECTS.

SHAPES, e.g. *en kraftig kurva* 'a sharp curve', are viewed as a type of ABSTRACTION.

[5] A leprechaun is (in Irish stories) a fairy in the shape of a little old man (Oxford Advanced Learner's dictionary)

4 A lexical theory of adjectives

There have been many attempts to define and categorise property concepts and adjectives in the literature. This chapter will start by defining the grammatical category of adjectives, and then some different ways to categorise adjectives within the class will be reviewed. As it turns out, these categorisation models are not incompatible with each other; a model integrating the theories reviewed will be proposed.

4.1 The grammatical category of adjectives

"The class of adjectives is a notorious swing-category in languages" (Givón 1979:13). The grammatical category of adjectives is not easy to define. It is hard to draw the line between participles that take adjectival function and adjectives, and there are many features that some adjectives have but others do not. Quirk et al. (1985:231) list features that are generally considered to be characteristic of English adjectives:

Adjectives can

- freely occur in attributive position;
- freely occur in predicative position;
- be pre-modified by the intensifier *very*; and
- take comparative and superlative forms, inflectionally or by using a pre-modifier such as *more* or *most*.

However, there are counterexamples to all the above criteria. Some adjectives can take only attributive position, e.g. *past*, and some only predicative position, e.g. *asleep*. Many adjectives cannot take comparative form, e.g. **asleepier* and **more asleep*, nor be pre-modified by the intensifier *very*, e.g. **very asleep*. Swedish adjectives display parallel problems.

The fuzziness of the grammatical category of adjectives, what Givón refers to as their being a "swing category", is reflected in the fact that it is a word class that not all languages have. There are languages that have no adjective class at all, e.g. Yurok,[6] and languages with a small and closed minor class of adjectives,

[6] Yurok is an Algonquian language spoken in California.

e.g. Igbo[7] where the adjective class is constituted by only eight adjectives (Dixon 1977). However, it is possible to express adjectival meaning in these adjective-deficit languages, it is just that they use other means than for example Swedish, which has a major adjective word class. There are two strategies to express adjectival concepts in the adjective-deficit languages: through verbs and through nouns.

4.1.1 Temporal stability

Givón (1984:51) claims that the categorisation of a word according to word class is a direct reflection of the time-stability of the word. Nouns are the most time-stable words since they are bound in space and time, cf. Figure 3. Verbs, which are bound neither in space nor in time, are the least time-stable. Intermediate between those two classes we find the adjectives.

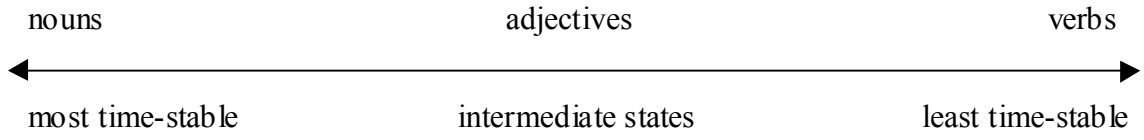


Figure 3: The time-stability scale according to Givón (1984).

In the class of adjectives, both temporally stable and temporally unstable concepts are found: e.g. *long* is (normally) a temporally stable property, while *noisy* is more sensitive to temporality. Givón asserts that adjectives and their attendant properties are cognitively, phenomenologically, and categorically a mixture of nouns and verbs. He claims that nouns and verbs, but not adjectives, are semantic primitives. Given the fact that not all languages have a class of adjectives, he is probably right.

4.1.2 Modifiers as discourse anchors

An alternative view of the nature of property concepts takes a more functional approach, as proposed by Hopper & Thompson (1984) and further described by Thompson (1988). They suggest that nouns and verbs are to be viewed as "discourse-manipulable participants" and "reported events", respectively. The function of adjectives is to anchor referents in discourse in terms of information flow. Thompson distinguishes between predicative and attributive adjectives. These two types are used to signal different functions in discourse: attributives are used to assign properties to new (though not brand new) discourse participants, while predicatives are used to assign properties to already-established discourse participants.

[7] Igbo, spoken e.g. in Nigeria, belongs to the Kwa subgroup of the Niger-Congo family.

4.1.3 Modifiers as single properties

Wierzbicka (1986) claims that both nouns and adjectives denote properties and that the difference between them is a matter of scope. Nouns denote a kind, e.g. a person, animal, or thing, with all the properties that are inherent in the noun in question. An adjective, on the other hand, denotes a single property, e.g. *old car* (where the property of AGE is singled out). An adjective singles out a property and does not refer to kinds in the way nouns do. Frawley (1992:441) exemplifies the contrast with the following example:

(a) Bill is an Irishman.

(b) Bill is Irish.

According to Frawley, example (a) gives much more information than example (b). It tells us not only that Bill is from Ireland, but also implies that Bill has a certain cultural background, a certain appearance, a certain behaviour, etc. – properties that are inherent in the noun. The adjective in (b) singles out the feature of nationality: it denotes one property independently.

4.2 Adjectival features

Most Swedish adjectives are clearly distinguishable morphologically, syntactically, and semantically. General features of adjectives will be illustrated here through Swedish examples.

4.2.1 Gradability

One of the most typical features of adjectives is gradability. An adjective is *gradable* if it can express comparison, e.g. *lång—längre—längst* 'long—longer—longest', and if it can be intensified, e.g. *mycket längre* 'much longer'. Gradability is fundamentally a semantic feature and it implies the existence of a scale in the structure of the adjective (Rusiecki 1985:3). The above example proves the existence of a scale denoting LENGTH, to which also e.g. *kort* 'short', *småväxt* 'short', *kortvuxen* 'short', *högväxt* 'tall', and *reslig* 'stately' belong.

Gradability is a well-studied field; see for example Bierwisch & Lang (1987). For Swedish material, see Lundbladh (1988).

4.2.2 Syntactic functions

Adjectives in general can take attributive or predicative function, e.g. *en stark pojke* 'a strong boy', *pojken är stark* 'the boy is strong', or function as the head

of a noun phrase that lacks a noun head, e.g. *de unga* ‘the young’, *de blinda* ‘the blind’.

4.2.3 Morphological features

Swedish adjectives are easy to distinguish thanks to the relatively rich morphology of the language. Adjectives agree with the nouns they modify; they can be inflected according to gender, number, and definiteness – see the examples below.

ett grön-t äpple a green-NEU apple	*en grön cykel* a green-UTR bike
två grön-a äpplen two green-PL apples	*två grön-a cyklar* two green-PL bikes
det grön-a äpplet the green-DEF apple	*den grön-a cykeln* the green-DEF bike
de grön-a äpplena the green-PL-DEF apples	*de grön-a cyklarna* the green-PL-DEF bikes
äpplet är grön-t the apple is green-NEU-DEF	*cykeln är grön* the bike is green-UTR-DEF
flera äpplen är grön-a several apples are green-PL	*flera cyklar är grön-a* several bikes are green-PL
äpplena är grön-a the apples are green-PL-DEF	*cyklarna är grön-a* the bikes are green-PL-DEF

Gradable adjectives can normally occur in absolute, comparative, and superlative forms. Comparison takes place either inflectionally, e.g. *stark* ‘strong’—*stark-are* ‘stronger’—*stark-ast* ‘strongest’, or by using the pre-modifiers *mer* ‘more’ and *mest* ‘most’, e.g. *ekonomisk* ‘economical’—*mer ekonomisk* ‘more economical’—*mest ekonomisk* ‘most economical’.

4.3 Adjective typologies

There have been many attempts to categorise adjectives, but apart from the traditional differentiation between descriptive and classifying adjectives, no theory has been pervasive enough to become normative. The classification systems presented are basically of three types:

(1) syntactic classifications;

(2) semantic classifications;

(3) a mixture of the two: semantico-syntactic classifications.

A review of some theories belonging to each category will follow, and then an attempt to bring them together into a single model will be presented.

4.3.1 Syntactic classifications of adjectives

Given the many syntactic features mentioned above, one might expect that it would be easy to build a system of syntactic criteria to classify adjectives in different groups. Syntactic features of adjectives that can be taken into account are (Quirk et al. 1985):

- syntactic function (attributive/predicative position);
- transitivity;
- ability to take comparison;
- ability to be modified by intensifier;
- ability to stand as direct object; and
- position within the adjective phrase.

Using each of the features as a classification criterion on its own is not very satisfying. This produces unbalanced groups where the majority ends up in one group and only a few "exceptions" in the other. Quirk et al. (1985:404) use some of the syntactic features above to build the following matrix:

	Attributive	***Predicative use after 'seem'***	***Intensifier ('very')***	***Comparison***
hungry	+	+	+	+
infinite	+	+	–	–
old	+	–	+	+
afraid	?	+	+	+
utter	+	–	–	–
asleep	–	+	–	–
soon	–	–	+	+
abroad	–	–	–	–

Table 2: Matrix for syntactic classification of adjectives (Quirk et al. 1985).

Quirk et al. distinguish eight different adjective classes using these features, as shown in Table 2. These classes can be divided into three groups: central adjectives, e.g. *hungry*, *infinite*, peripheral adjectives, e.g. *old*, *afraid*, *utter*, *asleep*, and adverbs, e.g. *soon*, *abroad*. The central adjectives display all the

characteristic features of adjectives, while the more peripheral an adjective is, the fewer of the characteristic features it displays.

Many semantic features are reflected through syntax, and taking the syntactic features of an adjective into account gives important information about the word. But considering the many exceptions mentioned above, it is not possible to use only syntactic criteria to build a differentiated classification system for adjectives.

4.3.2 Semantic classifications of adjectives

Classification of adjectives based on semantic criteria has been more fruitful. The traditional categorisation of adjectives as descriptive or classifying is based on reference. This is the only model to have reached some kind of general approval. Quirk et al. (1985) and Noreen (1904) have proposed other semantically based models.[8]

4.3.2.1 Classification based on reference

The traditional differentiation between descriptive and classifying adjectives is based on different types of reference. A descriptive adjective necessarily modifies an inherent property,[9] e.g. *a small apple*, *a sour apple*, *a red apple*. Size, taste, and colour are properties inherent in apples and the adjectives in the above noun phrases are therefore descriptive.

Classifying adjectives, on the other hand, do not refer to a property inherent in their head nouns. What is modified is some non-inherent feature of the noun. For example, in *a French apple* the adjective indicates where the fruit was grown, but the property of NATIONALITY is not inherent in the noun *apple*. Further examples with classifying adjectives are *scientific methods*, *medical device*, and *medieval church*.

4.3.2.1.1 Descriptive adjectives

The prototypical adjectives, the ones that first come to mind when we think of adjectives, are descriptive adjectives, sometimes also called "bona fide adjectives" and "proper", "central", or "predicating adjectives". In the WordNet documentation descriptive adjectives are defined as follows: "A descriptive adjective is one that ascribes an attribute to a noun. That is to say, *x* is *Adj* presupposes that there is an attribute *A* such that $A(x) = Adj$. To say *The package*

[8] For a review of Noreen's semantic classification system for adjectives, see Malmgren (1990:62).

[9] All entities are assumed to be characterised by or possess certain perceptible or otherwise intelligible properties, see Lyons (1977:110).

is heavy presupposes that there is an attribute WEIGHT such that WEIGHT(package) = heavy" (Fellbaum, Gross & Miller 1993).

Descriptive adjectives most faithfully display the "adjectival" characteristics listed by Quirk et al. (1985): in general, they may occur in both attributive and predicative positions, and they are gradable, i.e. they can take comparison and be pre-modified by an intensifier such as *very*. However, there are exceptions to the rule. There are descriptive adjectives which, like classifying adjectives, have an absolute meaning and therefore are not gradable. Some examples are *tresidig* 'three-sided', *rätvinklig* 'right-angled' and *död* 'dead'. Warren (1984) suggests that there is a dichotomy between "either-or attributes/effects", e.g. *tailed*, *bearded*, *additional*, and "more-or-less attributes/effects", e.g. *talented*, *sensational*, where the former are absolute values impossible to grade while the latter are gradable. Thus, gradability cannot be used to distinguish descriptive adjectives from classifying ones.

The syntactic restrictions on descriptive adjectives are more liberal than those on classifying adjectives. The adjective phrase is a popular example among computational linguists because of its recursiveness, i.e. the possibility to pile adjective phrases in front of a head noun. However, the adjectives are not placed in random order, cf. Loman (1956). If an adjective phrase consists solely of descriptive adjectives, these can be co-ordinated in free order by means of conjunctions. Judging from my data, classifying adjectives tend to come after the descriptive adjective(s) in a multiword adjective phrase:

**en finansiell stor flopp* 'a financial grand failure' but
en stor finansiell flopp 'a grand financial failure'

**en politisk ny idé* 'a political new idea' but
en ny politisk idé 'a new political idea'

**en politisk bred enighet* 'a political wide consensus' but
en bred politisk enighet 'a wide political consensus'

4.3.2.1.2 Classifying adjectives

Classifying adjectives are characterised by absolute meaning and are therefore not gradable, e.g. *svensk* 'Swedish', *politisk* 'political', and *grevlig* 'of a count'. Nor can an intensifier such as *mycket* 'very' normally pre-modify them in the intensifying sense. In the examples below *mycket* means 'typically' rather than 'very'.

?*ett svenskare anfall* 'a more Swedish attack'
?*ett mycket svenskt anfall* 'a very Swedish attack', i.e. 'a typically Swedish attack'

?*en politiskare/mer politisk idé* 'a more political idea'
?*en mycket politisk idé* 'a very political idea', i.e. 'a typically political idea'

?*ett kungligare slott* 'a more royal castle '
?*ett mycket kungligt slott* 'a very royal castle', i.e. 'a typically royal castle'

Classifying adjectives are generally derived from noun stems; there are several productive endings used to form classifying adjectives in Swedish, e.g. *-ig*, *-mässig*, *-isk* (Teleman 1972:76).

Syntactically, classifying adjectives tend not to occur in predicative position, unless the intention is to bring out a contrastive meaning. Classifying adjectives often occur as (pre)modifiers of abstract nouns and in these cases often in a use corresponding to a genitive or noun premodifier.

Sometimes it is hard to categorise an adjective as descriptive or classifying. Many adjectives actually show up in both groups, cf. the examples below.

nervösa rubbningar 'nervous disturbances'
Han blev nervösare och nervösare 'He became more and more nervous'

manligt arbete 'masculine work'
Lundström gör ett mer manligt intryck än Rundkvist
'Lundström makes a more masculine impression than Rundkvist'[10]

In the first case, we clearly have a classifying adjective, telling us what type of disturbance we are dealing with. In the second case *nervös* 'nervous' modifies an attribute of the referent, and in this case *nervös* must be classified as a descriptive adjective.[11] Whether the referent is modified must be taken into account for correct classification of these adjectives to be possible.

In the examples with *manlig* 'masculine', the adjective has been derived from *man* 'man' to form a classifying adjective. This adjective has, owing to a semantic shift, evolved to function as a descriptive adjective, too.

A key property of classifying adjectives is that they are reminiscent of noun-noun compounds in that they form syntactic and semantic units with their heads.

[10] Example from Teleman 1972:79.

[11] *Nervös* is a borrowing from French, and it is uncertain whether it was borrowed as a classifying adjective with a subsequent semantic shift in Swedish or whether both types of meaning came with the loan word. The SAOB does not make any distinction.

4.3.2.1.3 Reference-modifying adjectives

"Reference-modifying adjectives" is a term introduced by Bolinger in 1967. He contrasted them with "referent-modifying adjectives", which in WordNet correspond to descriptive adjectives. For example, in the nominal phrase *den förre kungen* 'the former king', *förre* does not modify the referent, but rather its reference. Reference-modifying adjectives can occur only in attributive position and the nouns they modify generally denote a function or a social relation. They form a closed class of only a few dozen adjectives (Fellbaum et al. 1990) and will not be further discussed in this book.

4.3.2.1.4 Identifying adjectives

Any adjective can serve as an identifying adjective; cf. the examples below.

> Hand me the red book, please!
> Hand me the medical book, please!

However, there are some adjectives that take exclusively identifying function, e.g. *enda* 'sole', *särskild* 'particular' (Warren 1984). Identifying adjectives display some special features, but since they are not within the scope of this thesis, the interested reader is directed to Warren (1984:102).

4.3.2.2 Three dimensions for classification

Quirk et al. (1985:434ff) give an alternative classification system to the one mentioned above. Three different scales are applicable to adjectives: *stative—dynamic*, *gradable—non-gradable*, and *inherent—non-inherent*.

Most properties do not change over time, i.e. they are *stative*. But there are properties of a more temporal character, e.g. *abusive*, *awkward*, *foolish*, and those are categorised as *dynamic*.

Gradability as defined above is present in all dynamic and most stative adjectives. Those stative adjectives that are non-gradable correspond to the category of classifying adjectives.

The third scale divides the adjectives according to whether they modify a referent directly or not, i.e. inherently or non-inherently. An inherent adjective applies to the noun directly, e.g. *a firm handshake*. A non-inherent adjective modifies some extension of the basic sense of the noun, e.g. *a firm friend*, which means 'a friend whose friendship is firm' rather than 'a friend who is firm'.

Again, the different features are summarised in a matrix:

Type	*Example*	*Gradable*	*Inherent*	*Stative*
Central	*She is a brave woman.*	+	+	+
Dynamic use of central adjective	*She is being very brave.*	+	+	–
Peripheral adjectives: non-inherent	*He is a firm friend.*	+	–	+
Dynamic use of stative adjective	*This actor is being wooden tonight.*	+	–	–
Peripheral adjective: non-gradable and non-inherent	*She is a medical student.*	–	–	+

Table 3: Classification of adjectives according to gradability, stability, and inherentness (Quirk et al. 1985).

4.3.2.3 A semantic network: WordNet

WordNet takes another approach to the description of adjectives. Four different adjective types are coded in WordNet: *descriptive adjectives* (e.g. *little*, *wet*), *classifying adjectives* (e.g. *fraternal*, *nuclear*), *reference-modifying adjectives* (e.g. *former*, *alleged*), and *colour adjectives* (e.g. *blue*, *purple*).

4.3.2.3.1 Descriptive adjectives

The descriptive adjectives are grouped around antonymous pairs (e.g. *stark—svag* 'strong—weak'), quite differently from nouns and verbs, which are organised in hierarchies with a separate file for each hierarchy. Each antonymous pair designates a particular scale, and each adjective in an antonymous pair has sets of synonyms, so-called "synsets", linked to it.

The following figure shows the opposed concepts *stark—svag* in the middle with synsets (in the circles) linked to them.

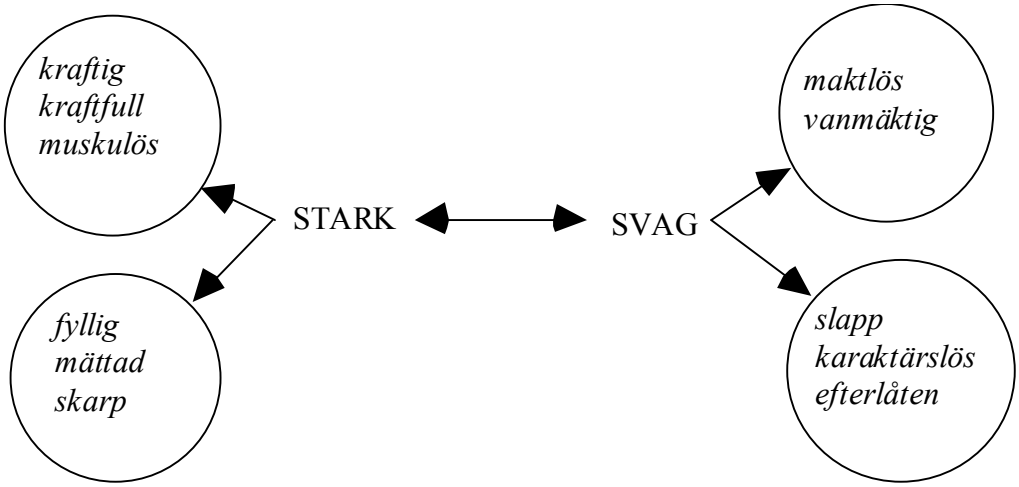


Figure 4: An example of the organisation of descriptive adjectives in a Swedish version of the Princeton WordNet.

Stark and *svag* are *direct antonyms*, while opposed concepts within the cluster around the scale that are not directly connected are called *indirect antonyms*, e.g. *stark—maktlös*, *svag—kraftig*, and *kraftig—slapp*.

4.3.2.3.2 Classifying adjectives

Fellbaum et al. (1993) use the term "relational adjectives" for classifying adjectives. As the term suggests, this type of adjective is associated with something, i.e. the classifying adjectives pertain to nouns in the sense that they are derived from noun stems (sometimes in other languages from which they were subsequently borrowed). Examples from English would be *sisterly* as in *sisterly love*, and *dental* as in *dental hygiene*.[12] Examples of relational adjectives in Swedish are derivations from Greek or Latin nouns such as *termisk* 'thermal' and *manuell* 'manual', but there are also very productive rules using the suffixes *-mässig,*[13] *-ig*, and others (Teleman 1972).

The fact that a relational adjective is derived from a noun is reflected in WordNet through a link from the adjective to the noun stem from which it is derived.

4.3.2.3.3 Reference-modifying adjectives

The reference-modifying adjectives are treated as descriptive adjectives. Some of them have a direct antonym, e.g. *förre—nuvarande* 'former—current', *möjlig—omöjlig* 'possible—impossible', and those that lack a direct antonym usually have an indirect one.

4.3.2.3.4 Colour adjectives

Colour adjectives are treated as a special case in the Princeton WordNet documentation, but there are no colour terms coded in the lexical database. Colour adjectives are descriptive, according to the reference-differentiation criterion, but they are problematic to code in WordNet's system of descriptive adjectives since they do not group around antonymous pairs in the same way as the descriptive adjectives do.

4.3.2.3.5 Participles

There is actually a fifth category not documented in Miller et al. (1990) but implemented in WordNet 1.5, namely participles. Participles are verb derivations with adjectival functions. They are listed as adjectives in WordNet, but their close relation to verbs is maintained by linking them to their respective verb root.

[12] Swedish often uses compounding for this type of construction, e.g. *systerkärlek* 'sister-love' for *sisterly love* and *tandhygien* 'tooth-hygiene' for *dental hygiene*.

[13] For a thorough description of the suffix *-mässig*, see Söderbergh (1964).

4.3.3 Semantico-syntactic classifications of adjectives

The third type of approach brings semantics and syntax together. There have been many attempts in this direction; suggestions by Dixon (1977), Warren (1984), and Malmgren (1990) will be reviewed below.

4.3.3.1 Universal semantic categories

Dixon (1977) studied 20 languages from different language families, focusing on how the different languages express adjectival meaning. He found that there are three types of languages in this respect: languages with a major class of adjectives (e.g. Swedish), languages with a minor class of adjectives (e.g. Igbo), and languages with no adjective class at all (e.g. Yurok). He found that among the languages with a class of adjectives, major or minor, the semantic content of the adjectives is fairly constant. Those universal semantic categories are:

- dimension (e.g. *large*, *little*);
- colour (e.g. *black*, *white*);
- value (e.g. *good*, *bad*, *pure*); and
- age (e.g. *new*, *young*, *old*).

Beyond the four universal semantic categories he distinguishes another three to cover the vocabulary of descriptive adjectives in English:

- physical properties (e.g. *hard*, *cold*);
- human propensity (e.g. *jealous*, *happy*, *clever*); and
- speed (e.g. *fast*, *slow*).

The similarity in semantic content between the languages with major adjective classes and the similarity of content between minor classes across languages suggest the existence of some type of syntactico-semantic universals. The majority of the syntactic properties of a lexical item can be predicted from the semantic description of that item. Dixon suggests three different levels of description:

- *Universal semantic level*: all dictionary items in all languages belong to a certain universal semantic type.
- *Basic/deep level*: the semantic type to which an item belongs will be associated with a specific word class in the language in question.

- *Surface level*: extensional properties such as derivational membership of other word classes.

There is no reason to believe that the set of seven semantic categories covering the English vocabulary of descriptive adjectives does not cover the Swedish descriptive adjectives as well. For applications of Dixon's universal semantic categories to Swedish adjectives, see Viberg 1994 and Stroud 1979:186–188.

4.3.3.2 Semantic relations between adjective and noun

Yet another approach is to consider the referent as well in the classification system. Malmgren (1990) and Warren (1984) build up syntactico-semantic systems bringing the referent into the picture.

Warren (1984) thinks in terms of components and uses role labels in her analysis. She distinguishes the following semantic relations between adjectives and nouns:

Role combination	***Connecting link***	***Example***
SOURCE–RESULT	constituted by	*criminal case*
RESULT–SOURCE	constituting	*criminal assault*
NORM–ADHERENT	in accordance with	*conventional methods*
COMPARANT– COMPARED	resembling	*Roman nose*
WHOLE–PART	belonging to	*vocal tone*
PART–WHOLE	having	*rational creature*
PLACE–OBJ	occurring in/on	*celestial bodies*
OBJ–PLACE	containing	*magnetic field*
ORIGIN–OBJ	deriving from	*domestic sewage*
TIME–OBJ	occurring in/at	*nocturnal illumination*
OBJ–TIME	during which – prevails/prevailed	*nuclear age*
AFFECTED OBJ–ACTOR	dealing with	*medical officer*
CAUSER–RESULT	caused by	*electric shock*
RESULT–CAUSER	causing	*pathetic boy*
GOAL–INSTRUMENT	be for	*athletic equipment*

Table 4: Semantic relations between adjectives and nouns according to Warren (1984).

Classifying adjectives are very complex and all the semantic relations mentioned can apply to classifying adjectives. On the other hand, only a subset of them are applicable to descriptive adjectives.

4.3.3.2.1 Descriptive adjectives

According to Warren (1984), descriptive adjectives basically indicate attributes, or "our assessment of the effect or alternative identity of that which is indicated by the noun head" (Warren 1984:88). Inspired by Aarts & Calbert (1979), Warren suggests a number of patterns that descriptive adjectives adhere to:

Relation type	***Example***
RESULT–SOURCE	*problematic law*
RESULT–CAUSER	*healthy air*
NORM–ADHERENT	*normal behaviour*
PART–WHOLE:	
FEATURE–WHOLE	*big house*
EXPERIENCE–EXPERIENCER	*angry man*
MANIFESTED EXPERIENCE/FEATURE–"MEDIUM"	*sad eyes*
APTITUDE/INCLINATION–POSSESSOR	*musical boy*
OBJECT–PLACE	*stony garden*
OBJECT–TIME	*dangerous time*
COMPARANT–COMPARED	*childish man*

Table 5: Semantic relations between descriptive adjectives and nouns according to Warren (1984).

Warren's system is elaborate and covers the English adjective-noun relations relatively well. Interestingly enough, her system is also applicable to noun-noun compounds.[14] However, its finely grained classification is applicable primarily at a low level of the lexicon.

4.3.3.3 Classification of the semantic relations

Malmgren (1990) categorises the semantic relations between adjective and noun into eight different groups. He obtains the following categories:

A. Attributives

(1) (a) The attribute modifies a property of the head noun directly, e.g. *en röd bil* 'a red car'.

(b) The attribute modifies a head noun other than the one explicitly mentioned, e.g. *ett vänligt brev* 'a friendly letter', *ett tveksamt förslag* 'a dubious suggestion'.

(2) The adjective is classifying, e.g. *den ekonomiska politiken* 'economic policy'.

(3) The head noun has the case role of object, e.g. *ett lätt problem* 'a simple problem'.

(4) The head noun has the case role of agent, e.g. *han är villig att åka* 'he is willing to go'.[15]

[14] Warren's system of semantic relations covering the adjective-noun combinations actually sprang from an extensive study of the semantic patterns of noun-noun compounds (Warren 1978).

[15] The subject of *villig* 'willing' is usually classified not as agent but as experiencer; Andersson (1984) notes that the use of agentivity in Malmgren's classification system probably is intended to sort out constructions with for example the adjectives *rädd* 'afraid' and *glad* 'happy' and in *Han var rädd för att förlora jobbet* 'He was anxious of losing his job'.

(5) The modification is made according to a relative norm depending on the head noun (unique to each noun), e.g. *en liten elefant* 'a small elephant'.

B. Adverbials

(6) The head noun is modified as to its referent's action by an evaluating adjective, e.g. *han var dum som for* 'he was foolish to leave'.

(7) The adjective modifies the referent as a function of something, e.g. *en flitig student* 'a diligent student'.

(8) The adjective modifies the head noun's descriptive element rather than the denotative one, e.g. *en sann patriot* 'a true patriot'.

Like Warren, Malmgren has used the work of Aarts & Calbert (1979) as a platform for his study. However, since he gathers the different semantic role relations between adjectives and nouns into groups, his model can be viewed as existing at a level above Warren's semantic relations.

4.4 Bringing the theories together

The reviewed systems for the classification of adjectives are quite diverse. Looked at more closely, they seem to treat different levels of the lexicon. Bringing them together gives a more unified picture of property concepts and their lexical characteristics; see Figure 5.

The division according to reference is a good starting point, since adjectives behave quite differently depending on their type of reference. Within each reference type, the adjectives are organised accordingly: descriptive and reference-modifying adjectives are organised around antonymous pairs and synsets, while classifying adjectives are connected to the roots they pertain to. The lexical structure common to all three types of adjectives constitutes the actual meaning of the lexemes. Little is known about how this is encoded, and it will not be further discussed in this thesis. Instead, the focus will be on lexical relations.

WordNet does not categorise adjectives any further than according to reference. The lexical features of an adjective are described through its place in the net. This approach gives a rough idea of the meaning of the adjectives, but it is impossible to understand how they should actually be used. I suggest that a

layer of top concepts be added, to which each individual scale is linked.[16] This would add important semantic information. Furthermore, each lexical entry in the adjective lexicon can be linked to the different nouns or groups of nouns that it can modify. This connection to the noun lexicon would reflect the semantic range of the adjective. The links can also carry information about what type of relation obtains between the adjective and the noun, e.g. SOURCE–RESULT, following Warren (1984).

The bidirectional arrow from the classifying adjectives to the noun lexicon in Figure 5 reflects the fact that the adjectives in this group are linked to the nouns they can modify as well as to the nouns they pertain to.

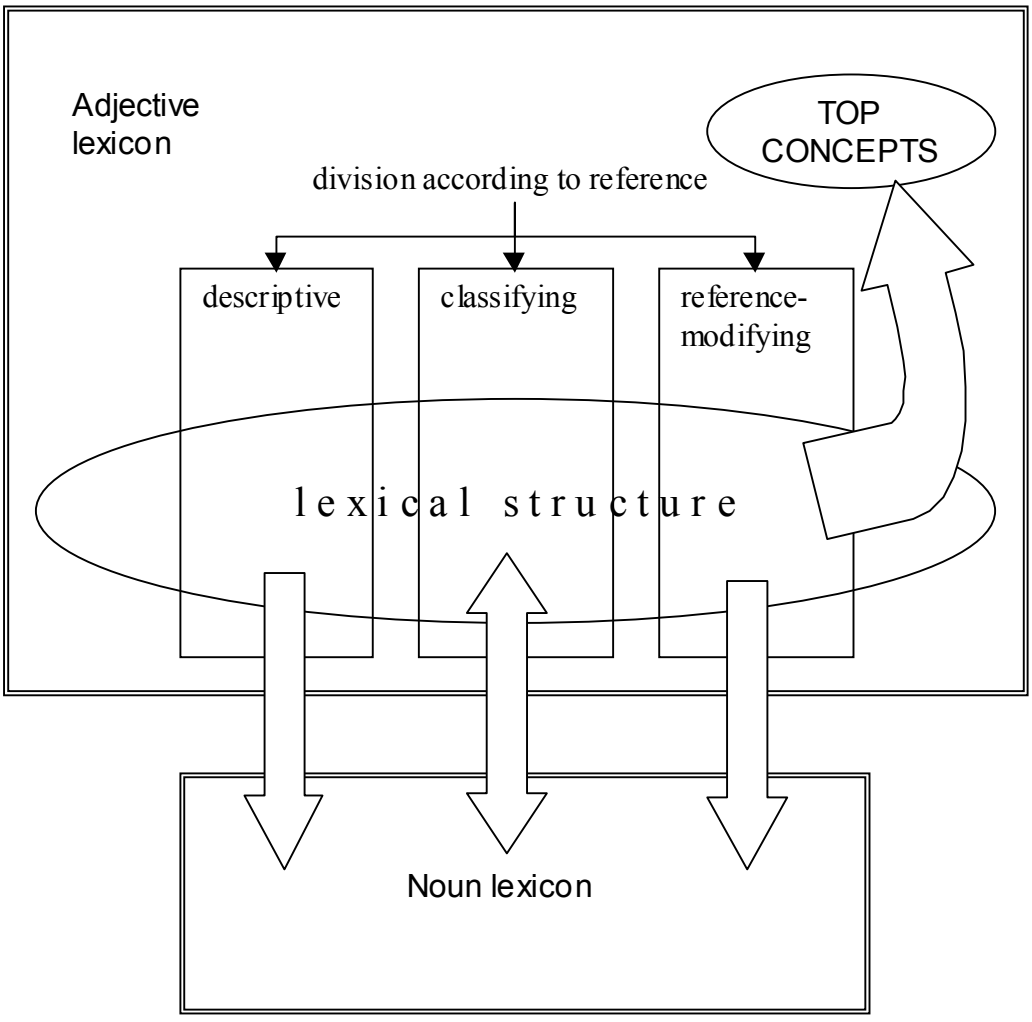


Figure 5: A lexical model for adjectives.

The top concepts can serve as a model for categorisation according to semantic type. Dixon's seven basic semantic types for English adjectives can be used as top concepts; they cover most of the descriptive adjectives. Each scale (cluster

[16] This categorisation is possible only for descriptive and most referent-modifying adjectives.

of words) points to a top concept, in analogy with the noun hierarchies. The scale of LENGTH, *short—long*, would thus be linked to DIMENSION, and the scale of "HAPPINESS", *happy—sad* is a type of HUMAN PROPENSITY and would be linked accordingly (see examples in Figure 6).

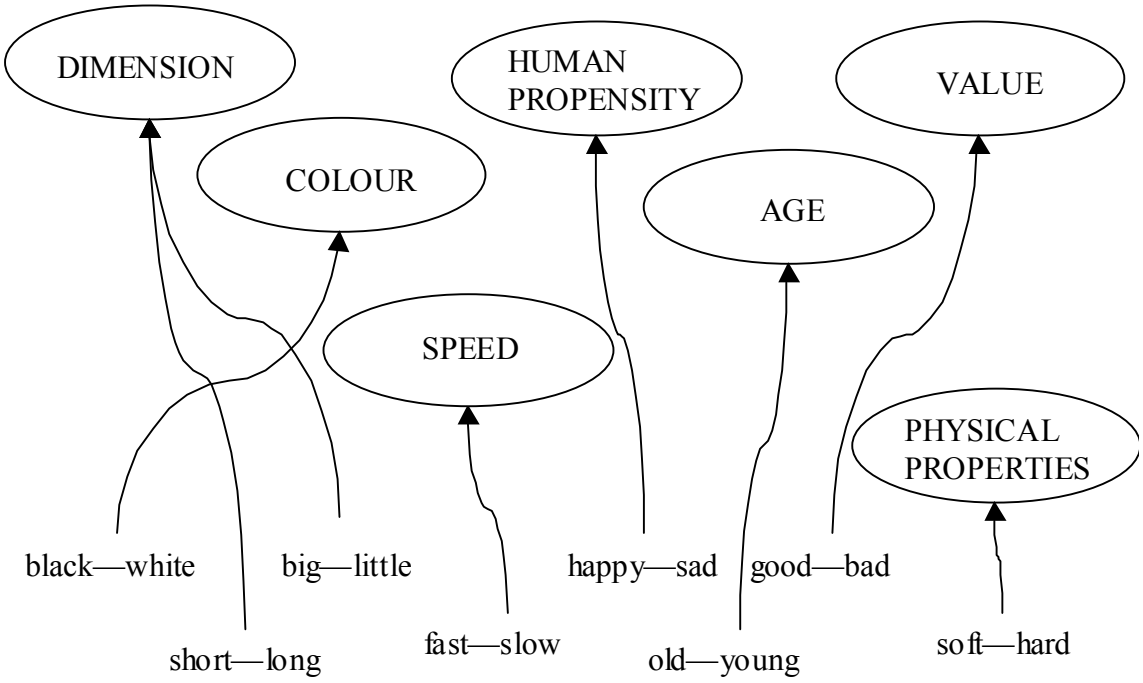


Figure 6: Model of linking scales to top concepts in WordNet.

The bottom layer would be part of an interface to the noun taxonomies in WordNet. Connections from an adjective to the nouns it can modify reflect its semantic range as a modifier. Such links can carry information about the type of semantico-syntactic relation existing between members of the adjective-noun pair, e.g. RESULT–SOURCE or PART–WHOLE.

The rest of this thesis will concentrate on two different levels of the lexical description: lexical relations and semantic range.

4.5 Summary of Part I

Some semantic definitions were presented in Chapter 2: Words express concepts, and concepts are interrelated. A group of interrelated words may form a lexical field, e.g. colour words. Gradable words belong to a semantic scale, which in turn belongs to a lexical field. For adjectives, gradability can be realised through comparison, e.g. *strong—stronger*, but it should be noted that a scale to which *strong* belongs contains also its synonyms and antonyms, such as *robust* and *weak*.

Language is governed by grammar, but the use of a word is also restricted semantically. The words that an adjective can modify reflect its *semantic range*.

The basic lexical relations organising adjectives in our mental lexicon are *synonymy* and *antonymy*. The weakened, context-sensitive definition of synonymy will be used in this book: "two expressions are synonymous in a context *C* if the substitution of one for the other in *C* does not change the truth-value of *C*."

Direct antonyms (*polar antonyms*) are two words from the same semantic field that are gradable, incompatible, and at least semi-reciprocal.

Chapter 3 dealt with the organisation of nouns in WordNet. The basic semantic relation organising the nouns in WordNet is hyponymy. Nouns are organised in semantic hierarchies in such a way that a lexical inheritance system is created. There are 25 such semantic hierarchies stored in separate files. The semantic categories are further grouped under 9 different base concepts, which will be used as semantic tags in the studies presented in this book. In addition, some semantic categories are coded with a finer grain: e.g. animals, humans, and creatures are coded as such, though they are all ORGANISMS.

The last chapter in Part I dealt with different theories of modification. Property concepts are complicated; they are not realised as adjectives in all languages, though this is the case in Indo-European languages. Adjectives are generally gradable, can occur in attributive or predicative position, can be premodified by an intensifier such as *very*, and can take comparative and superlative forms.

Many attempts have been made to categorise adjectives, basically using three different approaches: syntactic classifications, semantic classifications, and semantico-syntactic classifications. A number of adjective typologies were reviewed and it was concluded that one classification system does not exclude the other; rather, they describe different levels of the lexicon.

The studies in this book primarily adhere to the Princeton WordNet framework, but this is a flexible system and it is fairly easy to add new types of knowledge at various levels of the lexicon. A merger of the described adjective classification systems is proposed using the Princeton WordNet as its basis.

In the model I propose, adjectives are primarily divided into three different groups according to their type of reference: descriptive, classifying, and reference-modifying adjectives. Descriptive and reference-modifying adjectives are organised around antonymous pairs, while classifying adjectives are linked

to the noun lemmas they pertain to. Additional information about the meaning of an adjective is available through links to Dixon's basic semantic types, i.e. *dimension*, *colour*, *speed*, *human propensity*, *age*, *value*, and *physical property*.

The semantic range of an adjective is reflected by a connection from the adjective to the nouns it can modify. Each link carries information about what type of relation obtains between the adjective and the noun, according to Warren (1984).

I believe that there is much more information to be added to the lexicon, and the model I propose allows for expansion. However, this is as far as I will take it in this book.

Part II: Methods of Corpus Research

5 Corpus-based methods

5.1 Why corpus-based methods?

There are a number of ways to collect data to study semantics, such as elicitation, experiments, casual observation, intuition, and corpora.

All different ways of collecting data are important, but all of them have their limitations. Elicitation and experiment require complicated methods and are arts of their own, but done right they can yield interesting results. Casual observation is of limited use because of the sampling problem, but it is an important complement to intuition, which is in many cases the starting-point of an investigation. Large corpora provide the opportunity to study semantic and syntactic phenomena on a larger scale.

5.2 Corpora

5.2.1 What makes text a corpus?

A corpus is basically a collection of texts in an electronic database. This is a rather wide definition, and there are other more demanding ones, e.g. that corpora must consist of structured collections of text specifically compiled for linguistic analysis, that they must be large, and that they must be representative of a language as a whole (Kennedy 1998:3, Biber et al. 1998:246). Biber et al. assert that a collection of text must be compiled with a specific purpose to be called a corpus. Furthermore, they require that the collection of texts be representative of a language or a part of a language to qualify as a corpus.

The text that makes up the corpus may be plain or pre-processed. It is common for the text to be tagged with morphosyntactic information, but there are also corpora annotated with e.g. prosodic and semantic information.

5.2.2 English corpora

A great many researchers from various countries have been working on English corpora since the 1960s, and this has lead to the existence of a very large number of corpora available for English. English corpora have been compiled for several different purposes. There are corpora for lexicography, for the study

of spoken language, for research on language acquisition, and for diachronic studies, as well as a large number of corpora compiled for special purposes.

Some of the studies in this thesis have made use of English corpora, namely the Brown corpus and the British National Corpus (BNC).

5.2.2.1 The Brown corpus

The Brown corpus was the first electronic corpus compiled and has set a standard for corpus-based research. It is a small corpus of 1 million words, genre-balanced across 15 categories and consisting of 500 text extracts of about 2,000 words each. The words are tagged with word-class information. Of the many versions of the Brown corpus available, the one distributed by ICAME[17] has been used for the studies in this book.

5.2.2.2 The British National Corpus

The British National Corpus (BNC) is a corpus of 100 million words, of which 90 million are written language and the remaining 10 million spoken language. Most of the texts are of informative character, i.e. they are chosen from the fields of applied sciences, arts, belief and thought, commerce and finance, leisure, natural and pure science, social science, and world affairs. About 25 per cent of the texts are of imaginative character, i.e. they are taken from literary and creative works. The corpus has been automatically word-class tagged. Further information about the BNC can be found at http://info.ox.ac.uk/bnc/ (as of March 2001).

5.2.3 Swedish corpora

Though Sweden has had many distinguished researchers working in the field of corpus linguistics, the flora of available machine-readable tagged Swedish corpora is not large. Most of the Swedish corpus linguists' efforts have concerned the English language, e.g. the London-Lund Corpus. There is only one manually tagged corpus of Swedish text, the Stockholm-Umeå Corpus (SUC), which is of fairly high quality. All other sizeable tagged Swedish corpora, such as Parole, are tagged using programs trained on the SUC, and of course nothing can improve on its model.

The SUC and Parole are the Swedish corpora used in the empirical studies underpinning this thesis.

[17] Resources from ICAME (the International Computer Archive of Modern English) can be requested from the Norwegian Computing Centre for the Humanities (http://www.hd.uib.no/, as of March 2001).

5.2.3.1 The SUC

The Stockholm-Umeå Corpus (SUC) is a genre-balanced corpus consisting of 1 million Swedish word tokens, manually tagged for word class and lemma. The balancing of genres and the physical partitioning of the corpus follow the principles used for the Brown corpus. The genres are listed in Table 6. The corpus and the tagset are further described in Ejerhed (1996) and Ejerhed et al. (1992).

Press: reports
Press: editorial
Press: reviews
Skills and hobby
Popular science
Biographies and novels
Miscellaneous
Scientific reports
Fiction

Table 6: Genres included in the SUC.

5.2.3.2 Parole

The Parole corpus is the largest tagged corpus of Swedish text freely available.[18] It consists of 25 million morphosyntactically tagged words. The texts are taken from over 36,000 different documents representing four different types of media: books, newspapers, journals, and "other", see Table 7.

Books
Newspapers
Journals
Other

Table 7: Genres included in the Parole corpus.

5.3 What can be squeezed out of a corpus?

There are many applications using corpora and many more will be developed. The corpus-based methods used in this thesis aim to gather knowledge about the lexicon, mainly through the study of different types of co-occurrence patterns.

[18] Concordances for individual words and expressions can be obtained at http://spraakdata.gu.se/lb/parole/ (as of April 2001).

5.3.1 Quantitative data

Plain frequency counts can give a hint about how common a word is. Examining the frequency of features across texts and registers makes it possible to compare the usage of different words in different domains.

Some of the studies in this book are based on the frequency of semantic categories rather than individual words.

5.3.2 Collocation vs co-occurrence

The two terms "collocation" and "co-occurrence" are closely related and the difference between them is not always clear. *Collocation* has been defined as "the occurrence of two or more words within a short space of each other in a text" (Sinclair 1991:170). "Collocation" often implies a habitual relationship of co-occurrence between the words (e.g. Stubbs 1995). Kjellmer (1984) adopts a stricter definition: "collocations are lexically determined *and* grammatically restricted sequences of words". The term "lexically determined" means that only recurring sequences of words are accepted as potential collocations. The second condition implies that only grammatically well-formed sequences are accepted as collocations.

My definition is more rigorous than Sinclair's (1991), and does not allow any intervening words between the two collocated words, but less strict than Stubbs's (1995) and Kjellmer's (1984) definitions above: "Collocation is the occurrence of two or more words adjacent to each other in a text". A word pair that co-occurs only once is also a collocation according to this definition – I do not require that the word pair be habitually co-occurring or, in Kjellmer's terms, be lexically determined.

Co-occurrence is the occurrence of two or more words within a specified unit of text. This unit can be a span of words of a certain type, e.g. a phrase, a sentence, a paragraph, a chapter, and so forth. When nothing else is specified, I use the sentence as the unit of text.

5.3.3 Strength of the association between words

Several statistical measures have been introduced to gauge the strength of the association between two words, e.g. t-score (Barnbrook 1996:97–98), mutual information, "MI" (Charniak 1993, McEnery & Wilson 1996, Oakes 1998), z-score (Berry-Rogghe 1973). Those measures are designed to determine the strength between words occurring at a close distance from each other, i.e. immediately next to each other or within a fixed window span. Research using

the sentence as its linguistic unit of study has also been presented. For instance, antonymous concepts have been shown to co-occur in the same sentence more often than chance would predict by Justeson & Katz (1991 and 1992) and Fellbaum (1995).

5.3.4 What is a sentence?

Several of the studies that will be presented focus on the sentence as the linguistic unit of interest. The tokenisation of sentences is problematic: the period (or full stop) is the most common type of punctuation ending sentences, but also the most ambiguous one (Grefenstette 1994). Apart from its use as punctuation, the period is also found in numerical expressions, e.g. *13.5%*, in alphanumerical references, e.g. *5.2.4.7*, in dates, e.g. *2001.01.01*, and in abbreviations, e.g. *e.g.* The exclamation mark and the question mark are almost always used as sentence-final punctuation. The semicolon is used to separate list elements and sometimes functions as a sentence separator. Disambiguation of punctuation is not trivial.

The corpora used in the empirical studies presented in this book have been pre-processed; a tagger has dealt with the problems of disambiguating the punctuation marks and tagging them correctly. Thus, distinguishing sentences correctly has not been a problem in the studies presented in this book.

6 Tool kit

There are many tools and programming languages available that specialise in text manipulation and linguistic applications. I will here account for two programs that I have developed to facilitate the studies in this book: *Para* and *Klassa*.

6.1 *Para* – a pairing program

The pairing program *Para* is designed to find the head of each modifier and pair the modifier with its head. It was written in Perl. The corpus used as input should be tagged with word-class and morphological information, which is used to check that the modifier agrees with the head.

The program reads a sentence and creates a list with one word in each cell. The list is processed step by step and for each adjective found, the noun it modifies is searched for. First the right-hand side of the context is examined, and for each step to the right it is checked whether the word in the current cell qualifies in an adjective phrase, i.e. other adjectives, adverbs, conjunctions, and punctuation are accepted, or whether it is a noun. If it is a noun the agreement is checked. The adjective must agree in gender, number, and definiteness.

If a noun is found that agrees with the adjective and there are no words unacceptable in an adjective phrase, e.g. a verb, between the two words, the adjective and the noun are written as output to a file together with the sentence analysed. The fact that the adjective is in attributive position is also recorded. If the program fails to find a qualifying noun in the right-hand context, the left-hand context is examined as well. If the adjective is a predicative there should be a copula in the sentence, and therefore the program searches for a possible copula verb, e.g. *vara* ‘be’, *bliva* ‘become’, in a special list.

That list of verbs accepted as copulas is based on Bolander (1980). If a copula and a noun agreeing with the adjective are found, a word pair consisting of the adjective and the noun is produced as output marked “PRED” for predicative; see the example below. If no possible head noun is found, the sentence is written to the output file marked “problem” so as to be easily distinguished in the manual analysis.

full	*sysselsättning*	ATT	‘full employment’
full	*korgarna*	PRED	‘full baskets’
full	*blom*	ATT	‘full blossom’
…			

The output from *Para* is imported to a database in FileMaker Pro, where semantic tagging is performed manually or using the script *Klassa*.

6.1.1 What Para *does not do*

The current version of the pairing program finds the correct head for about 80% of the adjectives in the corpus. Among the 20% of cases where it fails are:

- Adjectives modifying a word in another sentence or adjectives occurring in headlines, e.g.

  Aldrig fri. ‘Never free.’

- Cases of incorrect tagging of the corpus, causing the checking of agreement to fail, e.g.

  …gamla/SG-DEF pass/PL-INDEF

  ‘…old/SG-DEF passports/PL-INDEF’

- Objective predicatives, e.g.

  Man klassade honom som för gammal.

  ‘He was classified as too old.’

- Resultative adjectives, e.g.

  Ricardas är dock en av dem som lyckats göra sig fri.

  ‘Ricardas, however, is one of those who have succeeded to make themselves free’

- Nominalisations of adjectives, e.g. *de döda* ‘the dead’, *de anhöriga* ‘the relatives’.

- Gradable measurement adjectives, e.g. *gammal* ‘old’, *hög* ‘high’ , *låg* ‘low’, which can take pre-modification by a noun phrase, e.g.

  …säger Ricardas, idag 52 år gammal.

  ‘…says Ricardas, today 52 years old.’

- Predicative attributes, e.g.

  Ivriga öppnade de dörren.

  ‘Eager, they opened the door.’

A failure rate of 20% may sound quite high and it is of course possible to improve the program significantly, at least for the cases where the failure is not

due to incorrect tagging of the input sentences. However, the effort it would cost to reach an error level of 10% or less is not worthwhile, since in the studies where *Para* was used, all the data were also manually studied and classified in a database.

6.2 *Klassa* – a classification program

Klassa is a Perl script that aids all types of classification. It is generic in the sense that it is possible to use any type of input and any set of labels for the classification.

For the studies in this book, the output from *Para* was used as input to *Klassa*, i.e. a list with three columns: an adjective, a noun and a syntactic function. The sentence in which the words appeared was also included. The program *Klassa* uses a database, which in the studies presented in this book consists of words and their semantic classification. For each word in the input list the program checks if it is listed in the database, and if so suggests a semantic category accordingly. The human interactor is asked to confirm the tag or suggest another one. If the word is not found in the database, the last semantic category previously confirmed is suggested.

For each word in the input file *Klassa* asks the human interactor either to confirm a suggested tag or to return a new one. The word and the tag are then saved in an output file and stored in a general "semantic lexicon". The semantic lexicon is used for the suggestion of semantic classifications. The flow chart in Figure 7 gives an idea of how the scrip *Klassa* works.

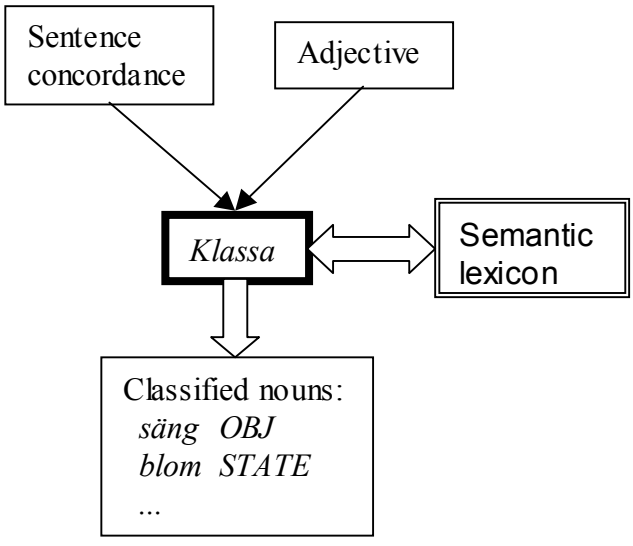


Figure 7: Flow chart of the classification script Klassa.

Klassa is an aid for "manual tagging". In this book, the tags are semantic, but the tag set is arbitrary. It is also possible to use it for other types of input. The strength of *Klassa* is that it builds up a database, which it consults for every new entry. The database grows as new material is tagged, and the ability of *Klassa* to make correct suggestions grows accordingly.

6.3 Summary of Part II

The use of corpus-based methods has grown concurrently with the use of computers in linguistic research. The main advantage of corpus-based methods is that research can be done on very large amounts of empirical data.

The studies in this thesis are based on English and Swedish corpora: the Brown corpus, the BNC, the SUC and the Parole corpus. The rest of the book will focus on the co-occurrence of words, i.e. the occurrence of two or more words within a specified unit of text, typically a sentence.

Two programs that are used to facilitate corpus studies are described in this chapter: *Para* and *Klassa*.

Para finds the adjectives in a sentence and attempts to locate the nouns they modify. It starts by assuming that the adjective is in attributive position and looks in the right-hand context for a noun that agrees in gender, number, and definiteness. If this fails, *Para* assumes that the adjective functions as a predicative and starts looking for a copula earlier in the sentence. If a copula is found, the program looks for a noun, proper name, or pronoun that can function as the head of the adjective. Where no possible head is found in a sentence, the output is marked with the tag "problem" so that these cases can be examined manually.

Klassa is an aid for manual classification. The program bases its suggestions of tags on a database, which grows as more material is analysed. Each suggestion has to be confirmed by a human tagger; if the tag suggested is incorrect it is possible to enter the correct one. The output is a list of classified nouns as well as an enlarged database.

Part III: Empirical Studies

It is well known that some words habitually collocate with other words (Firth 1957). Knowing this, one may wonder: is it possible to distinguish semantic relations through the study of habitual co-occurrence? In this section of the thesis, several empirical studies concerning these issues will be presented.

First, empirical studies of adjectives co-occurring with other adjectives will be presented and it will be shown that direct antonyms co-occur significantly more often than chance would predict. The types of context in which the antonyms co-occur will also be studied and results supporting the substitutability hypothesis will be presented. Furthermore, empirical data will be presented showing that direct antonyms co-occur significantly more often than other adjective pairs in opposition on the same scale. In the second part of the empirical studies, co-occurrence patterns of adjectives with nouns will be presented and the semantic ranges of two antonymous words will be described and compared. Semantic range will also be used to group a number of synonyms according to meaning. It will be suggested that semantic range be coded in SWordNet.

7 Adjectival co-occurrence

This chapter will discuss the semantic relation of antonymy, how antonyms behave in co-occurrence patterns and the possibility to distinguish them using statistical analysis of corpora.

Despite the problems in defining antonymy, language users agree on which words are direct antonyms. How are they learned? Is there an easy way to distinguish antonym pairs?

Antonymy is not a semantic relation specific to adjectives, but is found in all word classes and also between classes (cf. Fellbaum 1995). This study is, however, limited to descriptive adjectives.

7.1 Antonyms in discourse

7.1.1 Adjectival compounds in Chinese

In Chinese it is fairly easy to distinguish the direct antonyms, since there is a rule for compounding that applies only to antonym pairs.

Compounding is a universal method for creating new words. Mandarin Chinese has a special rule by which it is possible to create a noun by combining two antonymous adjectives.[19] The meaning of the resulting compound is a quality whose bipolar extremes are signalled by the two words making up the compound (Li & Thompson 1981:80).

ha[illegible]o-huài	‘good-bad = quality’
dà-xia[illegible]o	‘big-small = size’
cháng-dua[illegible]n	‘long-short = length’
le[illegible]ng-rè	‘cold-hot = temperature’
ga[illegible]o-a[illegible]i	‘tall-short = height’
kuài-màn	‘fast-slow = speed’
hòu-báo	‘thick-thin = thickness’
zhe[illegible]n-jia[illegible]	‘true-false = truthfulness’

The compounds are formed using the following formula, i.e. with the constituent perceived as indicating the positive pole preceding the negative one.

$$A^{+} + A^{-} = N_{\text{compound designating quality}}$$

[19] There is a continuing discussion whether there is a word class of adjectives in Chinese or not; Li & Thompson (1981) treat adjectives as a separate class.

No Indo-European languages have rules of this type. But there may be other features of antonym pairs that can be used to distinguish them.

7.1.2 Antonymy among the Walbiri

The Walbiri are an aboriginal people of Central Australia, extensively studied by anthropologists and linguists. Young Walbiri men are taught the secret language Tjiliwiri after their initiation to adulthood. This ritual language is an "upside-down mirror" of Walbiri: all nouns, verbs, and pronouns of ordinary Walbiri are substituted with their antonyms. This means that 'you are tall' is expressed by 'I am short' in Tjiliwiri, and 'I am sitting on the ground' by 'you are standing in the sky'. (Hale 1971)

This is a nice example showing that some people are explicitly aware of the semantic relation of antonymy.

7.1.3 Speech errors

Slips of the tongue often substitute a target word with a closely related word. Söderpalm (1979) reports several examples of substitution of semantically similar concepts in her dissertation on speech errors. She collected a corpus of normal as well as pathological speech errors. Below are some examples of spontaneous speech errors made by informants suffering from aphasia.

Utterance	***Target***
månad 'month'	*halvtimme* 'half an hour'
ett elnät 'an electricity network'	*spis* 'stove'
frun 'the wife'	*mannen* 'the husband'
ja körde bussen 'I was driving the bus'	*åkte buss* 'was going by bus'
när jag läser 'when I'm reading'	*jag skriver* 'I'm writing'

Table 8: Spontaneous speech errors made by informants suffering from aphasia (Söderpalm 1979:87–89).

Each substitution above was made within the same semantic field, and in the last three cases the target is actually substituted with its antonym.

Most databases of speech errors are based on diary notes (cf. Linell 1982, Fromkin 1973, 1980). This is an unreliable method and makes it impossible to estimate the frequency of different types of speech errors. Söderpalm's corpus of pathological speech errors consists of 110 examples. Of these, 38 are listed as paradigmatic substitutions of words, i.e. where the speaker makes a mistake by choosing the wrong morpheme from the right paradigm.[20] Three of these paradigmatic substitutions can be classified as antonym substitutions (depending

[20] Söderpalm defines *paradigmatic substitutions* of segments as "…one segment is replaced by another one, and the error is not due to any obvious influence from other segments in the utterance but is the result of a mistake in the choice of units from the paradigm, a similarity disorder." (Söderpalm 1979:75)

on how liberal the classification criteria are). However, it is clear from her corpus that semantically related words do appear in substitution errors.

There is also one example of substitution in Linell (1982), where the speaker happens to say *en svara å fråga på* 'an answer to question' when he means *en fråga å svara på* 'a question to answer'. Linell has a larger corpus of slips of the tongue – around 700 items – and a large percentage of them are substitutions ("exchanges" according to Linell). However, only 10 of the substitutions are at the level of content words; most of them involve segments, e.g. *slutat snöa* vs *snutat slöa* 'stopped snowing' vs 'snopped stowing'. He does not dwell further on the types of substitutions found at the word level, so it is unclear whether there are more examples of antonym substitution to be found in his corpus.

7.1.4 Word-association tests

A closely related word can be elicited in psycholinguistic experiments. Deese (1965) used word-association tests to obtain a measure of semantic similarity. A type of stimulus–response test was used, where words functioned as stimuli and the subjects were told to respond with the first word that came to mind. Not surprisingly, words similar in meaning evoked the same responses. Counting the stimulus word itself as a response by each subject, the coefficient of correlation between pairs of words was computed as the intersection of the two distributions of responses and interpreted as a measure of semantic similarity.

Word-association tests have been used in a number of studies of Swedish vocabulary. Among others, Einarsson & Hultman (1984:37ff) compare the language of female and male teenagers using word-association tests, and Abelin (1996) uses this method to study onomatopoetic and other words. However, studies of semantic similarity focusing on the lexical relation of antonymy have, to my knowledge, not been made for Swedish.

Clark (1970) questions what word-association tests actually show: they do indicate that certain words are clearly associated in some way, but not how. He stresses that "any successful explanation of word associations must be formulated in terms of syntactic and semantic features". A stimulus word that has an antonym will elicit that antonym more frequently than any other word. Clark's explanation of this is his "minimal-contrast" rule. Though the results of word-association tests should be interpreted with care, they clearly show that there is a strong relation between the opposites in an antonymous pair.

7.2 The co-occurrence hypothesis

It is clear that people are aware of antonym relations in their language. In some languages, such as Chinese and Walbiri, the awareness of antonyms is explicit, while in others, such as Swedish and English, it is revealed only implicitly through slips of the tongue. In both cases it is possible to show that people consider antonyms to be closely related through word-association tests.

"How are specific words selected as direct antonyms? Or perhaps more answerable: how do people learn that the direct antonymy of wet/dry is somehow different from the same conceptual opposition when it is expressed by, say, soggy and arid?" (Charles & Miller 1989)

Charles & Miller suggest two plausible answers to this question, one based on frequency, the co-occurrence hypothesis, and the other based on context, the substitutability hypothesis:

> The co-occurrence hypothesis for antonymous adjectives: Two adjectives are learned as direct antonyms because they occur together in the same sentences more frequently than chance would allow.

> The substitutability hypothesis for antonymous adjectives: Two adjectives are learned as direct antonyms because they are interchangable in most contexts, i.e. because any noun phrase that can be modified by one member of the pair can also be modified by the other.

They go on to show that substitutability does not seem to be a cue for forming associations between antonymous concepts; rather, their co-occurrence in the same sentence seems to be the cue.

Justeson & Katz (1991) present empirical data that confirm the co-occurrence hypothesis. Using English corpora, they show that antonymous concepts do co-occur in the same sentences more often than chance would predict. They do this by calculating the expected number of co-occurrences in a text on the assumption that all the words in a corpus are randomly distributed and then comparing this with the actual number of co-occurrences. Studying 35 antonymous pairs previously identified by Deese (1965), they find that, overall,

antonym co-occurrence takes place 8.6 times more often than expected. Table 9 shows an extract of their study.

Sentential occurrences of individual adjectives				***Sentential co-occurrences***				
Word1		***Word2***		***Observed***	***Expected***	***Ratio***	***Rate***	***Probability***
1001	new	569	old	28	10.40936	2.7	1/20.3	$3.07*10^{-6}$
122	left	231	right	28	0.51505	54.4	1/4.4	$1.27*10^{-40}$
347	large	504	small	26	3.19623	8.1	1/13.3	$4.33*10^{-16}$
146	black	243	white	22	0.64839	33.9	1/6.6	$2.84*10^{-27}$
407	high	137	low	20	1.01904	19.6	1/6.9	$3.95*10^{-20}$

Table 9: Extract from Deese's adjective pairs and their sentential co-occurrences in the tagged Brown Corpus (Justeson & Katz 1991): the five most frequent adjectival co-occurrences.

The individual words and their number of sentential occurrences are listed in the left-hand part of the above table. It should be noted that what is recorded is the number of sentences, not the total number of occurrences of a word. Thus, the values here may deviate from the total number of occurrences of a word in the corpus; for example, the word *stark* 'strong' occurs 404 times in the SUC but in only 392 different sentences.

The right-hand part of the table lists sentential co-occurrences. The *Observed* column indicates the number of sentences containing both *Word1* and *Word2*.

The next slot, *Expected*, indicates the number of sentences with the two words co-occurring that chance would predict.

Ratio is the ratio between observed and expected co-occurrences, and *Rate* (1/*n*) indicates that one sentence out of *n* sentences with the less frequent adjective will contain its antonym as well.

The last column, *Probability*, shows the probability of finding the number of co-occurrences actually observed or more under the (clearly erroneous) null hypothesis that the co-occurrences are due to pure chance alone.

Justeson & Katz also present empirical data in support of the substitutability hypothesis. They investigate the contexts where the direct antonyms occur and find that "excluding the accidental cases, 63% (139/219) of antonym co-occurrences are in lexically identical structures. In 42% (58/139) of these co-occurrences, the antonyms themselves are simply conjoined."

They conclude that "co-occurrence takes place via substitution, substitution yields antonym alignment, and alignment leads to association."

7.3 On the probability of co-occurrence

Justeson & Katz model the co-occurrence problem using hypergeometric distribution to calculate the number of co-occurrences expected and the probability of finding at least the number of co-occurrences actually observed.

The null hypothesis is that words are randomly distributed throughout the text. The null hypothesis will be proved wrong; the linguist's task is to explain why some words co-occur more often than chance predicts. Justeson & Katz use this model to show that words designating opposite concepts co-occur more often than they would be expected to if their distribution were random.

One problem is that quite obviously words are not randomly distributed. Grammar constrains what words or types of words are used together. Concepts belonging to the same semantic field co-occur and so do words used in idiomatic expressions. Collocation and co-occurrence methods have been used to identify phrases and syntactic features of a word, as well as in "semantic profiling" to distinguish different meanings of a word or to compare two closely related words. The linguist's task is to explain the reason for collocation or co-occurrence from case to case.

This problem, however, is less serious for the Justeson & Katz study since it concerns sentential co-occurrence regardless of in what order the words come in the sentences. This means that the impact of grammar on the order of the words is of less importance, and for that reason the model used is acceptable for that type of study.

7.3.1 Accounting for variation in sentence length

The studies of sentential co-occurrence use the sentence as the unit of study. Justeson & Katz assume that all sentences are of equal length. However, this is quite obviously not so, and variation in sentence length affects both the expected number of sentential co-occurrences and the probability value (or "p-value"). Holtsberg & Willners (2001) argue that the hypergeometric model is too crude an approximation and compute the p-value using Poisson distribution. Furthermore, given a sentence S with a length L and containing the adjective A, the calculation of the probability of finding the antonym of A in the same sentence must account for the fact that one of the positions in the sentence is already taken by A. The positions where the antonym of A can be found are then L–1.

These factors are accounted for in the studies below. The statistical aspects of the method are described in greater detail in Holtsberg & Willners (to appear).

7.4 Experiments

A computer program, *Coco*, was developed by Anders Holtsberg in co-operation with the author to calculate the expected and observed co-occurrences as well as the probabilities, taking sentence-length variation into account (Holtsberg & Willners to appear. It was written in the programming language Icon (Griswold & Griswold 1983) and run on a SUN workstation. Its input is a list of words, i.e. the adjectives of interest, and a corpus. The output is a table presenting the expected number of co-occurrences, the number actually observed and the probability, according to the random-distribution hypothesis, of observing at least that number of co-occurrences – see Table 9 for an example. Where the probability value falls short of 10^{-4}, it is rounded to 0.

The co-occurrence hypothesis was proved to hold for data from both English and Swedish. First, Justeson & Katz's study was replicated using the Brown corpus, and accounting for variation in sentence length. The results for English were confirmed with data from the BNC. The SUC and Parole were used as test data for Swedish.

Coco was originally developed to analyse a whole corpus. However, the computing capacity of the machines available did not allow for such a treatment of the BNC, and the Parole corpus was available only through the Internet. The distribution of sentence lengths was obtained for both corpora, as well as the sentential occurrences and co-occurrences of the words in the test set. *Coco* was modified to perform the calculations in two steps: first the calculations to account for variation in sentence length in the corpus and then the calculations of *Expected*, *Ratio*, and *Probability* for each of the words in the test set.

7.4.1 Deese's adjectives, controlled for sentence-length variation

7.4.1.1 Corpus and test set

As in Justeson & Katz's study, a tagged version of Brown was used as test corpus. The test set used was the same 35 antonym pairs that Justeson & Katz used, which had previously been identified as antonyms by Deese (1965).

Words		*Sentential occurrences*		*Sentential co-occurrences*			
Adj. 1	***Adj. 2***	***N1***	***N2***	***Obs.***	***Exp.***	***Ratio***	***Prob.***
active	passive	86	11	2	0.02	99.03	0.0002
alive	dead	57	161	2	0.20	10.21	0.017
back	front	28	78	3	0.05	64.34	0
bad	good	127	694	16	1.88	8.50	0
big	little	312	275	13	1.83	7.10	0
black	white	152	250	23	0.81	28.35	0
bottom	top	3	70	0	0.00	-	-
clean	dirty	46	37	1	0.04	27.52	0.036
cold	hot	137	122	8	0.36	22.42	0
dark	light	148	62	5	0.20	25.52	0
deep	shallow	84	14	0	0.03	-	-
dry	wet	54	45	2	0.05	38.55	0.0013
easy	hard	109	138	0	0.32	-	-
empty	full	63	215	1	0.29	3.46	0.25
far	near	36	16	1	0.01	81.32	0.012
fast	slow	32	49	1	0.03	29.87	0.033
happy	sad	95	35	1	0.07	14.09	0.068
hard	soft	138	59	3	0.18	17.13	0.0008
heavy	light	110	62	1	0.15	6.87	0.14
high	low	418	138	19	1.23	15.43	0
inside	outside	6	38	0	0.00	-	-
large	small	351	505	26	3.78	6.87	0
left	right	67	214	13	0.31	42.47	0
long	short	522	191	12	2.13	5.64	0
narrow	wide	61	145	2	0.19	10.59	0.016
new	old	1024	629	30	13.75	2.18	0.0001
old	young	629	359	17	4.82	3.53	0
poor	rich	101	74	7	0.16	43.87	0
pretty	ugly	39	20	0	0.02	-	-
right	wrong	214	113	8	0.52	15.50	0
rough	smooth	40	35	1	0.03	33.46	0.029
short	tall	191	55	1	0.22	4.46	0.2
sour	sweet	4	63	1	0.01	185.88	0.0054
strong	weak	189	29	3	0.12	25.64	0.0002
thick	thin	66	90	1	0.13	7.89	0.12

Table 10: Sentential co-occurrences of Deese's adjective pairs in the tagged Brown Corpus. Probability values lower than 10^{-4} are rounded down to 0.

7.4.1.2 Results

Table 10 shows the sentential co-occurrence data found for Deese's adjectives. Five of the antonym pairs listed by Deese do not co-occur at all in the Brown Corpus: *bottom—top*, *deep—shallow*, *easy—hard*, *inside—outside*, and *pretty—ugly*. For most of the remaining antonym pairs, the number of co-occurrences is statistically significant: 25 of them are significant using a significance level of 0.05; 19 at the 0.01 level; and 14 at 10^{-4}.

The co-occurrence values observed differ somewhat from Justeson & Katz's, and so do the p-values. Unlike their study, the present one includes adjectives appearing in titles, e.g. *On active service*, in which case all the words involved are tagged to indicate that they appear in a title, e.g. ON/IN-TL ACTIVE/JJ-TL SERVICE/NN-TL. This explains why some of the observed occurrences are higher in this study than in Justeson & Katz's one.

The p-values calculated by Justeson & Katz are consistently lower than the p-values in this study, which take sentence-length variation into account. The differences in p-value are not great, but the overall ratio of *Observed* to *Expected* is 7.0, compared with Justenson & Katz's 8.6.[21] Despite this lower overall ratio, however, the co-occurrence hypothesis still appears to hold.

7.4.2 Confirming the results with a larger corpus

7.4.2.1 Corpus

The results from the study on the Brown corpus were confirmed using the BNC. The same method, accounting for variation in sentence length and for the fact that one position in the sentence is taken, was used.

The BNC is tagged for word class using the BNC Basic (C5) Tagset, which consists of 61 different tags. Most tags apply only to one-word units but some multi-word phrases are tagged as one word, e.g. 'in general', 'a little', 'such as', 'on behalf of', 'according to'. The corpora used for the Swedish studies do not recognise multi-word phrases, so there is a slight difference, but it should not be significant.

[21] The overall ratio is obtained by dividing the sum of all the expected values with the sum of all the observed ones.

Words		*Sentential occurrences*		*Sentential co-occurrences*			
Adj. 1	***Adj. 2***	***N1***	***N2***	***Obs.***	***Exp.***	***Ratio***	***Prob.***
active	passive	7289	1445	245	2.63	93.02	0
alive	dead	4252	11676	248	12.42	19.98	0
back	front	4501	9696	262	10.91	24.01	0
bad	good	26204	124542	1603	816.11	1.96	0
big	little	33688	29300	408	246.84	1.65	0
black	white	19998	19184	3162	95.94	32.96	0
bottom	top	3476	14651	492	12.74	38.63	0
clean	dirty	5161	2789	81	3.60	22.50	0
cold	hot	11596	9445	575	27.39	20.99	0
dark	light	12907	12396	393	40.01	9.82	0
deep	shallow	9818	1512	83	3.71	22.36	0
dry	wet	5463	3732	237	5.10	46.48	0
easy	hard	18822	18212	177	85.72	2.06	0
empty	full	5409	28529	134	38.59	3.47	0
far	near	7395	3782	39	6.99	5.58	0
fast	slow	6707	5760	133	9.66	13.77	0
happy	sad	12847	3587	80	11.52	6.94	0
hard	soft	18212	6626	246	30.18	8.15	0
heavy	light	10537	12396	183	32.66	5.60	0
high	low	56971	28903	2120	411.78	5.15	0
inside	outside	592	4750	40	0.70	56.88	0
large	small	47184	51865	2946	611.98	4.81	0
left	right	13892	40420	1640	140.42	11.68	0
long	short	40723	19703	1138	200.65	5.67	0
narrow	wide	5338	16812	138	22.44	6.15	0
new	old	115504	64991	3627	1877.22	1.93	0
old	young	64991	37259	1563	605.55	2.58	0
poor	rich	16579	7706	722	31.95	22.60	0
pretty	ugly	3925	1436	8	1.41	5.68	0.0001
right	wrong	40420	15512	753	156.79	4.80	0
rough	smooth	3507	3052	44	2.68	16.44	0
short	tall	19703	5302	138	26.12	5.28	0
sour	sweet	635	3537	47	0.56	83.68	0
strong	weak	19550	4523	376	22.11	17.00	0
thick	thin	5119	5536	128	7.09	18.06	0

Table 11: Sentential co-occurrences of Deese's adjective pairs in the British National Corpus. Probability values lower than 10^{-4} are rounded down to 0.

7.4.2.2 Results

All 35 word pairs studied co-occur significantly more often than expected. The overall ratio of *Observed* to *Expected* is 4.31. This is less than what was found in the investigation carried out on the Brown corpus, but it shows the same tendency.The ratios of a few word pairs are fairly low even though they are still statistically significant: *good—bad*, *big—little*, and *new—old* have ratios of less than 2.0. In all three cases the scales involved are quite complex, and some of the words actually designate the end point of more than one scale. *Old* is a true antonym not only of *new*, but also of *young* (which also shows a fairly low ratio of 2.6). *Big—little* is one of the scales involved in describing the field of SIZE, where we also find the scale *large—small*. The complexity of this field will be further discussed in the case study of *stor* 'large' and its synonyms. As for *good—bad*, *good* is an antonym of *evil* as well. The highly diverse usage of these words may partly explain their low ratios.

7.4.3 Data from Swedish

It is clear that direct antonyms co-occur more often than chance would predict in English. In the next section it will be shown that the co-occurrence hypothesis holds for Swedish, too.

7.4.3.1 Test set and test corpus

Justeson & Katz studied adjective pairs listed by Deese (1965) as antonymous. I attempted to translate this list into Swedish, but I soon ran into problems. Some of the words do not translate as Swedish adjectives, e.g. 'inside' and 'outside', whose equivalents in Swedish are the pronouns *inre* and *yttre* or the noun prefixes *inner-* and *ytter-*. In a couple of cases it was hard to find truly antonymous concepts in Swedish, e.g. 'near' and 'far', for which I was very hesitant to use either one of *nära* and *fjärran* and *nära* and *långt borta* ('far away'). Other problematic pairs were 'dirty—clean' and 'rough—smooth'. The translation of 'clean' is quite clearly *ren* but whether the direct antonym of *ren* is *smutsig* or *oren* is unclear. The other pair, 'rough' and 'smooth', may be rendered as e.g. *ojämn—jämn*, *ojämn—slät*, *bucklig—slät*.

To get round this problem, I chose to use a set of word pairs defined as antonyms by Lundbladh (1988):

bred—smal	‘wide—narrow’
djup—grund	‘deep—shallow’
gammal—ung	‘old—young’
hög—låg	‘high—low’
kall—varm	‘cold—warm’
kort—lång	‘short—tall’
liten—stor	‘small—big’
ljus—mörk	‘light—dark’
långsam—snabb	‘slow—fast’
lätt—svår	‘easy—difficult’
lätt—tung	‘light—heavy’
tjock—tunn	‘thick—thin’

The Stockholm-Umeå Corpus (SUC) served as the body of Swedish text searched for antonym co-occurrences.

7.4.3.2 Results

The results are presented in Table 12. All of the word pairs investigated co-occur more often than expected. There are three word pairs with a p-value over 0.0001: *bred—smal* ‘wide—narrow’, *djup—grund* ‘deep—shallow’, and *lätt—svår* ‘easy—difficult’.

The adjectives *bred* and *smal* mark the ends of the scale of WIDTH. However, *smal* is often used about people, e.g. *hon är smal* ‘she is thin’, in which case the direct antonym of *smal* is *tjock* ‘fat’ rather than *bred*. This may explain the somewhat higher p-value for *bred—smal*. The field of WIDTH is highly complex and there are several scales involved. *Tjock* is also a direct antonym of *tunn* ‘thin’; this is another relation showing significance in the table below.

The words *djup* and *grund* co-occur only once, reflecting the infrequent use of *grund*. Another word on the same scale that may be involved as an antonym of *djup* is *ytlig* ‘superficial, shallow’.

In the case of *lätt—svår*, once again, one of the members of the word pair has several direct antonyms. *Lätt* shows up as an end point on the scale of DIFFICULTY as well as on the scale of WEIGHT, where its opposite is *tung* ‘heavy’.

Looking at the ratio column instead of the probability column, we see that the ratio for *liten—stor* is very low compared with those for the other word pairs. Both words in this pair are very frequent and also highly polysemous. Moreover, the semantic field of SIZE is rich in expressions – the Strömberg thesaurus lists in all 77 synonyms of *stor* and *liten*. This may have an effect on the level of the ratio.

The overall ratio between *Observed* and *Expected* is 3.12, i.e. direct antonyms can be said to co-occur 3.12 times more often than chance would predict. If we exclude *liten—stor* from this calculation, the overall ratio is 8.17.

Words		***Sentential occurrences***		***Sentential co-occurrences***			
Adj. 1	***Adj. 2***	***N1***	***N2***	***Obs.***	***Exp.***	***Ratio***	***Prob.***
bred	*smal*	113	55	2	0.12	17.39	0.0061
djup	*grund*	117	17	1	0.04	27.17	0.036
gammal	*ung*	1050	455	47	8.84	5.32	0
hög	*låg*	760	333	47	4.68	10.04	0
kall	*varm*	102	102	12	0.19	62.32	0
kort	*lång*	262	604	21	2.93	7.17	0
liten	*stor*	1344	2673	111	66.48	1.67	0
ljus	*mörk*	84	126	7	0.20	35.82	0
långsam	*snabb*	55	163	4	0.17	24.11	0
lätt	*svår*	225	365	5	1.52	3.29	0.020
lätt	*tung*	225	164	7	0.68	10.25	0
tjock	*tunn*	53	85	4	0.08	47.98	0

Table 12: Sentential co-occurrences of Lundbladh's antonymous adjective pairs in the SUC. Probability values lower than 10^{-4} are rounded down to 0.

7.4.4 Confirming the results for Swedish with a larger corpus

7.4.4.1 The Parole corpus

The same test set as above was studied in the Parole corpus. This is tagged with morphosyntactic information, but not lemmatised, so each word form was counted individually; *N1* and *N2* are the sums of the occurrences of the individual word forms. The existence of several word forms had to be taken into account in the search for the sentential co-occurrences as well. Parole was searched for any form of *Adj. 1* co-occurring with any form of *Adj. 2* in any order.

7.4.4.2 Results

The results are presented in the table below. The findings from the study on the SUC are supported.

As in the SUC, the words *djup* and *grund* 'deep' and 'shallow' co-occur only once in the Parole corpus.

There are some extreme values in the table: *liten—stor*, which behaved differently from the other word pairs in the study above as well, is actually expected to co-occur 3.5 times more often than it does. This is due to the very high frequency of both words, especially *stor*, which is the most frequent adjective in Swedish. The scale of SIZE will be excluded from the further study

of sentential co-occurrence. A thorough study of the semantic range of *stor* will be reported it in the next chapter.

Excluding the field of SIZE, antonym co-occurrence takes place in 3.07 times as many sentences as expected, overall. Going one step further and excluding also all the word pairs where more than one antonym is involved, i.e. *gammal—ung*, *lätt—svår*, and *lätt—tung*, sentential co-occurrence of antonyms takes place in 4.54 times as many sentences as expected.

Words		***Sentential occurrences***		***Sentential co-occurrences***			
Adj. 1	***Adj. 2***	***N1***	***N2***	***Obs.***	***Exp.***	***Ratio***	***Prob.***
bred	*smal*	2241	870	23	1.59	14.46	0
djup	*grund*	1732	69	1	0.1	10.26	0.093
gammal	*ung*	18667	10022	172	152.58	1.13	0.07
hög	*låg*	14923	6358	205	77.38	2.65	0
kall	*varm*	1819	2430	71	3.60	19.69	0
kort	*lång*	5568	11395	164	51.74	3.17	0
liten	*stor*	18909	57989	254	894.31	0.28	1.0
ljus	*mörk*	1026	2001	47	1.67	28.07	0
långsam	*snabb*	711	3121	25	1.80	13.81	0
lätt	*svår*	5111	11646	56	48.55	1.15	0.16
lätt	*tung*	5111	3392	49	14.14	3.47	0
tjock	*tunn*	1002	1108	8	0.90	8.84	0

Table 13: Sentential co-occurrences of Lundbladh's antonymous adjective pairs in Parole. Probability values lower than 10^{-4} are rounded down to 0.

7.4.5 Conclusion

All four case studies above confirm that direct antonyms co-occur sententially more often than chance would predict. The first study is a replication of Justeson & Katz's (1991) study of Deese's antonyms in the Brown corpus, using a further-developed version of their method – controlling for sentence-length variation and the decreased number of possible positions available for a given lemma in a sentence already containing another given lemma. The results of the first study were confirmed using data from the BNC.

The two subsequent studies apply the same method to Swedish corpora – the SUC and Parole – and show that the co-occurrence hypothesis holds for Swedish as well.

The overall ratios between *Observed* and *Expected* range from 3.17 to 7.0. Even though the choice of corpus for the study and the choice of word pairs included in the test set can have a certain impact, it is clear that direct antonyms co-occur in the same sentence more often than chance would predict.

7.5 Rewriting the co-occurrence hypothesis

7.5.1 Introduction

Language is not random. The order of the words in a sentence is governed by grammar and there are semantic restrictions on the use of each word. There may be various reasons why two words co-occur significantly often. Concerning antonyms, I think it is important to compare the results for each antonym pair with co-occurrence frequencies for other word pairs on the same scale.

In this section I will discuss the importance of comparing the co-occurrence frequency of direct antonyms with that of indirect antonyms. The co-occurrence hypothesis will be rewritten and data will be presented supporting the rewritten hypothesis.

7.5.2 Reasons for co-occurrence

In the previous section it has been shown that antonymous adjectives co-occur more frequently than they would be expected to if their distribution were random. However, there are other reasons for adjectival co-occurrence than the lexical relation of antonymy.

Idiomatic expressions give rise to highly frequent co-occurrence patterns that do not necessarily involve an antonym relation. Bendz (1965) carried out an extensive study on co-ordinated word pairs such as *liv och död* 'life and death', *dag och natt* 'day and night' in Swedish as well as eight other languages from Romance and other Indo-European language families. He distinguishes three different types of word pairs:

- *antonyms*, e.g. *liv och död* 'life and death', *gammal och ung* 'old and young', *stort och smått* '(things) large and small';
- *enumeratives*, e.g. *adel, präster, borgare och bönder* 'the nobility, the clergy, the burghers, and the peasants';[22]
- *synonyms*, e.g. *tyst och stilla* 'quiet and calm', *lugn och ro* 'peace and quiet', *lös och ledig* 'loose and free', *(inom) lås och bom* '(under) lock and key'.

Bendz does not categorise the different word pairs according to word class, but apart from the semantic aspects he discusses some formal aspects as well. The word pairs are often held together by alliteration, e.g. *stor och stark* 'big and strong', assonance, where the vowel is the bearing unit, e.g. *stor och grov* 'large

[22] This is not strictly speaking a *word pair*, though it obviously qualifies according to Bendz's definition.

and heavy, and rhymes such as *rätt och slätt* 'right and plain' and *smått och gott* 'little and good = nice little things'.

The words are most often conjoined with *och* 'and'; the order between the words is generally frozen. When antonymous words are involved, the positive, most important, closest, or most obvious concept precedes the other, e.g. *rik och fattig* 'rich and poor', *hög och låg* 'high and low', *glädje och sorg* 'joy and sorrow'. The length of the words also matters. Usually both have the same number of syllables, but when they are asymmetrical the combination monosyllabic word + disyllabic word appears in both orders, e.g. *tigga och be* 'beg and plead' and *ler och långhalm* 'clay and straw = as thick as thieves'. In combinations of mono/disyllabic + polysyllabic (more than two syllables) words, the shorter word almost always comes first, e.g. *bitter och besviken* 'bitter and disappointed'.

Though antonymous concepts tend to show up in idiomatic expressions, so do other words. Bendz's three types of word pairs have one feature in common: the words forming the pairs almost always belong to the same semantic field.

Apart from antonyms appearing in idiomatic expressions, opposing concepts can take other functions in discourse. Jones (1998) has made an extensive classification of contrastive meanings according to function in text. He distinguishes the following classes of co-occurring antonyms in texts:

- *Ancillary antonymy*: multi-contrast sentences where the antonymous pair helps to signal another less obvious contrast in the sentence.

  There is no doubt that the legal department was ***right*** and the social workers ***wrong***.

- *Umbrella antonymy*: expresses inclusiveness or exhaustiveness and the antonymous pair is conjoined by *and* or another conjunction.

  There is a valid argument that leaders need as much information as possible, both ***bad*** and ***good***, before allowing a customer to borrow money.

- *Distinguished antonymy*: when the distinction beween the two antonyms is explicit, usually indicating a *difference*, *gap*, or *division* between them.

  God has given us the ability to distinguish between ***right*** and ***wrong*** and to shoulder responsibility.

- *Idiomatic antonymy*: the antonymous pair takes part in an idiom.

  The same ability to blow ***hot*** and ***cold*** was also displayed in August when he invited journalists for a friendly get-together on a Hong Kong beach.

- *Comparative antonymy*: the antonyms are weighted relative to one another.

  Sometimes I feel more ***masculine*** than ***feminine*** and I don't like it.

- *Extreme antonymy*: a type of umbrella antonymy where the antonyms refer to the extremes of the scale rather than exhausting the entire scale.

  This tough little bug, scientists told us, would die only in a very ***hot*** or a very ***cold*** temperature sustained over a period of time.

- *Transitional antonymy*: the sentences express a change on the scale designated by the antonymous pair.

  He turns ***bad*** ball into ***good*** ball and invariably makes the right decisions.

- *Oppositional antonymy*: sentences on the form *X, not Y*.

  For a while, no one would buy any cheese made with unpasteurised milk, despite the fact that listeria cases were caused by ***soft***, not ***hard*** cheeses.

The co-occurrence of antonyms can signal many different functions. We can conclude that there are many reasons for words in general and antonyms in particular to co-occur.

7.5.3 Something to compare with: The rewritten co-occurrence hypothesis

Muehleisen (1997) approaches the question of differences in co-occurrence behaviour between direct and indirect antonyms in her dissertation:

"The reason why some words have no clear antonym is because they simply do not co-occur often enough with any semantically contrasting word for this lexical association to develop, and presumably the explanation why some words have more than one antonym must be that these words frequently occur with more than one semantically contrasting word. However, I have not found any research that directly tests this by looking at the co-occurrence patterns of words that do not have antonyms or word that have two antonyms."

I have thought along the same lines, but with the following reasoning: since there are many possible explanations why two words co-occur more frequently than chance would allow, showing that direct antonyms co-occur more often than would be expected if their distribution were random does not actually prove anything. Words belonging to the same semantic field often co-occur and since the two members in an antonymous pair also belong to the same semantic field it would actually be quite surprising if they did not co-occur rather often!

A more fruitful way to go about this problem is to compare the frequency of co-occurrence of direct antonyms, such as *strong* and *weak*, with that of indirect

antonyms, such as *powerful* and *weak*. The co-occurrence hypothesis for antonymous adjectives could then be rewritten as follows.

> The co-occurrence hypothesis for direct antonyms vs indirect antonyms:
> Two adjectives are learned as direct antonyms because they occur together in the same sentences more frequently than indirect antonyms on the same scale.

7.6 Proving the rewritten co-occurrence hypothesis

To prove that direct antonyms co-occur significantly more often than indirect antonyms do, eleven semantic fields were chosen. These are the ones that Lundbladh gives as examples of true antonymy (1988:25) and that I have previously shown to co-occur more often than chance would predict. As previously mentioned, the scale designated by *liten—stor* was excluded from the test set. The Strömberg (1995) thesaurus was used to find the indirect antonyms of each field. Because of the tokenisation of the SUC, multi-word phrases were excluded, but apart from that there was no further analysis of the synonyms listed by Strömberg. The words studied in each semantic field are listed in the Appendix. Expected and observed co-occurrences, as well as the probabilities for the indirect antonyms, were calculated using the previously mentioned program, *Coco*.

All synonyms of the direct antonyms were used in the study and they were combined to form all possible indirect-antonym pairs. The sentential co-occurrences of over 7,000 indirect antonyms were studied.

The columns for indirect antonyms in the table below show the sum of all the observed and expected sentential co-occurrences of all the indirect antonyms in the field, respectively. The totals for observed and expected co-occurrences were used to calculate the ratio of *Observed* to *Expected* and the probability of finding as many co-occurrences as were actually observed or more.

Scale	*Sentential co-occurrences*			
	Observed	*Expected*	*Ratio*	*Prob.*
bred—smal	48	27.11	1.77	0.0002
djup—grund	1	0.20	5.13	0.1813
gammal—ung	101	54.69	1.85	0
hög—låg	360	263.86	1.36	0
kall—varm	6	0.86	6.93	0.0003
kort—lång	169	121.92	1.39	0
ljus—mörk	57	26.42	2.16	0
långsam—snabb	6	2.03	2.95	0.018
lätt—svår	16	18.60	0.86	0.76
lätt—tung	18	14.05	1.28	0.18
tjock—tunn	23	7.65	3.00	0

Table 14: Observed and expected sentential co-occurrences of indirect antonyms in eleven different semantic fields in the SUC. Probability values lower than 10^{-4} are rounded down to 0.

Most of the indirect antonyms, like the direct ones, co-occur significantly more often than chance would predict: 8 of the 11 scales are significant using a level of 0.05, 7 at 0.01, and 5 at 10^{-4}.

However, the ratios are fairly low. For the majority of the semantic fields, the number of observed co-occurrences of indirect antonyms is quite close to the number expected with random distribution. In one case the observed value falls below the expected one: for *lätt—svår*. We expect to find 18.6 sentential co-occurrences of indirect antonyms from the scale *lätt—svår*, but find only 16. However, this deviation would not seem to be significant.

Indirect antonyms from the field of DEPTH co-occur only once but yet yield a ratio of 5.13, owing to the low number of expected co-occurrences. The co-occurrence in this field is an accidental one: *mörk* 'dark' and *flack* 'flat' co-occur in the phrase *flacka slätter täckta med mörk lavasand* 'flat plains covered in dark lava sand'. *Mörk* and *flack* are indirect antonyms through *djup—grund* 'deep—shallow', cf. the Appendix.

The number of co-occurrences of indirect antonyms exceeds the number of co-occurrences of direct antonyms on all scales but two, i.e. *varm—kall*, where the direct antonyms co-occur twelve times and indirect ones six times, and *djup—grund*, where both direct and indirect antonyms co-occur only once each.

The sentential co-occurrences of direct vs indirect antonyms are compared in Table 15.

Scale	*Sentential co-occurrences*					
	Direct antonyms			*Indirect antonyms*		
	Observed	*Expected*	*Ratio*	*Observed*	*Expected*	*Ratio*
bred—smal	2	0.12	17.39	48	27.11	1.77
djup—grund	1	0.04	27.17	1	0.20	5.13
gammal—ung	47	8.84	5.32	101	54.69	1.85
hög—låg	47	4.68	10.04	360	263.86	1.36
kall—varm	12	0.19	62.32	6	0.86	6.93
kort—lång	21	2.93	7.17	169	121.92	1.39
ljus—mörk	7	0.20	35.82	57	26.42	2.16
långsam—snabb	4	0.17	24.11	6	2.03	2.95
lätt—svår	5	1.52	3.29	16	18.60	0.86
lätt—tung	7	0.68	10.25	18	14.05	1.28
tjock—tunn	4	0.08	47.98	23	7.65	3.00

Table 15: Observed and expected sentential co-occurrences in the SUC for direct and indirect antonyms in eleven different semantic fields.

The indirect antonyms co-occur overall 1.45 times more often than would be expected if their distribution were random, while the direct antonyms co-occur overall 3.12 times more often than expected. Based on these eleven semantic fields it is clear that direct antonyms co-occur significantly more often than indirect antonyms.

The differences in ratios between indirect and direct antonyms are visualised in the diagram below.

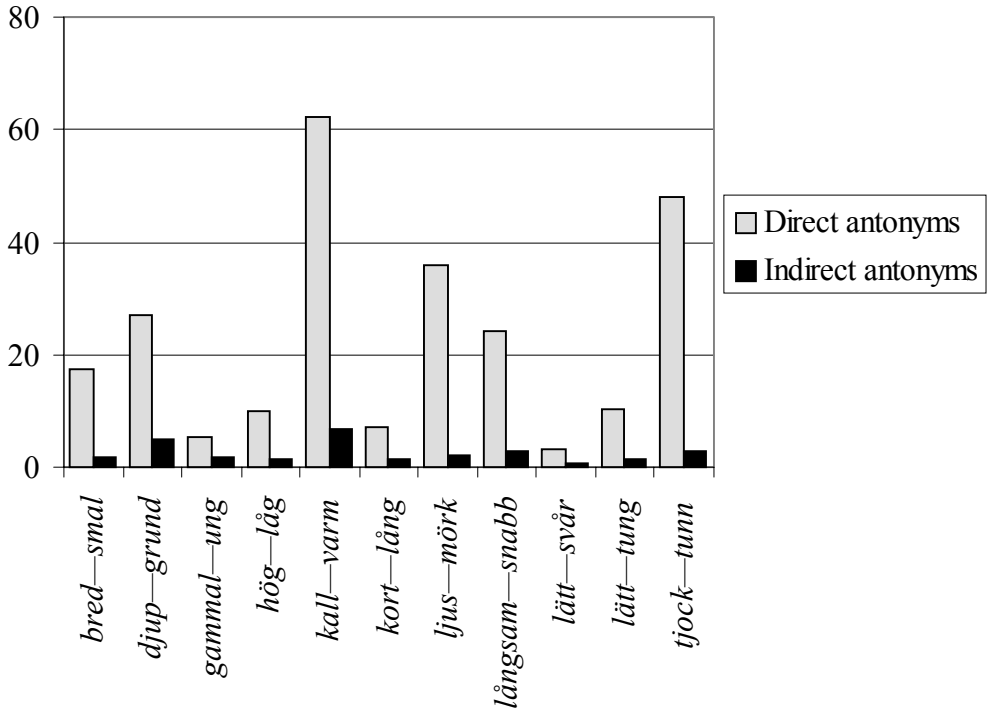


Figure 8: A comparison of the ratios of direct and indirect antonyms on the same scales in the SUC.

7.6.1 Discussion

The results show that both direct and indirect antonyms co-occur sententially more often than chance would predict. Indirect antonyms co-occur slightly more often than would be expected if their distribution were random, 1.45 times more often than expected in the SUC, while the overall ratio for direct antonyms is 3.12. Adjectives with more than one antonym tend to pull down the ratio for all the word pairs it is part of, e.g. *lätt,* which appears in the antonym pairs *lätt—tung* 'light—heavy' and *lätt—svår* 'easy—difficult'.

The adjectives in the study were not categorised according to meaning. It is likely that separate treatment of e.g. *lätt* in its two meanings would produce statistically significant results for both antonym pairs involved.

Chan (2000) has recently replicated this study for English. He has elaborated the calculation of the ratio, but he does not account for variation in sentence length. His studies of co-occurring direct and indirect antonyms in the BNC confirm the findings reported above.

It is clear that antonym pairs where neither word is a member of several antonym pairs co-occur significantly more often in the same sentence than other word pairs on the same scale. This may be a cue for the language learner to realise that two words are antonyms. The substitutability hypothesis may provide another cue.

7.7 The substitutability hypothesis

Justeson & Katz (1991) also investigated the sentences in which the co-occurring antonyms were found and presented data for English that confirm the substitutability hypothesis, i.e. that the co-occurring antonyms seem to be interchangeable in most contexts. I have studied the sentential co-occurrences of Lundbladh's twelve antonym pairs in the SUC, in all 269 sentences, and classified them as either antonym co-occurrences or accidental co-occurrences. In the accidental co-occurrences, there is no obvious relation between the noun phrases in which the two antonyms occur, e.g.

> *Universitetet är ett helt* ***litet*** *samhälle för sig – med ljus, liv och hemkänsla* ***större*** *delen av dygnet.*
>
> 'The university is a *small* community in itself – with lights, life and "home feel" during the *greater* part of both day and night.'

The contexts in which the opposites occurred were of several different types. To avoid a substantial number of sentences in the group *Other*, the set of categories used by Justeson & Katz (1991) was supplemented with the categories *Elliptic*, *Semantically related NP*, and *Semantically related N*. The 269 sentences were classified using the following categories:

- *Co-ordinated construction*, e.g.

  *…skillnaderna mellan **korta** och **långa** a-ljud...*

  '…the differences between *short* and *long* a-sounds…'

- *Identical NP*, where the whole phrase is identical, i.e. the syntactic construction of the NP as well as the constituent words are identical, though differences in inflection are permitted, e.g.

  *Studier […] visar att den **snabbaste** användaren ofta löser givna ordbehandlingsuppgifter fem gånger snabbare än den **långsammaste** användaren.*

  Studies […] show that the *quickest* user often solves given word-processing tasks five times faster than the *slowest* user.

- *Identical N*, where the heads of the noun phrases are the same, but the rest of the NP may diverge, e.g.

  I ett experiment jämförde han en grupp lässvaga äldre elever med en grupp läsnivåmatchade yngre elever...

  'In one experiment he compared a group of *older* pupils who had a reading disability with a group of *younger* pupils matched for reading level…'

- *Identical PP*, where the antonyms are part of identical prepositional phrases, e.g.

  *En skärm med **hög** upplösning […] kommer att generera flacka linjer på ett bättre sätt än en skärm med **låg** upplösning.*

  'A screen with a *high* resolution […] will generate flat lines in a better way than a screen with a *low* resolution.'

- *Elliptic*, where the head of one of the noun phrases is left out, e.g.

  ***Yngre** patienter dyrare än **äldre**.*

  '*Younger* patients more expensive than *older* (ones).'

- *Semantically related NP*, a sentence where the antonyms co-occur in identical syntactic constructions and where the head nouns are not identical

but semantically related. They may be antonyms, synonyms, hypo/hypernyms, or meronyms, e.g.

*Då försäkringstagaren betalat in en **liten** premie men försäkringsbolaget lovar att betala ut en **stor** summa till de efterlevande om försäkringstagaren dör är bolagets risk stor.*

'Since the policy holder has paid a *small* premium but the insurance company promises to pay a *large* amount to the survivors if the policy holder dies, the company's risk is big.'

- *Semantically related N*, sentences where the antonyms do not appear in parallel constructions syntactically, but the nouns they modify are semantically related, e.g.

  *…i centrum av filmen står också tre munkar; en mycket **ung** novis, en man som tvekar om sin livsväg och en **gammal** döende mästare.*

  '…in the centre of the movie, there are also three monks: a very *young* novice, a man doubting whether he has made the right choice, and an *old* dying master.'

- *Other*, sentences not falling into any of the categories above but where the co-occurrence is not accidental, e.g.

  *På **hög** höjd är lufttrycket **lågt**.*

  'At *high* altitudes, atmospheric pressure is *low*.'

It is clear from the table below that the major part of the observed co-occurring antonyms are used as opposites and not just accidentally show up in the same sentences. Of the 269 sentences examined, 11% (30) were categorised as accidental occurrences of antonymous concepts.

The result of the classification is presented in the table below.

Syntactic context	***Sentential co-occurrences***	***Sentential co-occurrences (%)***
Semantically related N	60	22.3%
Co-ordinated adjectives	52	19.3%
Elliptic	46	17.1%
Accidental	30	11.2%
Identical N	23	8.6%
Semantically related NP	17	6.3%
Identical NP	16	5.9%
Identical PP	2	0.7%
Other	23	8.6%

Table 16: Semantic contexts of antonym co-occurrences in the SUC corpus.

In 49.3% of the sentences studied the antonymous adjectives appeared in parallel syntactic constructions, i.e. identical NPs, identical PPs, co-ordinated adjectives, elliptic constructions, and semantically related NP (where the

syntactic construction is parallel but the head nouns of the two noun phrases are only semantically related). 30.9% are identical or semantically related nouns that do not appear in parallel constructions. The remaining 19.8% are classified as "accidental" or "other". Altogether, 51.6% of the studied sentences were constructions with an identical head, i.e. where both adjectives have the same head, as in the co-ordinated and elliptic constructions, or where their heads are two identical instances of the same word in two different noun phrases.

The majority of the co-occurring antonyms in the SUC are found in identical or near-identical contexts. Thus, it can be concluded that the substitutability hypothesis holds also for Swedish. The frequency of co-occurrence may play a role in learning direct antonyms, but it is clear that the direct antonyms are actually used in parallel contexts in the sentences where they co-occur.

7.8 Prosody as a cue for word association

The studies just presented concerning antonyms and their behaviour in corpora show that antonyms do co-occur more often than chance would predict, and significantly more often than other opposing word pairs on the same scale. Furthermore, to a large extent the co-occurring members of word pairs are found in parallel contexts. This may help the language learner to form an association between the two words in an antonymous pair. However, these studies have been made on written corpora, which children learning their mother tongue do not normally come into contact with.

Preliminary studies show some interesting characteristics of co-occurring antonyms in spoken language. Studying the child-directed adult speech in Strömqvist's and Richthoff's longitudinal corpus of 5 Swedish children from around 18 to 48 months old (Richthoff 2000, Strömqvist et al. forthcoming) a clear trend is found: antonyms co-occurring in the same utterance are emphasised with focal accents.

Together with the high frequency of co-occurring antonyms and the parallelism of the contexts in which they appear, I think the prosodic cue is a strong help for the language learner to form the association between two antonymous words. Since a focal accent is easier to distinguish than the characteristic features of the co-occurrence hypothesis and the substitutability hypothesis, the prosodic cue may be even more important to the language learner than frequency and substitutability.

7.9 Distinguishing antonyms by looking at co-occurrence patterns

In the chapter above it has been shown that direct antonyms co-occur more often than indirect antonyms on the same scale. Since this is the case it should be possible to distinguish direct antonyms automatically by comparing the probabilities and ratios of all word pairs. This reasoning is circular, but it may still be useful.

7.9.1 Experiment

The Icon program described above, *Coco*, was run on a list of all the adjective lemmas that occur five or more times in the SUC, i.e. about 1,500 of the total 5,759 adjective lemmas. The SUC was used as test corpus.

The results were sorted on fallingrising probability.[23] 357 word pairs were significant at a significance level of 10^{-4}, and all of them share a semantic feature of some type. Table 17 below shows the ten most significant word pairs found.

1.	*höger* 'right'	*vänster* 'left'
2.	*kvinnlig* 'female'	*manlig* 'male'
3.	*svart* 'black'	*vit* 'white'
4.	*hög* 'high'	*låg* 'low'
5.	*inre* 'inner'	*yttre* 'outer'
6.	*svensk* 'Swedish'	*utländsk* 'foreign'
7.	*central* 'central'	*regional* 'regional'
8.	*fonologisk* 'phonological'	*morfologisk* 'morphological'
9.	*horisontell* 'horizontal'	*vertikal* 'vertical'
10.	*muntlig* 'oral'	*skriftlig* 'written'

Table 17: The 10 top co-occurring adjective pairs when sorted on rising p-value

The majority of the word pairs are classifying adjectives, and they share the same semantic range, and most often they belong to the same semantic field as well, e.g. the linguistic terms *fonologisk—morfologisk* 'phonological—morphological' and the terms denoting university faculties *humanistisk—samhällsvetenskaplig* 'of Arts—of Social Sciences'. There are 59 antonym pairs among the 357 word pairs that are significant at the chosen level.

The list produced as described above does not find only antonym pairs, but also other lexically related words. Interpreted with care such a list is a useful tool for a lexicographer.

[23] I.e. the probability of finding the number of co-occurrences actually found or more, see definition above.

7.10 Chapter summary

This chapter deals with the co-occurrence of adjectives. It takes as its starting point Justeson & Katz's study of the co-occurrences of antonymous adjectives. Confirming their results, it is shown that direct antonyms co-occur significantly more often than chance would predict in both English and Swedish. However, since there are various reasons for words to co-occur, the co-occurrence hypothesis is revised: direct antonyms co-occur more often than indirect antonyms. Studies of indirect antonyms in the SUC show that they co-occur less frequently than direct antonyms do.

The substitutability hypothesis says that "two adjectives are learned as direct antonyms because they are interchangeable in most contexts, i.e. because any noun phrase that can be modified by one member of the pair can also be modified by the other". The contexts of the co-occurring antonyms are examined and it is found that the antonyms are generally found in identical contexts in the sentence. Justeson & Katz's findings from English are confirmed using Swedish data.

The language learner gets several types of cues to form the association between opposing concepts. The main studies in this chapter point out the importance of frequency and parallelism of contexts. However, I believe there is a third important factor involved, namely prosody. I hypothesise that the co-occurring antonyms are specially marked with a focal accent in spoken language, further facilitating the acquisition of the semantic relation of antonymy.

Studying all the word pairs of a certain type in a corpus will provide a list of statistically significant co-occurrences that is a useful tool for the lexicographer. All significant word pairs (adjective pairs) found here turned out to be semantically related. Apart from antonymy, relations such as synonymy and hyponymy were found. This supports the choice of synonymy and hyponymy as the main relations coded in WordNet. A list of significant co-occurrences can also be used to check important relations in the lexicon while it is being developed.

8 Co-occurrence of adjectives with nouns

"You shall know a word by the company it keeps"
(Firth 1957)

Though many words are said to have synonyms, it is not often that the words cover exactly the same semantic ground. True synonyms, such as *tussilago* and *hästhovsört*, which designate the same plant, 'coltsfoot' (*Tussilago farfara L.*), are uncommon. Nouns and adjectives in WordNet are organised in synsets, with synonymous or semantically similar words clustered together to form a set. Each synset signifies a concept – it does not explain anything about the concept, only signifies its existence.

Taking a closer look at the adjectives, it is clear that the words within a synset are not always synonyms. For example, *stark* and *kraftig* are synonyms in the following context: *stark kritik—kraftig kritik*, both meaning 'strong criticism'. But if we change the head of the noun phrase to *arbetsgivare* 'employer', the phrases do not share the same meaning anymore. *Stark arbetsgivare* means 'strong employer', where *stark* 'strong' says something about the employer's power and skills in leading the company, i.e. not as a person but as an organisation, while *kraftig arbetsgivare* says something about the physical constitution of the employer as a person, and rather makes you think of an overweight man in his mid-fifties. The problem of polysemy is reflected in WordNet by the word's appearance in several synsets. *Stark* and *kraftig* in the first example belong to the same synset {stark, kraftig} headed by the antonym pair *stark—svag*, while in the second example, *kraftig* is coded as a synonym of *tjock* 'fat'.

My hypothesis is that data on the frequencies of the semantic categories an adjective modifies are useful when carrying out a semantic analysis of the adjective. The frequency information can be used to group adjectives according to meaning. It is also possible to add knowledge to the lexicon about what types of nouns an adjective can modify.

In the studies of sentential co-occurrence of direct antonyms, several problematic word pairs were found. Some word pairs did not behave as expected according to the co-occurrence hypothesis, i.e. word pairs that were assumed to be direct antonyms were not found to co-occur more often expected under the assuption that the words in the corpus are randomly distributed. One of these word pairs was *full—tom* 'full—empty', of which not a single sentential co-

occurrence was found in the one-million-word corpus SUC and only a few non-significant ones in the 25-million-word corpus Parole.

It has been argued that two words must correspond in semantic range to qualify as direct antonyms (Muehleisen 1997). The semantic ranges of *full* and *tom* will be investigated and described in the following chapter. These semantic ranges will then be compared to see if and how they overlap.

The dimension of SIZE is known to be problematic in English; there exist two different scales: *big—little* and *large—small*. In some contexts both scales are applicable, in others not (Muehleisen 1997 and Biber et al. 1998:43ff). Swedish has only one scale for SIZE, *stor—liten*, but even though these two words are highly frequent, they do not co-occur significantly often in Parole, and their number of co-occurrences in the SUC is significant but the ratio is unusually low. There may be only one scale for SIZE in Swedish, but there are in fact many words describing dimensional properties. An investigation of the semantic ranges of *stor* and its synonyms will be presented and the results will be used to group the synonyms according to semantic range.

The rest of this chapter will present two case studies concerning semantic range: one concerning *full—tom* and one concerning *stor* and 28 synonyms of *stor*.

8.1 A case study: The semantic ranges of *full* 'full' and *tom* 'empty'

8.1.1 Introduction

The word pair *full—tom* did not behave as expected in any of the studies of the sentential co-occurrence patterns of antonyms. In both the SUC and Parole, these two words co-occur less often than expected, as can be seen in Table 18. In the SUC they actually do not co-occur at all. In Parole, they would be expected to be found in the same sentence seven times, but in fact co-occur only five times.

Corpus	***Word1***	***Word2***	***Sentential occurrences***		***Sentential co-occurrences***			
			N1	***N2***	***Obs.***	***Exp.***	***Ratio***	***Prob.***
SUC	*full*	*tom*	164	64	0	0.19	0.00	1.0
Parole	*full*	*tom*	3327	1635	5	7.15	0.70	0.84

*Table 18: Sentential co-occurrences of **full—tom** in the SUC and Parole.*

Of the five sentential co-occurrences of *full* and *tom* in Parole, only the following one is a clear case of opposition:

*Och lika plötsligt som baren blev **full**, blev den **tom**.*

'And as suddenly as the bar had become *full*, it became *empty*.'

In the remaining four sentences, which are listed below, the word pair is not used in clear opposition, i.e. the nouns that the two adjectives modify are not identical or closely semantically related. However, the use of antonymous concepts may be a way to build up a discourse even if the antonyms do not modify the same noun (Jones 1998). In the first example below, *full* is used in the meaning 'drunk', whose antonym is *nykter* 'sober' rather than *tom* 'empty', but it is nevertheless playfully used as an antonym of *tom*. There is a difference in terms of content: the person is empty of feelings, but full of alcohol. In the other examples, the heads modified by the two antonyms belong to quite different semantic categories: *ögon fulla av* 'eyes full of'—*tomt vemod* 'empty sadness', *full aktivitet* 'full activity'—*skolan låg tom* 'the school lay empty', *tomma fickor/mage* 'empty pockets/stomach'—*full käft* 'full mouth'.

***Tom**, tänkte hon, jag är alldeles **tom** fast jag är **full**...*

'*Empty*, she thought, I am totally *empty* though I'm *drunk* [lit. *full*]…'

*Det såg ut som om hans ögon skrattade, men hon visste att de inte gjorde det, för om man såg in i dem såg man att de var **fulla** av vemod, inte ett **tomt**, själlöst vemod utan ett vemod som var fyllt av insikt.*
'It looked as if his eyes were laughing, but she knew that they were not, because if you looked into them you saw that they were *full* of sadness, not an *empty* soulless sadness but a sadness filled with insight.'

*Lokalvårdarna skurade inför terminsstarten och på expeditionen var det **full** aktivitet, men för övrigt låg skolan fortfarande **tom** och öde.*
'The cleaners were scrubbing the floors with the new term about to begin and there was *full* activity in the office, but otherwise the school still lay *empty* and deserted.'

*Jag vet vad de går för, **tomma** fickor, **tom** mage, men käften **full** av stora ord.*
'I know their sort, *empty* pockets, *empty* stomach, but their mouths *full* of big words.'

8.1.2 Previous work on full *and* tom

No word-association tests that cover the word pair *full—tom* have, to my knowledge, been carried out for Swedish. It would be interesting to see if subjects would actually respond with *full* when given the stimulus *tom*. However, the corresponding English antonym pair, *full—empty*, is found among the 35 word pairs listed by Deese (1965) as antonyms in English. In the study of sentential co-occurrences of antonyms in the Brown corpus by Justeson & Katz (1991), where Deese's antonyms form the test set, *full* and *empty* do not co-occur significantly often either. One sentence containing both words is found in Brown, while the expected value is calculated at 0.25. This yields a ratio of 4.0, but the probability of finding one sentence or more with the two words co-occurring is not significant; it is reported as 0.22.

Most speakers of Swedish would intuitively agree that the opposite of *tom* is *full*, so why do these words not co-occur sententially as often as other antonyms?

8.1.2.1 Rusiecki's and Lundbladh's work

Rusiecki (1985) gives *full—empty* as an example of "bounded-scale" antonym pairs. Words forming a pair on such a scale are incompatible and fully reciprocal adjectives; both ends of the scale are bounded and the scale is symmetric. It can be visualised as follows:

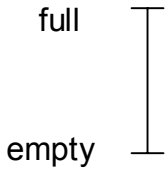


*Figure 9: The **full—empty** scale according to Rusiecki (1985).*

This category of word pairs is characterised by "the fact that the value of the measure function is not a denominate number". These words cannot be used in constructions such as 'NP is *n* units *A*', e.g. **the bottle is four litres/teacups full*, but are used in sentences of the type 'NP is *n A*', e.g. *the bottle is two-thirds full*.

Another feature is that neither member is marked. Both examples below are equally acceptable:

How full is the tank?

How empty is the tank?

Rusiecki provides only one example of bounded-scale antonym pairs, namely *full—empty*. Lundbladh (1985:32) does not agree with Rusiecki's classification of *full—empty*. In his taxonomy, *tom* and *full* are the end points of two separate scales, both belonging to the "non-binary asymmetric" type of scale, along with e.g. *fullständig* 'complete', *fullkomlig* 'perfect', *fulländad* 'perfect', *komplett* 'complete', and *slutgiltig* 'definitive'. Using Swedish examples, he claims that the comparative forms of *full* and *tom*, *fullare* and *tommare*, are not semi-reciprocal, and suggests instead two separate scales. The "EMPTINESS" scale is characterised by the concept *empty* as its end point, while the "FULLNESS" scale has *full* as its end point, see Figure 10.

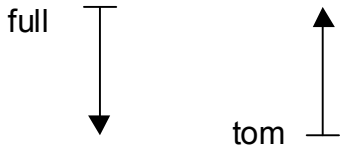


Figure 10: Scales of FULLNESS and EMPTINESS according to Lundbladh (1988).

Lundbladh's claim that *full* and *tom* do not belong to the same scale and are not antonyms is supported by the finding that these two words do not display the same co-occurrence pattern as non-disputed antonym pairs.

In a dissertation by Victoria Muehleisen (1997) semantic range is used to show why two words are antonyms. She claims that words in an antonym pair must have the same semantic range. Could the key to the problem be that *full* and *tom* do not have the same semantic range?

The semantic range of a word is "a reflection of a word's meaning in terms of its typical range of use" (Muehleisen 1997:201). Through the study of a word's collocation patterns it is possible to gain more knowledge about the word's meanings and usage. For an adjective, a study of the nouns it modifies should give a fairly good idea of its semantic range. Muehleisen studied collocation patterns for a couple of word pairs by calculating the of mutual information ("MI"), which measures the strength of the association between two words. Unfortunately, the traditional calculation of MI postulates a fixed window span. Because of this, adjectives that are not very close to their head noun, e.g. adjectives in predicative position, are not covered.

No measure of the strength of association will be calculated in this study. Raw frequency data on what semantic categories the adjectives were found to modify will be presented.

The goal of the study is to describe and compare the semantic ranges of *full* and *tom*. My hypothesis is that the semantic ranges of the two words are significantly diverse.

8.1.3 Clues from dictionaries

The modern Swedish dictionary *Nationalencyklopedins ordbok* ("NEO") lists the following meanings for *full*:

(1) containing as much as possible

(2) occurring to a great extent, e.g. *full storm* 'full gale', *full sommar* 'high summer'

(3) intoxicated

The meanings listed for *tom* are:

(1) lacking concrete content

(2) lacking meaningful content

Comparing the entries for *full* and *tom* in the NEO yields some obvious differences. *Full* is more polysemous than *tom*, with three meanings as opposed to only two for *tom*. This is not surprising: it has been mentioned earlier that high-frequency words are generally more polysemous than low-frequency words, and *full* is more than twice as frequent as *tom* in the two corpora studied. The lexical entries in the NEO do not say anything about the semantic ranges of the two words.

The entry for *full* in *Svenska Akademiens ordbok* ("SAOB") is more exhaustive than the one in the NEO. The SAOB describes Swedish from a diachronic perspective based on data from almost five centuries. Twelve meanings for *full* are listed, of which two are obsolete. The SAOB distinguishes four main meanings: (1) the notion of container "containing as much as it could", (2) "rich in", (4) "without cavity", and (11) "intoxicated". The remaining meanings are more or less closely related to meanings (1), (2), (4), and (11). The entry for *tom* has not yet been published. Generally, the SAOB gives quite a good idea of the semantic ranges of the words it describes in a diachronic perspective. This study concerns the synchronic meanings of *full* and *tom*. Therefore, modern material from the SUC and Parole will be used to describe the semantic ranges of the two words, which will then be compared to see if they overlap.

8.1.4 Method

The data for the first study were taken from the SUC: all forms of *full* and *tom* were extracted. They were then run through *Para* and imported to a database in FileMaker Pro where they were categorised semantically. The data from the Parole corpus was treated the same way.

8.1.4.1 Semantic classification

The semantic categories of the Princeton WordNet were used for the classification of the nouns.

First, the concepts were classified at a fairly low level, e.g. *säng* 'bed' was classified as *furniture*. After the first analysis, it was possible to cluster the words in groups belonging to nodes higher up in the taxonomy, until the top nodes were reached – see the tree structure below. In this way *bed* ended up in the category OBJECT.

säng 'bed'
=> furniture, piece of furniture, article of furniture
=> furnishings
=> instrumentality, instrumentation
=> artifact, artefact
=> object, physical object

Semantic classification is always problematic and schoolbook examples are rare in real corpora. Although many nouns could be categorised in several different ways, in all cases one category was consistently chosen.

8.1.5 Full *and* tom *in SUC*

8.1.5.1 The semantic range of full

Different forms of *full* occur 164 times in the SUC. All were included in the study. *Full* was found to modify nouns from the semantic categories OBJECT, HUMAN, ABSTRACTION, HUMAN ACTION, PSYCHOLOGICAL FEATURE, STATE, MONETARY REPRESENTATION, and PHENOMENON. The nouns modified by *full* were distributed across the semantic categories as follows.

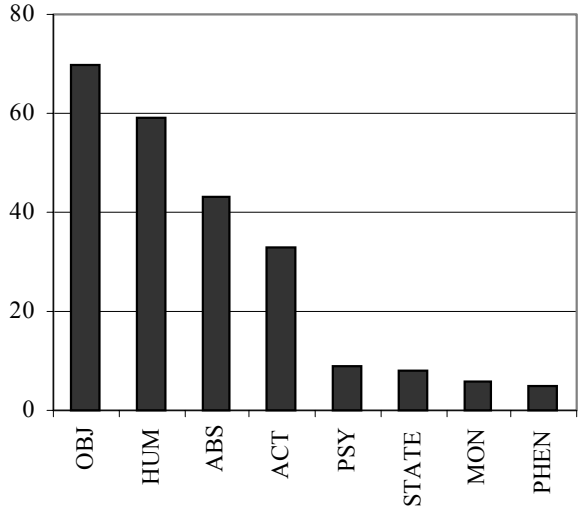


Figure 11: The semantic range of ***full*** *in the SUC. (The bars show absolute numbers in this diagram and the following ones unless otherwise indicated).*

The most common type of noun modified by *full* is OBJECT, e.g. *en korg full med frukt* 'a basket full of fruit' and *diskhon var full med porslin* 'the sink was full of dishes', closely followed by HUMAN.

As a curiosity, it can be mentioned that instances of the category HUMAN are by default full of alcohol. If there is no prepositional phrase indicating otherwise, the human subject is drunk as in the examples below.

Gammelfarfar var ***full*** *och senil när han förfalskade testamentet!*

'Great-grandfather was *drunk* and senile when he forged the will!'

Och prästen [blev] allt ***fullare*** *med varje dopsup.*

'And the priest [got] more *drunk* with each christening dram.'

It is not normal to state explicitly what the content is in the case of intoxication. The constructed sentence below is considered strange (though possible) by native speakers.

?*Han var* ***full*** *av vin och öl.*

'He was *full* of [i.e. 'drunk on'] wine and beer.'

Other than alcohol, humans can be full of abstract phenomena such as enthusiasm, admiration, and questions; see the examples below.

*Björn är **full** av entusiasm över sin H-båt.*

'Björn is *full* of enthusiasm for his H-boat.'

*Jag var **full** av beundran för hans fru...*

'I was *full* of admiration for his wife…'

*Jag är **full** av frågor...*

'I am *full* of questions…'

The third-most common type of noun modified by *full* is ABSTRACTION, e.g. *brev* 'letter' (where the content, not the object, is modified), *fullt ansvar* 'full responsibility', and *full poäng* 'maximum score'.

HUMAN ACTION is also a fairly common type of head noun, e.g. *full sysselsättning* 'full employment' and *en blick full av avund* 'a look full of envy'.

Among the less common types found are: PSYCHOLOGICAL FEATURE, STATE, MONETARY REPRESENTATION, and PHENOMENON; see the examples below.

MONETARY REPRESENTATION:

*I år är det bara deltagare från Stockholms läns [sic] som fått **fullt** landstingsbidrag.*

'This year, only participants from the Stockholm county have received a *full* subsidy from the county council.'

PSYCHOLOGICAL FEATURE:

*Vill man uttrycka den **fulla** betydelsen av ett ord får man ta med fler detaljer än vad som redovisats här.*

'If you want to express the *full* meaning of a word, you have to include more details than given here.'

STATE:

*I stadens styrelse är det **fullt** kaos.*

'In the city council, there is *full* chaos.'

PHENOMENON:

*...— det blåste **full** storm — ...*

'…— there was a *full* gale blowing —…'

8.1.5.2 Containers and rods

Taking a closer look at the group of OBJECTS, it is obvious that the words in this group belong to two different types. Very often they refer to a container that can be filled with something, e.g. *en korg* 'a basket', *en diskho* 'a sink', or a *surface* that can be covered, e.g. *ett bord* 'a table', *en vägg* 'a wall'. Most of the examples of OBJECTS in the material are container*s*: 56 out of 70. The remaining 14 occurrences are *surfaces*.

Surfaces have many features in common with containers. A surface is a fixed area that can be covered with something. An object can be *on* the surface or not, while for a container, the key preposition is *in*. I will treat surfaces as a subgroup of container*s*; thus, surfaces will not be treated separately in this study.

The three meanings listed for *full* in the NEO were all represented in the material studied. The core meaning relates to containers, e.g. *Korgarna är fulla av godsaker* 'The baskets are full of sweets'. The container metaphor is central, and the instances found are characterised by the fact that the container exists whether it is filled or not. The baskets just mentioned exist independently of their contents. Some lexical items, such as *basket*, *bottle*, and *basin*, are fundamentally containers. However, most (if not all) words can function as containers in certain contexts.

The *container* meaning is commonly encountered in an abstract sense, modifying nouns from the categories HUMAN and PSYCHOLOGICAL FEATURE; see the following examples.

*Björn är **full** av entusiasm över sin H-båt.*

Björn is *full* of enthusiasm for his H-boat.

*... den **fulla** betydelsen av ett ord...*

…the *full* meaning of a word…

However, the container metaphor is not applicable when *full* modifies nouns from the categories MONETARY REPRESENTATION, PHENOMENON, HUMAN ACTION, and ABSTRACTION. *Storm* 'gale' in *full storm* 'full gale' cannot be seen as a container, and nor can the insurance term *efterlevandeskydd* 'survivor protection' in *fullt efterlevandeskydd* 'full survivor protection'. In such cases, the phenomenon starts to exist when the content starts to exist and there is no abstract container involved. I suggest a kind of "measuring rod" to account for this meaning; see Figure 12. I will call this meaning *the rod meaning*. It turns out to be the most frequent meaning of *full* in the SUC.

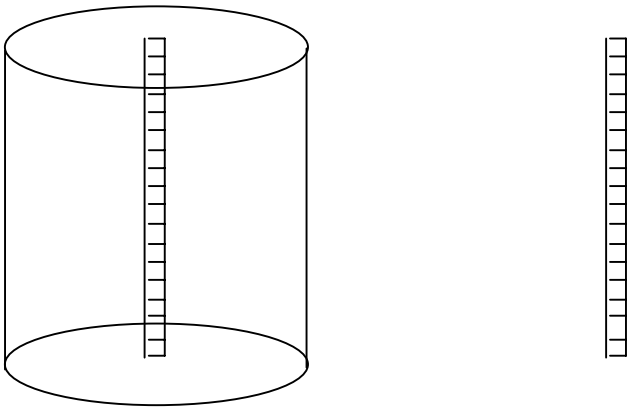


Figure 12: The container metaphor and the measuring rod

Concepts such as *full fart* 'full speed', *full poäng* 'maximum score', and *full sysselsättning* 'full employment' simply cannot be empty – then they would not exist. Therefore it is not possible to modify concepts from these categories with *tom*: **tom fart* 'empty speed', **tom poäng* 'empty score' and **tom sysselsättning* 'empty employment' – there is no container available. Instead *ingen* 'no' or *utan* 'without' is used to express such concepts, e.g. *ingen fart* 'no speed', *utan poäng* 'without score', *ingen sysselsättning* 'no employment'. Another way to express the absence of such concepts is to use the suffixes *-lös* '-less' and *-fri* '-free'. Nouns which can be modified by *full* but not by *tom* can usually form adjectives using *-lös*, e.g. *ansvarslös* 'irresponsible', *sysselsättningslös* 'without occupation', *poänglös* 'pointless', or *-fri*, e.g. *molnfri* 'cloudless', *benfri* 'boneless', *riskfri* 'safe' (Sigurd 1972; cf. the description of the use of the English suffixes *-ful* and *-less* by Holmqvist & P□uciennik 1996).

The third meaning of *full* is *intoxication*, which can be viewed as a special case of the *container* meaning with both the content and the container given, i.e. the content is alcohol and the container is a human body.

The meanings are distributed as follows in the SUC:

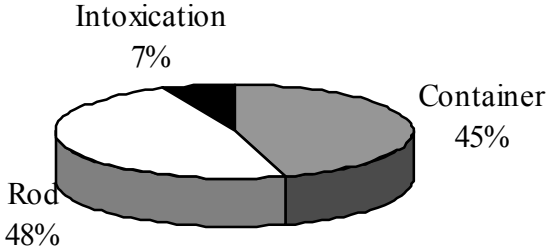


*Figure 13: The distribution of the meanings of **full** in the SUC.*

All instances of *full* in the *intoxication* meaning modify HUMAN nouns. The majority of the nouns modified by *full* in the *container* meaning belong to the category OBJECT, i.e. containers such as *säck* 'sack', *ficka* 'pocket', *grop* 'pit',

hårborste 'hair brush'. There are a number of instances of HUMAN and ABSTRACTION, and one example of a PSYCHOLOGICAL FEATURE modified by *full* in the *container* sense: *den fulla betydelsen av ett ord* 'the full meaning of a word'. The distribution of the semantic categories modified by *full* in the *container* sense is clear from the following diagram.

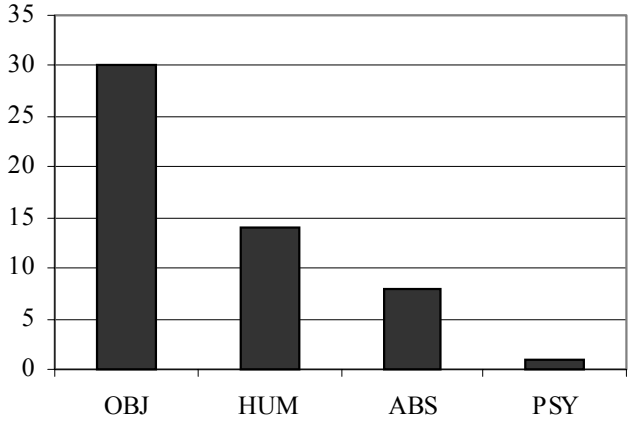


*Figure 14: Distribution in the SUC of the semantic categories modified by **full** in the container sense.*

The words modified by *full* in the *rod* sense most commonly belong to the categories of HUMAN ACTION and ABSTRACTION, e.g. *full produktion* 'maximum production' and *full poäng* 'maximum score'. There are also some examples of PSYCHOLOGICAL FEATURE: *full förståelse* 'full understanding', *fullt allvar* 'full seriousness'; PHENOMENON: *full orkan* 'full hurricane'; STATE: *i full blom* 'in full bloom'; and MONETARY REPRESENTATION: *fullt landstingsbidrag* 'full subsidy from the county council'. The diagram below visualises the distribution of semantic categories modified by *full* in the *rod* sense.

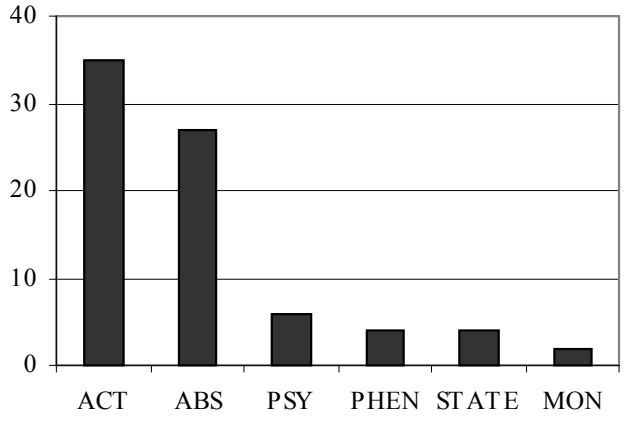


*Figure 15: Distribution in the SUC of the semantic categories modified by **full** in the rod sense.*

Two categories are modified by *full* in both the *container* sense and the *rod* sense: ABSTRACTIONS and PSYCHOLOGICAL FEATURES.

8.1.5.3 The semantic range of tom

There are 64 sentences with different forms of the adjective *tom* in the SUC. The most common type of head noun modified by *tom* is OBJECT, e.g.

***Tomma** tunnor skramlar ju mest!*

'As we all know, *empty* vessels make the greatest noise!'

Just as found in the examples with *full*, the OBJECTS modified include instances of surfaces, treated very much like containers, e.g. *den tomma parkeringsplanen* 'the empty parking lot', *bordet var tomt* 'the table was empty'. 48 of the instances of OBJECT are containers and 11 surfaces. There are also some occurrences of ABSTRACTION, PSYCHOLOGICAL FEATURE, and STATE; see below.

ABSTRACTION:

*Allihop har en **tom** och lite sorgsen blick...*

'They all have *empty* and somewhat sad eyes…'

*...en sådan **tom** uppvisning...*

'…such an *empty* performance…'

PSYCHOLOGICAL FEATURE:

*Descartes självmedvetande är visserligen omöjligt att sätta ifråga, men **tomt** på innehåll.*

'Descartes's self-consciousness is indeed impossible to question, but *empty* of content.'

*Denne läste med en **tom** känsla i bröstet sin fars sista ord.*

'He read, with an *empty* feeling in his chest, his father's last words.'

STATE:

*... varandet är **tomt** ...*

'…existence is *empty*…'

The semantic categories of the nouns modified by *tom* were distributed as follows.

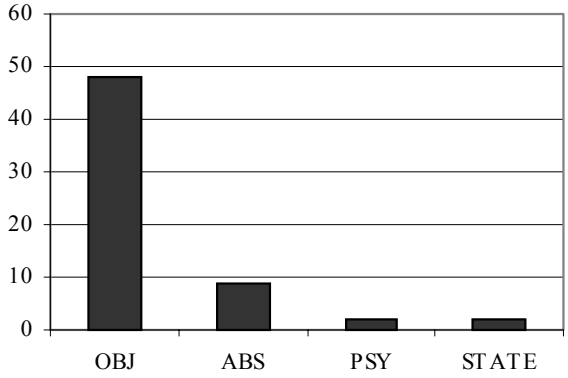


Figure 16: The semantic range of ***tom*** *in the SUC.*

OBJECT dominates the semantic range of *tom*. Of the OBJECTS, most are containers and about 20% are surfaces. In the material studied, *tom* was found to modify only a couple of examples of ABSTRACTION, PSYCHOLOGICAL FEATURE, and STATE.

The container metaphor is central to the meaning of *tom*. Even the examples of surfaces found can be viewed as containers: *scenen* 'the stage', *parkeringsplatsen* 'the parking lot'. They share the critical feature of having existence independently of their content. Thus the measuring-rod metaphor is not applicable to any of the cases of *tom* found in the SUC.

8.1.6 Full *and* tom *in Parole*

To test if the categories used above are sufficient to describe the semantic ranges of *full* and *tom*, another 500 random occurrences each of *full*+NOUN and *tom*+NOUN from the Parole corpus were classified in the previously created database. The results for each word will be presented before the semantic ranges of the two words are compared.

8.1.6.1 The semantic range of full

A search string that included the tag for adjectives was used to extract all the examples of *full* from the Parole corpus. However, nearly 200 of the instances found were actually not adjectival uses of *full*, but adverbs that had been incorrectly tagged as adjectives. The remaining 304 examples were categorised and analysed as in the previous study.

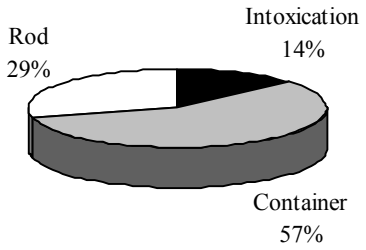


*Figure 17: The distribution of the meanings of **full** in Parole.*

While the *container* and *rod* meanings had about equal shares in the SUC, the *container* meaning is the most common one in Parole, with 57% of the *examples*; see Figure 17. The share for the *intoxication* meaning is higher too: 14% compared with 7% in the SUC. The instances found are distributed across semantic categories as shown in the diagram below.

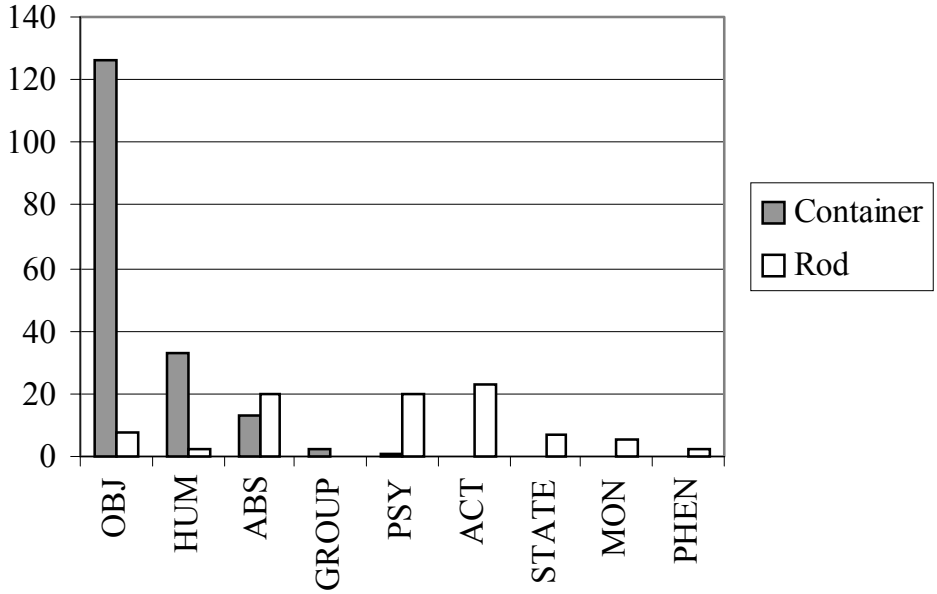


*Figure 18: The semantic range of **full** in Parole.*

The distribution of the semantic categories modified by *full* in Parole showed the same trend as in the SUC. The most common categories were the same: nouns from the categories OBJECT and HUMAN were most frequently modified by *full* in the *container* sense, while ABSTRACTION, PSYCHOLOGICAL FEATURE, and HUMAN ACTION were the ones most often modified by *full* in the *rod* sense. There were also some examples of OBJECTS, HUMANS, STATES, MONETARY REPRESENTATIONS, and PHENOMENA modified by *full* in the *rod* sense.

A few new categories of words were found to be modified by *full* in its different senses. Not surprisingly, nouns from the semantic category GROUP were found to be modified by *full* in the *container* sense, e.g. *fullt sortiment av insaltad ingefära* 'a full range of pickled ginger'. Also, a few occurrences of

substantivised adjectives were found, e.g. *den fullt påklädde* 'the fully dressed (man)'.

Unexpectedly, there was yet another group of nouns modified by *full* in the *rod* sense, namely a certain type of OBJECT. These nouns all share one feature: they are in fact a collection of things that together form a unit; see the examples below.

***full** uniform* 'full uniform'

***full** skrud* 'full garb'

***full** rustning* 'full armour'

***full** stridsutrustning* 'full combat gear'

***full** ornat* 'full canonicals'

***fullt** slagverk* 'full percussion set'

***full** pluton* 'full platoon'

***full** make up* 'full make-up'

The semantic ranges of the different meanings of *full* are summarised in the following diagram.

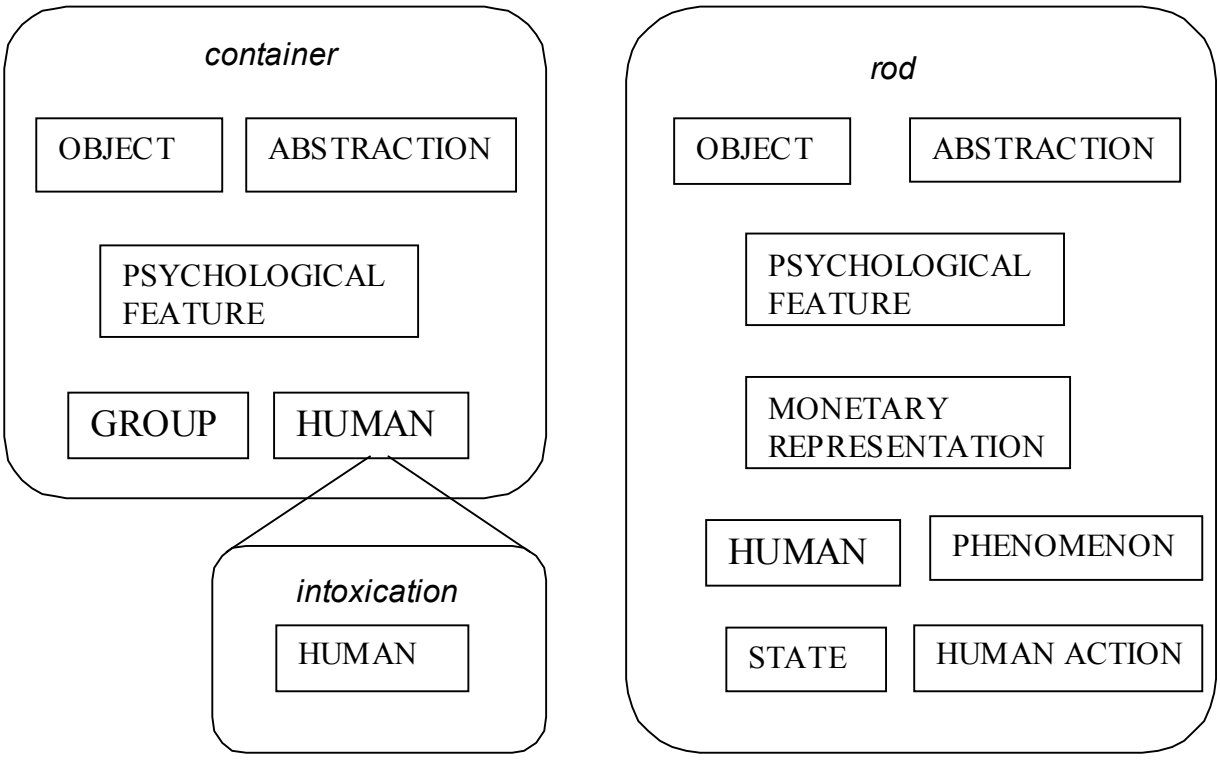


*Figure 19: The semantic range of **full** in Parole.*

The figure shows the two basic meanings of *full*: the *container* sense and the *rod* sense. The semantic range of *full* in the *container* sense covers OBJECT,

ABSTRACTION, PSYCHOLOGICAL FEATURE, GROUP, and HUMAN. The *intoxication* sense is a special case of the *container* sense when *full* modifies a HUMAN.

In the *rod* sense, the semantic range covers OBJECT, ABSTRACTION, PSYCHOLOGICAL FEATURE, MONETARY REPRESENTATION, HUMAN, PHENOMENON, STATE, and HUMAN ACTION. The semantic range of the *rod* sense is wider than that of the *container* sense. MONETARY REPRESENTATION, PHENOMENON, STATE, and HUMAN ACTION are semantic categories found to be modified by *full* in the *rod* sense but not in the *container* sense.

8.1.6.2 The semantic range of tom

Ambiguity does not yield the same problems for *tom* as it did for *full*, where two-fifths of the occurrences that were randomly extracted from Parole were adverbs rather than adjectives. The material studied for *tom* in the Parole corpus consists of 537 occurrences.

Tom was found to modify a few more categories in Parole than in the SUC. Apart from examples of OBJECTs (containers and surfaces), ABSTRACTIONs, and PSYCHOLOGICAL FEATUREs, a few examples of HUMAN, STATE, and HUMAN ACTION modified by *tom* were found; see the examples below.

HUMAN:

...en ***tom*** *människa...*

'an *empty* person'

Jag är ***tom***

'I am *empty*'

STATE:

ett vakuum ***tommare*** *än den* ***tomma*** *rymden*

'a vacuum *emptier* than *empty* space'

HUMAN ACTION:

tomma *utbildningsplatser*

'*empty* places in educational programmes'

One of the most difficult nouns to classify was the very uncommon word *staffage* 'staffage = accessory figures (in a landscape painting), (stage) décor', which occurs in the sentence *Dessa gengångare är dock ett så slentrianmässigt och* ***tomt*** *staffage i romanen...* 'These ghosts, however, are such hackneyed and

empty staffage in the novel…'. It has been coded as GROUP, forming a category with only one member.

The semantic categories were distributed as visualised in the diagram below.

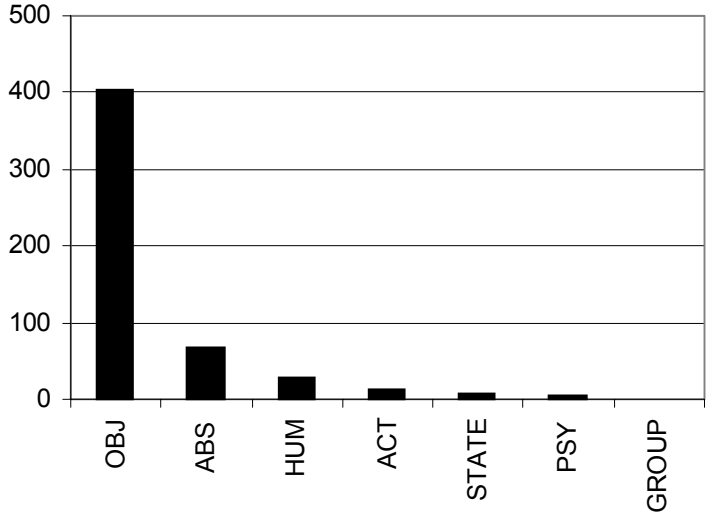


Figure 20: The semantic range of tom in Parole

The graph shows that the majority of the nouns modified by *tom* in Parole belong to the semantic category OBJECT. The modification of other categories is marginal; still, all the semantic categories represented in the diagram are part of the semantic range of *tom*, which is visualised below.

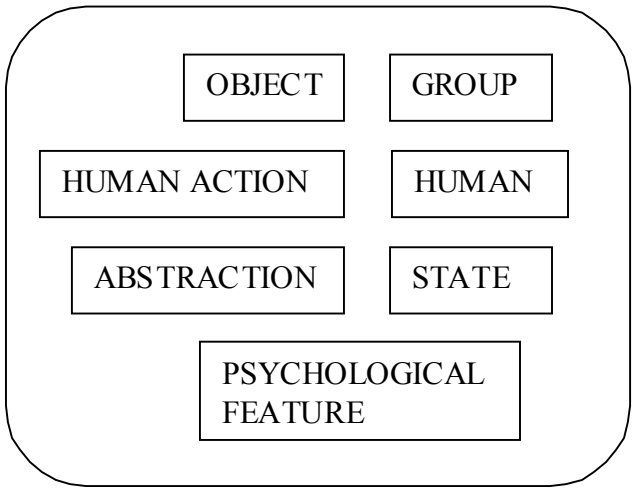


Figure 21: The semantic range of ***tom*** *in Parole.*

8.1.7 A comparison of the semantic ranges of **full** *and* **tom.**

If the semantic ranges of the adjectives *full* and *tom* are merged in the same diagram (see Figure 22) we can see that the semantic range of *tom* partially overlaps the semantic range of *full*. The overlap covers part of the *container* meaning of *full*: both *full* and *tom* were found to modify nouns belonging to the

categories OBJECT, ABSTRACTION, PSYCHOLOGICAL FEATURE, GROUP, and HUMAN (in the *container* sense).

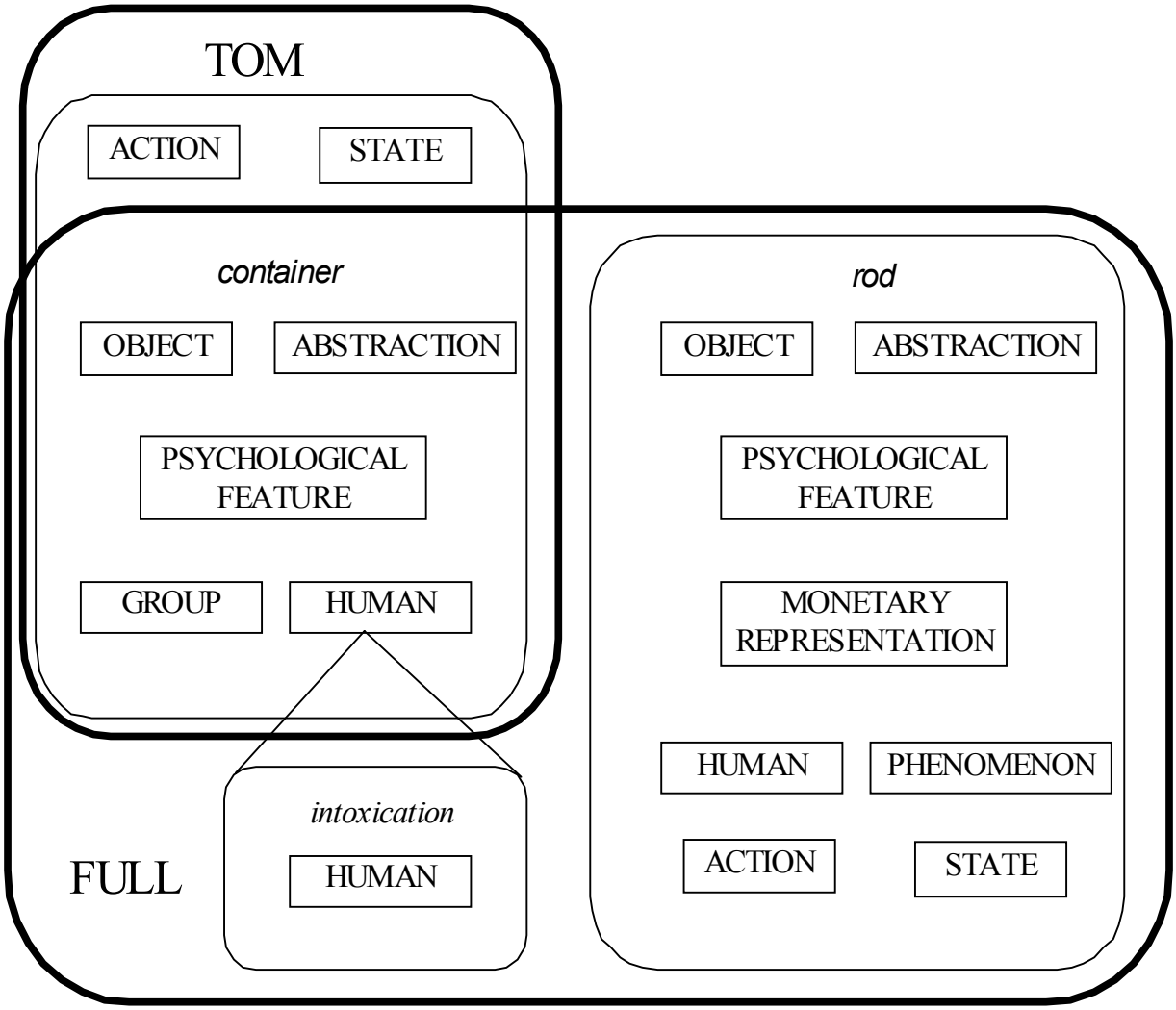


*Figure 22: Comparison of the semantic ranges of **full** and **tom**.*

Only *full* is used in the meaning *intoxication*, and only with HUMAN head nouns. I view intoxication as a special case of *full* modifying a HUMAN; see Figure 22.

There are few examples of *tom* modifying ACTIONs and STATEs. Even though no examples of *full* modifying these types of nouns were found in the study, such constructions are not impossible. The head noun *vakuum* 'vacuum' in the example *ett vakuum tommare än tomma rymden* 'a vacuum emptier than empty space' may be, as it were, empty by definition, but one counterexample is *den tomma evigheten* 'empty eternity' – it is not hard to imagine eternity being full of objects, events, feelings, etc.

It is striking that only *full* is used in the *rod* sense. The container metaphor is central to the meanings of *tom*, while *full* can be used when there is no container

involved but rather a type of measuring rod. However, the semantic range of *tom* is quite wide: the only two semantic categories that *full* can modify but *tom* cannot are MONETARY REPRESENTATION and PHENOMENON.

This study covers the semantic ranges of only two words, but both *full* and *tom* have other antonyms. Other antonymy relations involved are *full—nykter* 'drunk—sober', *full—inget* 'full—nothing', and *tom—fylld* 'empty—filled'. *Nykter* is the opposite of *full* in the *intoxication* sense, while in the *rod* sense of *full*, the other end of the scale is indicated by *inget*. That part of the range of *tom* that does not overlap with the semantic range of *full* – where *tom* modifies ACTIONS and STATES – probably overlaps with the semantic range of *fylld* 'filled' instead.

To get a clear view of the entire picture, it would be necessary to make a thorough investigation of the semantic ranges of the other antonyms, i.e. *fylld*, *nykter*, and *inget*.

The overlap in semantic ranges indicates that *full—tom* are antonyms in the *container* sense, but not in the *rod* sense. The deviating co-occurrence pattern previously observed for *full—tom* is probably attributable to the fact that there are several antonym relations involved in the field.

8.2 A case study: The semantic ranges of *stor* and its synonyms

The semantic ranges of *stor* and 28 of its synonyms will be described in the following section. It will be shown in Section 8.2.3 that the synonyms of *stor* can be classified in four groups according to the semantic categories that they frequently modify in the pilot study, which is based on data from the SUC: OBJECT, ABSTRACTION, HUMAN, and HUMAN ACTION.

A small adjustment of the semantic categories will be made: HUMAN ACTIONS, EVENTS, and STATES will be merged into SITUATIONS in Section 8.2.4. This merger gives a better view of the adjectives that frequently modify HUMAN ACTIONS, EVENTS, and STATES – the latter of these groups not being frequent enough on thier own.

The next study, presented in Section 8.2.5, is based on data from Parole and the words will be classified in four groups according to what semantic category of nouns they most frequently modify: OBJECTS, ABSTRACTIONS, HUMANS, and SITUATIONS, respectively. Within each group, the words will be further classified according to the other semantic categories that they frequently modify. The result will be eleven synonym sets in four main groups.

In the next Section, 8.2.6, we will take a closer look at the various meanings of *stor*. Four meanings of *stor* will be distinguished: *concrete dimension*, *countable quantity*, *uncountable quantity*, and *importance*. The semantic ranges for the four different meanings of *stor* will be described, and this will be used to group each of the synonym sets distinguished in 8.2.5 with a meaning of *stor*.

It was found that the most frequent meaning of *stor* used in the Parole corpus is *uncountable quantity*, not the core meaning *concrete dimension*.

The core meaing of a word is generally the most frequent meaning of a word. It is also the meaning first acquired by language learners, and based on the results from Section 8.2.6, it was questioned whether *concrete dimension* is in fact the core meaning of *stor*. A study of children's use of *stor* will be presented in 8.2.7, showing that children use exclusively the *concrete-dimension* meaning of *stor*. This indicates that *concrete dimension* is indeed the core meaning of *stor*.

A model for how the meanings of *stor* are metaphorically extended from the core meaning to those of *countable quantity*, *uncountable quantity*, and *importance* will be sketched in Section 8.2.8.

In Section 8.3, the results from the study of the semantic ranges of the synonyms of *stor* in the Parole corpus presented in Section 8.2.5.4 will be matched with the semantic ranges of the various meanings of *stor* presented in Section 8.2.6. Based on the empirical data, a grouping of four separate entries of *stor* in SWordNet will be suggested and the synonyms of *stor* will be linked to the various meanings of *stor* from eleven separate synsets. The structure will be discussed and improved. The entries for the adjectives will be connected to the noun lexicon; this is a way to code semantic range in SWordNet.

8.2.1 Introduction

The words in WordNet are defined by their places in the net. The meaning or meanings of a word are reflected in its coded relations to other words, e.g. its synonyms, antonyms, and hypernyms. Furthermore, the lexicon user gets a short lexical definition and a few usage examples. The various word classes form separate nets with very few connections between them. Some morphologically derived or otherwise similarly connected words (typically morphologically derived in another language and then borrowed) are linked; the adjectives of this type are called "relational" and are actually not stored in a net. They are listed in a file and from each adjective there is a pointer to the noun it pertains to, e.g. *musical* is connected with *music*, *criminal* with *crime*, and *dental* with *tooth*. Apart from morphological derivations, there are no links connecting words belonging to different word classes.

A search for the word *bombastic* in Princeton WordNet 1.6 yields the following result:

> The adjective "bombastic" has 1 sense in WordNet.
>
> 1. bombastic, declamatory, large, orotund, tumid, turgid -- (ostentatiously lofty in style; "a man given to large talk"; "tumid political prose")

There are semantic links to other words from *bombastic*, e.g. to its indirect antonyms, and to a familiarity count, which indicates how commonly used the word is. This is not much of a clue for the lexicon user, nor for a text-generation system making use of the electronic lexicon. For adjectives it would be helpful to have information on what types of nouns a given adjective can modify – i.e. the semantic range of the adjective.

Direct antonyms have the same semantic range, cf. Muehleisen (1997) and the study of *full* and *tom* in Section 8.1. The words forming an antonymous pair are similar in all respects but one, in which they are maximally opposed. I believe that similarity of semantic range is a feature valid for synonyms as well. The objective of this study is to describe the semantic ranges of words from the field of SIZE and to use these descriptions to classify the words in synonym groups of the kind used in the Princeton WordNet and SWordNet. Because of the large number of terms in this field, however, I have limited my study to cover terms from only one end of the scale, namely synonyms of *stor* 'large'. It would be very interesting to compare the semantic ranges of the words from either end of the scale, but that will be saved for future research.

The following study describes and compares the semantic ranges of *stor* and 28 of its synonyms. The information found has been used to organise those adjectives in SWordNet.

Several studies of collocation within the field of SIZE have been made for English, e.g. Muehleisen (1997) and Biber et al. (1998). Biber et al. present a small investigation of the synonyms *big*, *large*, and *great* to shed light on what collocation studies can be used for, and Muehleisen uses the collocation patterns of *big*, *large*, *small*, and *little* to determine antonymy relations between those words. Words describing size are highly frequent; the field is rich in synonyms, highly complex, and expected to be a fruitful object of study for Swedish as well.

Stor is the most frequent adjective in the SUC. In the frequency list, it is followed by *annan* 'other', *ny* 'new', and then another word expressing size: *liten* 'small', the antonym of *stor*, cf. Table 19. In accordance with *Svenska Akademiens grammatik* (Teleman et al. 1999: 5: §196–239), I would prefer to classify several of the top-ten adjectives in Table 19 as pronouns rather than adjectives, namely *annan*, *själv*, and *sådan*, but Table 19 reflects their classification in the SUC.

stor 'large'	2793
annan 'other'	2657
ny 'new'	1756
liten 'small'	1394
själv 'self'	1301
olik 'unlike'	1273
god 'good'	1167
gammal 'old'	1086
hel 'whole'	1082
sådan 'such'	1016

Table 19: The ten most common adjectives in the SUC and their number of occurrences.

There are many terms designating size in Swedish; the Strömberg thesaurus lists 43 synonyms of *stor* and 32 synonyms of *liten*. Apart from *stor—liten*, several adjectives mentioned as synonyms themselves mark the end points of scales, e.g. *bred—smal*, *lång—kort*, *stark—svag*. They all give information relating to size, but they are restricted to certain dimensions, e.g. *bred—smal* for width and *lång—kort* for length. Unlike in English, there is only one general antonym pair in Swedish among the 75 words designating size, namely *stor—liten* 'large—small'. In English we find both *big—little* and *large—small*, which indicates that it takes two separate scales to cover the field of SIZE in English. There seems to exist only one scale of SIZE in Swedish; however, there are considerable differences in semantic range between the various concepts designated by the synonyms of *stor*. The following study is an attempt to describe the semantic ranges of those synonyms and to group them according to their semantic ranges.

8.2.2 Test set: **stor** *and its synonyms*

The choice of words for the test set was based on the synonyms of *stor* listed in the Strömberg thesaurus. All words found there that are adjectives and occur at least 50 times in the Parole corpus were included in the study. Strömberg lists quite a few participles,[24] which were excluded, e.g. *betydande* 'considerable' and *omfattande* 'extensive'. There is probably no semantic difference between participles and "true adjectives", but the former were excluded so as to limit the test set. Furthermore, some infrequent adjectives such as *himmelsvid* 'huge' and *muskulös* 'muscular', of which not enough examples were found for meaningful studies of them to be possible, were left out. This leaves 29 words, including *stor*, to study:

[24] The definition of participles follows the convention in the SUC.

ansenlig 'considerable'	*mäktig* 'powerful, immense'
avsevärd 'considerable'	*märklig* 'strange, remarkable'
bred 'wide'	*oerhörd* 'tremendous'
diger 'voluminous'	*ofantlig* 'enormous'
dryg 'ample'	*otrolig* 'incredible'
enorm 'enormous'	*riklig* 'plentiful'
framstående 'prominent'	*ryktbar* 'famous'
förnäm 'distinguished'	*rymlig* 'spacious'
grov 'heavy'	*stark* 'strong'
inflytelserik 'influential'	*stor* 'large'
innehållsrik 'containing a great deal'	*storväxt* 'tall'
kolossal 'colossal'	*vidsträckt* 'extensive, vast'
kraftig 'strong'	*väldig* 'huge'
lång 'long'	*ädel* 'noble, magnanimous'
mycken 'much'	

*Table 20: Test set: synonyms of **stor** (according to Strömberg) included in the pilot study.*

The following words were excluded from the study:

berömd 'famous'	*omfångsrik* 'extensive'
betydande 'considerable'	*omåttlig* 'tremendous'
betydlig 'considerable'	*storsinnad* 'magnanimous'
fullvuxen 'full-grown'	*storsint* 'magnanimous'
himmelsvid 'huge'	*utmärkt* 'excellent'
högrest 'tall'	*utsträckt* 'extended'
muskulös 'muscular'	*voluminös* 'voluminous'
omfattande 'extensive'	

*Table 21: Synonyms of **stor** (according to Strömberg) not included in the pilot study.*

8.2.3 Pilot study: Semantic ranges of stor *and its synonyms in the SUC*

A pilot study was carried out to describe the semantic ranges of *stor* and its synonyms based on data from the SUC. The SUC is a manually – and thus very accurately – tagged corpus. Apart from word-class information, each word in the corpus is tagged as to lemma. This makes it easy to find all instances of the words in the corpus, even those characterised by irregularities such as a change of stem vowel, e.g. *lång—längre—längst* 'long—longer—longest', or a complete change of stem, e.g. *liten—mindre—minst* 'small—smaller—smallest'.

Each sentence containing a word from a lemma belonging to the test set was extracted from the corpus. The head of each modifier was then identified using the program *Para*, which created a list of adjective-noun combinations. The nouns were classified semantically according to the Princeton WordNet's unique beginners for nouns, using the program *Klassa*. Since SWordNet was not yet big enough to be used, the semantic categorisation of the English equivalent of the Swedish noun was used when tagging.

The classification results were analysed in terms of frequencies; see Table 22.

Adjective lemma	***Semantic category of the modified noun, number of occurrences***											
	ABS	ACT	ANI	EVENT	GROUP	HUM	OBJ	PHEN	MON	PSY	STATE	***Total***
stor[25]	161	88	5		35	24	118	6	10	54	7	508
lång	134	25			1	11	89	1	2	5	4	272
stark	63	25	2	1	23	25	30	18		44	7	238
mycken	47	43	3	4	5	17	55	3	4	20	5	206
kraftig	16	26	2			5	38	6		5	4	102
bred	15	14	1		10	1	36			5	2	84
grov	6	16					37			2		61
väldig	9	6	2		1	6	31		1		1	57
enorm	11	7					10	1	2	4	5	40
märklig	9	5	2	1	2	1	14	2		2	1	39
mäktig	5	4	1		5	5	7					27
oerhörd	12	4				1	1			3	4	25
dryg	11				1	2	3		1			18
ädel	3	1	1			3	9					17
riklig	5	3					3	1	2		1	15
rymlig						1	13					14
förnäm	3					4	2			3		12
avsevärd	4	2					1		1		2	10
otrolig	3					1	1		1	1		7
storväxt						7						7
framstående	1	1			1	3						6
inflytelserik					1	5						6
vidsträckt		1					2			2		5
ansenlig	4									1		5
ofantlig	1	3					1					5
diger	4											4
innehållsrik	1						3					4
ryktbar	1					3						4
kolossal	1											1

*Table 22: **Stor**, synonyms of **stor** and the frequency of the semantic categories they modify in the SUC.*

The semantic categories most frequently modified by the adjectives in the test set are ABSTRACTION, HUMAN ACTION, HUMAN, and OBJECT.

It is clear that for most of the adjectives in the test set, there is one semantic category to which the majority of the words they modify belong. For example, *lång* typically modifies an ABSTRACTION such as a time period or a distance, e.g.

[25] There are 2,793 occurrences of forms of *stor* in the SUC. This was considered to be too large a material for the pilot study. Thus, of the sentences containing any form of the word *stor*, slightly over 500 sentences were randomly chosen using a Perl program, *Slump*; this subset was included in the pilot study.

en lång sträcka 'a long distance', *Hur lång tid tar det?* 'How long time will it take?' The adjective *stark* also seems to prefer to modify an ABSTRACTION, while *kraftig*, *grov*, and *väldig* prototypically modify an OBJECT.

The results indicate that while *stor* modifies a wide range of semantic categories, the majority of its synonyms are much more restricted in what types of nouns they modify. The synonyms cluster according to the semantic category most frequently modified. Four groups can be distinguished: adjectives that most frequently modify OBJECTs, ABSTRACTIONs, HUMANs, and HUMAN ACTIONs, respectively; see Table 23. The words in each group share semantic features. This seems to be a fruitful way to group the words according to meaning, and it is worth looking at a larger material using this approach. EVENTs and STATEs are low-frequent in the study presented, cf. Table 22. It will be shown in the next Section, 8.2.4, that they can be merged with HUMAN ACTIONs.

Adjectives most frequently modifying			
OBJECT	ABSTRACTION	HUMAN	HUMAN ACTION
bred	*avsevärd*	*förnäm*	*ofantlig*
grov	*ansenlig*	*framstående*	
innehållsrik	*diger*	*inflytelserik*	
kraftig	*dryg*	*ryktbar*	
mycken	*enorm*	*storväxt*	
mäktig	*kolossal*		
märklig	*lång*		
rymlig	*oerhörd*		
vidsträckt	*otrolig*		
väldig	*riklig*		
ädel	*stor*		
	stark		

*Table 23: **Stor** and synonyms of **stor** grouped according to the semantic category most frequently modified.*

8.2.4 SITUATIONS

In this section it will be shown that some of the semantic classes used in the previous studies can be merged, namely HUMAN ACTION, STATE, and EVENT. STATEs and EVENTs were low frequent in the study based on material from SUC. By analogy with Mourelatos's (1978) fusion of Vendler's and Kenny's schemes of verb types (see also Verkuyl 1993), *states*, *processes,* and *events* can be merged under the hypernym SITUATION, cf. Figure 23. Developments and punctual occurrences are types of events.

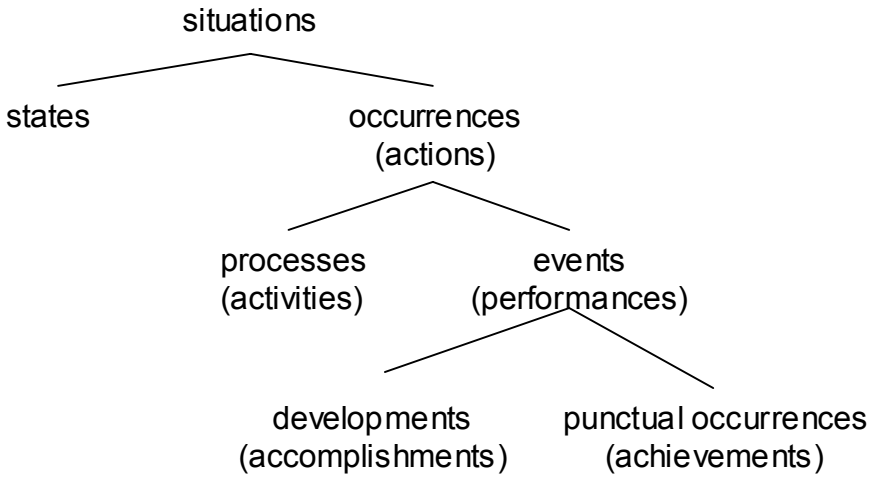


Figure 23: The Vendler-Kenny scheme of verb types (Mourelatos 1978).

The Princeton WordNet distinguishes between *activities*, *events*, and *states*. Mourelatos's *processes* correspond to *activities* in WordNet. *Activities* and *events* share the hypernym *occurrences*. *States* and *occurrences* are, in turn, hyponyms of *SITUATIONS*, cf. Figure 24.

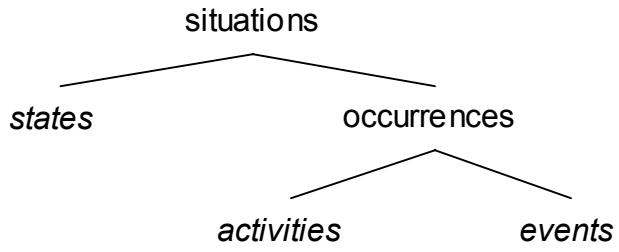


Figure 24: Taxonomy of SITUATIONS

Activities, *events*, and *states* will be presented as SITUATIONS in the following studies.

8.2.5 Semantic ranges of stor *and its synonyms in Parole*

The word-class tags of the SUC are of a high quality, but the corpus is quite small. In this study, the semantic ranges of *stor* and its synonyms will be described based on examples from a larger corpus, Parole.

8.2.5.1 Changes to the test set

During the study of the synonyms of *stor* in the SUC, it was observed that some adjectives in the test set were seldom used synonymously with *stor*. There are also some words in the test set that mark the end points of other scales than *stor—liten* and therefore deserve large studies of their own. Five words were excluded from the test set: *mycken*, *märklig*, *lång*, *bred*, and *stark*.

Mycken ‘much’ is highly polysemous, and in the sense of *stor* it seems to have to do with quantification. The adverbs *mer* ‘more’ and *mest* ‘most’, used to express comparison, belong to the lemma *mycken*, and the tagger used on the Parole corpus has had great difficulty in distinguishing between adjectives and adverbs. This yields some methodological problems, and because of the highly diverse usage of the word a more extensive study would be necessary.

Märklig ‘strange, remarkable’ does not appear to function as a synonym of *stor* in any of the examples found in the SUC. Though it is possible to use the word synonymously with *stor*, this use is obviously uncommon – presumably somewhat dated – and *märklig* will also be excluded from further study.

Several of the highly frequent adjectives in the test set denote the end point of scales other than *stor—liten* ‘large—small’: *lång*, *bred*, and *stark*. *Lång* ‘long’ and *bred* ‘wide’ are dimensional adjectives, but the dimension they relate to is included in the meaning of the word. *Lång* modifies the length of an OBJECT, e.g. *en lång bana* ‘a long track’, or, in a transferred sense, the length of something more abstract, e.g. *lång tid* ‘long time’. *Bred* modifies the width of something, e.g. *en bred väg* ‘a wide road’*, ett brett kunskapsområde* ‘a wide area of knowledge’. *Stark* ‘strong’ is usually the antonym of *svag* ‘weak’ and is synonymous with *stor* only in an abstract sense.

These three words are all highly frequent, highly polysemous, and only marginally synonymous with *stor*. One may also wonder why other dimensional words sharing the feature of specifying what dimension is modified are not listed by Strömberg, e.g. *tjock* ‘thick’. However, the polysemy of these words causes problems in the analysis of their semantic ranges; since they have several meanings, it is hard to generalise about their semantic ranges. Instead, it would be necessary to investigate the semantic range for each of the senses individually. Since there is no room for such studies in this thesis, *lång*, *bred*, and *stark* will be excluded from the test set.

The test set has thus been reduced to the 24 adjectives listed in Table 24.

ansenlig 'considerable'	*mäktig* 'powerful, immense'
avsevärd 'considerable'	*oerhörd* 'tremendous'
diger 'voluminous'	*ofantlig* 'enormous'
dryg 'ample'	*otrolig* 'incredible'
enorm 'enormous'	*riklig* 'plentiful'
framstående 'prominent'	*ryktbar* 'famous'
förnäm 'distinguished'	*rymlig* 'spacious'
grov 'heavy'	*stor* 'large'
inflytelserik 'influential'	*storväxt* 'tall'
innehållsrik 'containing a great deal'	*vidsträckt* 'extensive, vast'
kolossal 'colossal'	*väldig* 'huge'
kraftig 'strong'	*ädel* 'noble, magnanimous'

*Table 24: Test set: **stor** and synonyms of **stor** (according to Strömberg) included in the study.*

8.2.5.2 Experiment

It is possible to use word-class tags in the regular expressions when searching the Parole corpus, even though these tags are not displayed in the concordances. However, Parole is not lemmatised, so the search expression must allow for various endings. Alternatively, several searches have to be made for each word form to extract concordances covering all word forms. Where the total number of occurrences of a word in its various forms exceeded 1,000, a subset of the sentences was randomly chosen for the study, cf. Table 25. In these cases, the different word forms were weighted in relation to their frequency to make sure that the subset constituted a balanced sample.

Concordances for each of the words in the test set were obtained from the Parole corpus. It is not necessarily the case that the head of an adjective is found in the same sentence as the adjective itself. It can be in the sentence before or even earlier, and sometimes it is not explicitly mentioned at all. The study was confined to sentences where the adjective and the noun occurred in the same sentence. A regular expression was used to find complete sentences containing a form of the adjectives in the test set.[26]

The results of the Parole searches were saved in text files like the following one:

[26] [msd="FE"] [msd!="FE"]* [word="adjective.*" & msd="AQ.*"] [msd!="FE"]* [msd="FE"] (Using this regular expression, every line will start with a major field delimitor (.?!); this has to be taken into account when processing the sentence automatically.)

```
Swedish PAROLE (LE2-4017)

Search for [msd="FE"] [msd!="FE"]* [word="storväxt" & msd="AQ.*"]
[msd!="FE"]* [msd="FE"] in PAROLE

19 hits (1000 maximum)

 . " Hagaskräcken " - storväxt man från speciell stadsdel .
 . " Männen skrattade hånfullt och en storväxt karl med grått hår och
 bräckt näsa tog några steg fram och svarade :
 . Bredvid honom satt en storväxt gråhårig man på huk och vred sina händer
 --- Kavass väpnaren .
 . Den ene hade ansiktet nästan helt gömt i kragen på en stor pälsrock och
 den andre, en ovanligt storväxt karl, viskade något till honom medan
 han såg på mig .
 . Det är ett storväxt ryskt lag med fem man som mäter två meter eller
 längre i strumplästen och lika många som får vågen att pendla runt hundra
 kilo .
...
```

Figure 25: The result of a search on the word ***storväxt*** *'tall' in Parole.*

The concordances were pre-processed to delete information about the search at the beginning of each file as well as to strip off the major field delimitor starting each line.

The remaining sentences were input to the program *Para*. Since the corpus was tagged automatically, the accuracy of the tags is poor compared to the SUC. The program *Para* was therefore modified to use a simpler algorithm for suggesting head nouns, namely suggesting the word immediately next to the adjective as its head noun – a primitive method, but since all the sentences had to be read anyway, accuracy did not suffer. The sentences were classified using *Klassa*. The classified sentences were imported to a FileMaker Pro database, which was used for proof-reading the semantic tags and for further classification.

8.2.5.3 *The semantic range of* stor

The distribution of the semantic range of *stor* in Parole corresponds well with what was found for *stor* in the pilot study, cf. Figure 26. ABSTRACTIONS, OBJECTS, and SITUATIONS account for about 70% of the semantic range of *stor*. PSYCHOLOGICAL FEATURES and GROUPS cover about 10% each, while HUMANS, MONETARY REPRESENTATIONS, PHENOMENA, and ANIMALS constitute only minor portions of the semantic range.

Highly frequent words tend to be used in a variety of senses and the semantic restrictions on what words they can co-occur with are very loose. *Stor* is found to modify nouns from all semantic categories used in the Princeton WordNet. Even though three semantic categories (ABSTRACTION, SITUATION, and OBJECT)

represent 68% of the semantic range of *stor*, all semantic categories were found among the nouns modified by *stor* in the material studied. The dispersion of the semantic categories modified by *stor* is high, indicating that *stor* is a word with a wide range of usage.

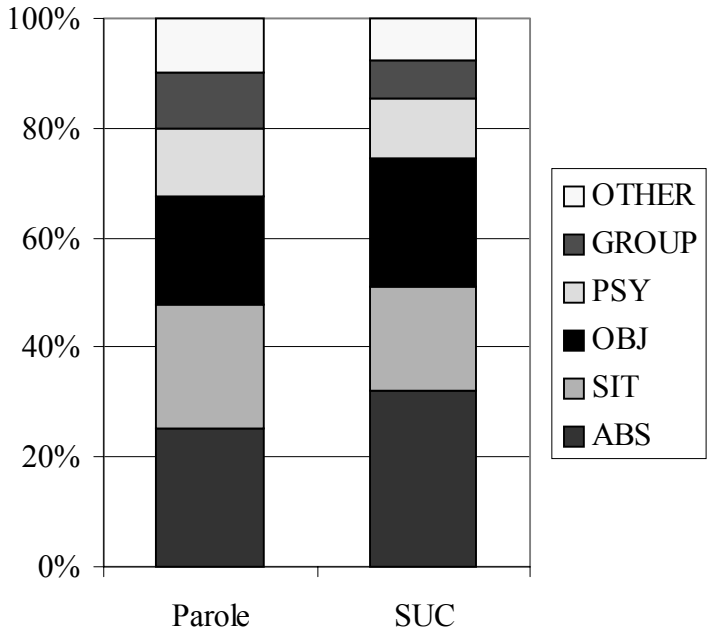


Figure 26: The semantic range of ***stor*** *in Parole and the SUC.*

8.2.5.4 *The semantic ranges of the synonyms of* stor

Naturally, the studied words vary in frequency. Table 25 gives an overview of the number of examples studied for each word, including *stor*. The material studied for the highly frequent words, i.e. *enorm*, *grov*, *kraftig*, and *stor*, are random samples of the total concordances for these words.

The synonyms of *stor* were grouped according to the semantic category or categories they most frequently modify. Following the suggestions in the discussion of semantic categories above, the ACTIVITY, EVENT, and STATE nouns were clustered in one group, SITUATION, for this study. Using this "parent node" for these closely related categories, the fourth-most frequently modified category becomes SITUATION (with seven adjectives) instead of HUMAN ACTION. The three most frequent categories are OBJECT, ABSTRACTION, and HUMAN.

Adjective	***Number of occurrences studied in***	
	SUC	***Parole***
ansenlig 'considerable'	5	101
avsevärd 'considerable'	10	133
diger 'voluminous'	4	96
dryg 'ample'	18	702
enorm 'enormous'	40	981*
framstående 'prominent'	6	226
förnäm 'distinguished'	13	142
grov 'heavy'	61	975*
inflytelserik 'influential'	6	149
innehållsrik 'containing a great deal'	4	52
kolossal 'colossal'	1	57
kraftig 'strong'	102	982*
mäktig 'powerful, immense'	27	686
oerhörd 'tremendous'	25	424
ofantlig 'enormous'	5	55
otrolig 'incredible'	7	351
riklig 'plentiful'	15	116
ryktbar 'famous'	4	54
rymlig 'spacious'	14	105
stor 'large/big'	508*	1000*
storväxt 'tall'	7	68
vidsträckt 'extensive, vast'	5	124
väldig 'huge'	57	981
ädel 'noble, magnanimous'	17	167

Table 25: Frequency of the words in the test set in the SUC and the Parole corpus. An asterisk (*) indicates that a word occurs more than 1,000 times in the corpus; in these cases, a subset of the concordance was randomly chosen for the study.

The frequencies of the semantic categories were ranked in ascending order for each adjective. The words were then grouped according to the semantic category or categories that they most frequently modified, as shown in the boxes below. Some of the words had high frequencies of several semantic categories, namely *avsevärd*, *enorm*, *förnäm*, *kolossal*, *mäktig*, *oerhörd*, and *otrolig*. These words will appear in several different groups.

Adjectives that frequently modify OBJECTS
förnäm 'distinguished'
innehållsrik 'containing a great deal'
kolossal 'colossal'
mäktig 'powerful, immense'
ofantlig 'enormous'
riklig 'plentiful'
rymlig 'spacious'
vidsträckt 'extensive, vast'
väldig 'huge'
ädel 'noble, magnanimous'

Adjectives that frequently modify ABSTRACTIONS
ansenlig 'considerable'
avsevärd 'considerable'
diger 'voluminous'
dryg 'ample'
enorm 'enormous'
kolossal 'colossal'
oerhörd 'tremendous'
otrolig 'incredible'

Adjectives that frequently modify SITUATIONS
avsevärd 'considerable'
enorm 'enormous'
grov 'heavy'
kolossal 'colossal'
kraftig 'strong'
oerhörd 'tremendous'
otrolig 'incredible'

Adjectives that frequently modify HUMANS
framstående 'prominent'
förnäm 'distinguished'
inflytelserik 'influential'
mäktig 'powerful, immense'
ryktbar 'famous'
storväxt 'tall'

*Table 26: Synonyms of **stor** divided into four primary groups according to the semantic category (categories) most frequently modified.*

Förnäm and *mäktig* appear in two groups: primary OBJECT modifiers and primary HUMAN modifiers. *Avsevärd*, *oerhörd*, *enorm*, and *otrolig* are primary modifiers of both ABSTRACTIONS and SITUATIONS. *Kolossal* will appear in three different groups: primary modifiers of OBJECTS, ABSTRACTIONS, and SITUATIONS, respectively.

Some words ended up in unexpected groups, e.g. *förnäm*, *mäktig*, and *ädel*. According to my intuition, their core meanings are to do with human qualities, but they turned out to modify OBJECTS more often than any other category. However, HUMAN is also a highly frequent category in the semantic ranges of these words. It is also a bit surprising to find *grov* and *kraftig* in the group of adjectives that most frequently modify SITUATIONS, rather than OBJECTS.

8.2.5.4.1 Adjectives that frequently modify OBJECTS

The physical properties of entities received considerable attention in the literature during the 70s and 80s, especially in the light of noun classifiers (Denny 1976, 1979; Allan 1977). "Extendedness", "interioricity", "size", "consistency", "arrangement", "quanta", "material", and "location" are some characteristic properties of OBJECTS that have been discussed (see Frawley 1992:121ff for a review).

One of the major physical properties of OBJECTs is *size*. In reverse, size is prototypically thought of as a feature of OBJECTs. If a person is asked to think of something small or large, the first that comes to mind are concrete OBJECTs and not distances, time periods, or actions. However, only 24% of the nouns modified by *stor* in the SUC are OBJECTs, while 31% are ABSTRACTIONs.

The group of synonyms that often modify OBJECTs is the largest; 10 of the 27 words belong to this group, see Table 27.

förnäm 'distinguished'
innehållsrik 'containing a great deal'
kolossal 'colossal'
mäktig 'powerful, immense'
ofantlig 'enormous'
riklig 'plentiful'
rymlig 'spacious'
vidsträckt 'extensive, vast'
väldig 'huge'
ädel 'noble, magnanimous'

Table 27: Adjectives that frequently modify OBJECTs in Parole.

Below are examples of each of the adjectives modifying an OBJECT.

*ett av världens **förnämsta** konserthus*	'one of the *finest* concert halls in the world'
*ett **innehållsrikt** rödvin*	'a red wine *containing a wide range of flavours*'
*en **kolossal** lejongul sammetssoffa*	'a *colossal* tawny velvet sofa'
*en **mäktig** glaciär*	'an *immense* glacier'
*en **ofantlig** murad skorsten*	'a *huge* brick chimney'
***riklig** mandelmassa*	'*plentiful* almond paste'
*en **rymlig** kupé*	'a *spacious* train compartment'
*det **vidsträckta** campusområdet*	'the *vast* campus area'
*ett **väldigt** portvalv*	'a *huge* doorway'
***ädel** linnedamast*	'*noble* linen damask'

Though the modified noun is an OBJECT in all of the examples above, there is considerable variation as regards what feature of the noun is modified. Only four of the adjectives clearly modify the size of the OBJECT, namely *kolossal*, *ofantlig*, *väldig*, and *mäktig*. In some of the examples, the focus is on the OBJECT as a container, either on its volume, e.g. *en rymlig kupé* 'a spacious train compartment', or on its content, e.g. *ett innehållsrikt vin* 'a red wine containing a wide range of flavours'. Closely related to the first group above, with modification of size, is the case of modification of extendedness, e.g. *det vidsträckta campusområdet* 'the vast campus area'.

One characteristic feature of OBJECTs is their ability to have both concrete and abstract properties, whereas for example ABSTRACTIONs do not have concrete properties. Two of the examples above are modifications of abstract properties, namely those with *förnäm* and *ädel*.

Ädel 'noble, magnanimous' is etymologically related to *adel* 'nobility' (Hellqvist 1948). Its synonymy relation with *stor* is apparent only when it modifies HUMAN noun-phrase heads, e.g. *en ädel man* 'a noble man'*, den ädle riddaren* 'the noble knight'. In this sense it is closely related to *förnäm* 'distinguished': the human is big in the sense that he has accomplished many good deeds or thoughts. In the examples with *ädel* modifying OBJECTs the concept of 'nobility' is more important than that of 'largeness'. For instance, *en ädel panna* 'a noble forehead' says nothing about the size of the forehead – it describes its shape, meaning that the forehead is high. *Ett ädelt väggparti* 'a noble wall section' similarly means that the wall is good in some sense, either in terms of material or in terms of shape.

From the discussion above it can be concluded that even though I have delimited this group to concern only OBJECTs, the features actually modified by the dimensional adjectives are quite diverse.

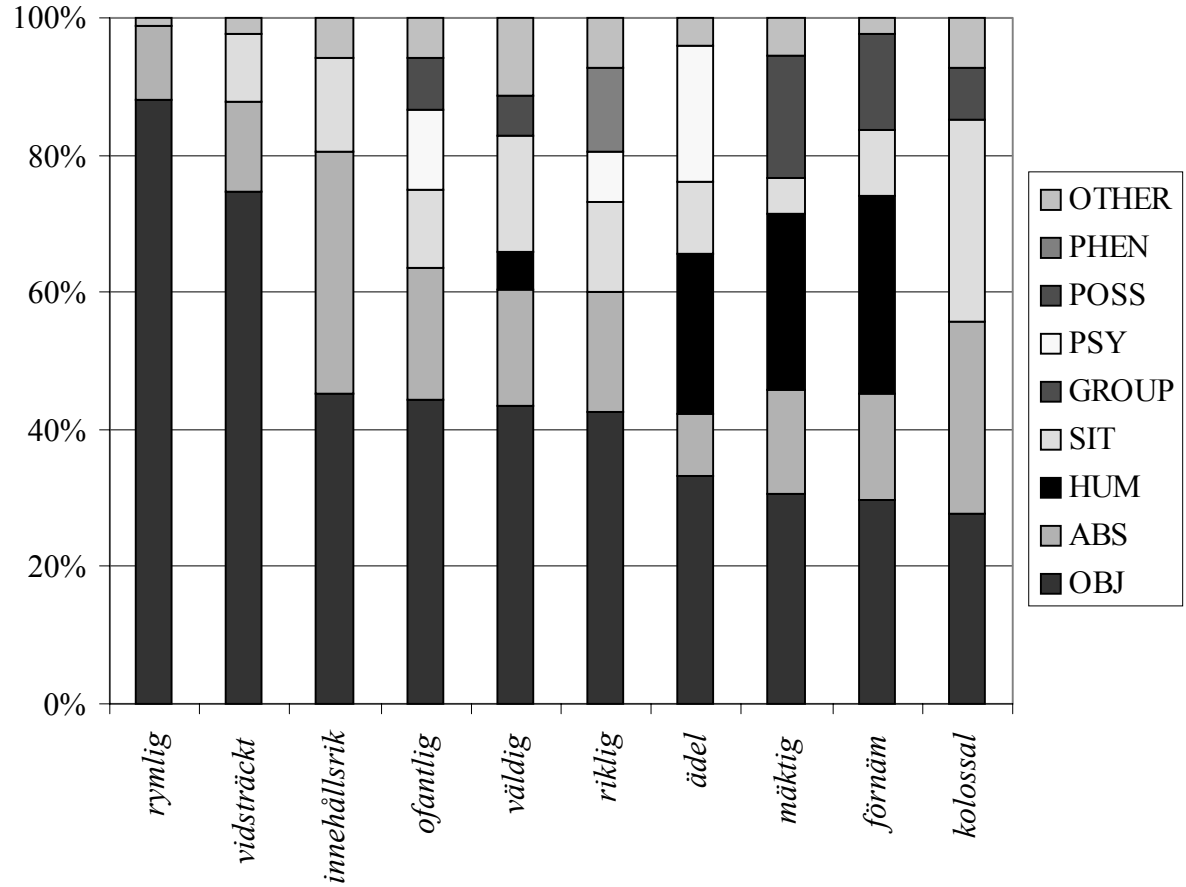


*Figure 27: The semantic ranges of synonyms of **stor** frequently modifying OBJECTs in Parole.*

The diagram in Figure 27 visualises the distribution of the semantic categories for the various adjectives is presented below. Comments on the semantic range of each of the words follow in the same order as the words are presented in Figure 27.

Rymlig 'spacious'. The majority of the nouns modified by *rymlig* are OBJECTS, and in most of the cases they refer to different types of containers: rooms, e.g. *rum* 'room', *husvagn* 'caravan', and vessels, e.g. *kärl* 'vessel', *kopp* 'cup', *sportbag* 'sports bag', *kaftan* 'caftan'. Also, instances of ABSTRACTIONS are modified in the *container* sense; see the examples below.

*Katt --- detta **rymliga** påsiga ord !*

'Cat --- this *spacious* baggy word!'

*Romanen är en **rymlig** genre,...*

'The novel is a *spacious* genre,…'

The first example above may seem strange at first sight, but it is taken from a monologue about what is in a word, more specifically in the word *katt* 'cat':

*...kunde han ens visa fram en katt som var en katt? Katt – detta **rymliga** påsiga ord! Han*

'…could he even show a cat that was a cat? Cat – this *spacious* baggy word! He

hade lärt sig att vara noga med orden. Men ju mer han synade dem, dess suddigare

had learned to be careful with his words. But the closer he looked at them, the fuzzier

blev de. Siames, angora eller bondkatt – vilket avsågs? Och katt eller katta?

they became. Siamese, Angora, or alley cat – which was intended? And tomcat or she-cat?

också kattan är katt. Han bemödade sig. Och under alla mödorna brast själva tråden som

the she-cat is a cat, too. He took pains. And all those pains tore the thread that linked the

länkade ordet till tinget! När han skrivit ner ordet och såg tillbaka på katten

word to the object! When he had written the word down and looked back at the cat

som gett upphov till det, då gick katten där och var helt oberörd –

that had caused it all, then the cat was walking about, supremely indifferent –

gav katten i vad som hänt!

couldn't care less about what had happened!'

In this example, a word is viewed as a container than can be linked to various referents, e.g. a cat can be a Siamese, an Angora cat, or an alley cat. *Genre* is also viewed as a container, which in this case is filled with different types of novels.

There is one example of *rymlig* modifying a SITUATION, actually a STATE: *rymlig tystnad kring tingen* 'spacious silence around the things'. Even this is a type of abstract container. Judging from the examples studied, it would seem that *rymlig* consistently modifies containers.

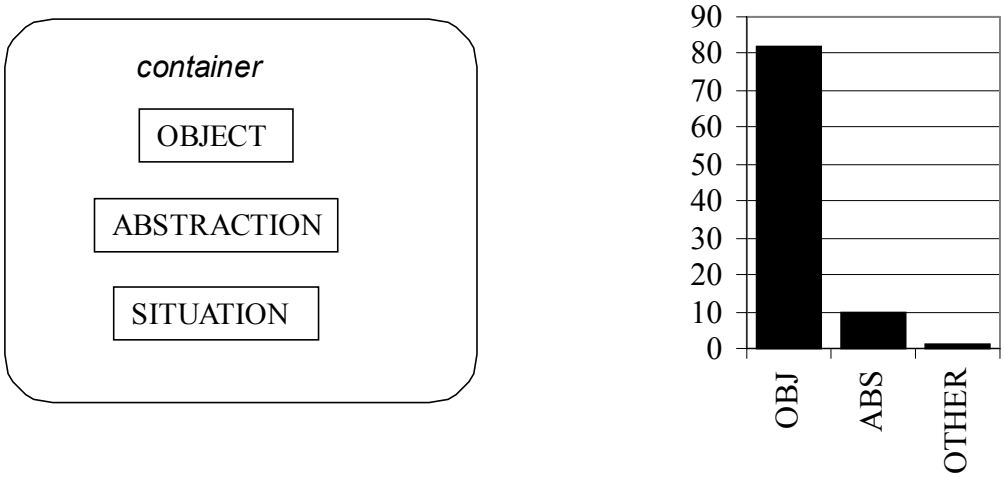


Figure 28: The semantic range of ***rymlig*** *'spacious'.*

Vidsträckt 'extensive, vast'. Like *rymlig*, *vidsträckt* most frequently modifies OBJECTs, but what is modified here is not a volume but rather a surface. There are many types of land or sea areas modified by *vidsträckt*, e.g. *Aultermyren* 'the Aulter swamp', *risfält* 'rice fields', *havsområden* 'marine areas', *lagun* 'lagoon'. Analogous examples of ABSTRACTIONs and SITUATIONs are *popularitet* 'popularity', *resor* 'trips'. A few instances modifying PSYCHOLOGICAL FEATUREs were found, e.g. *vidsträckta kunskaper* 'extensive knowledge'.

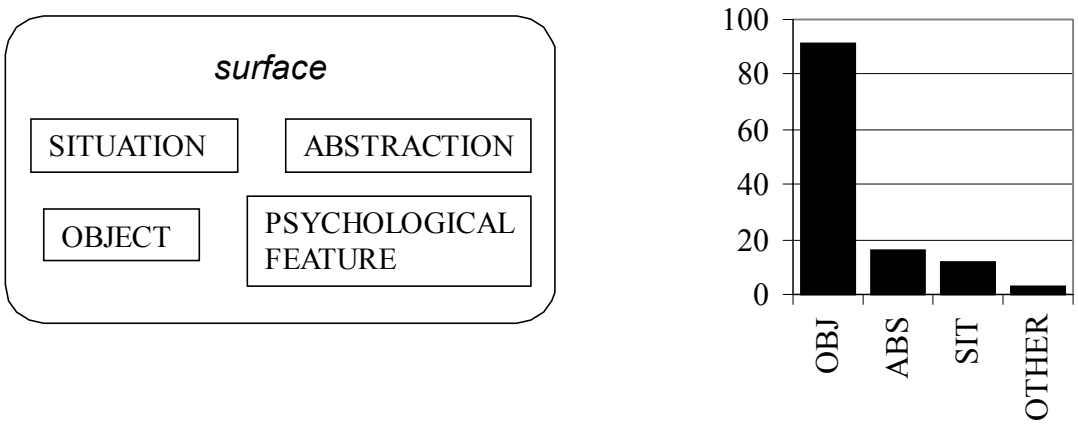


*Figure 29: The semantic range of **vidsträckt** 'extensive, vast'.*

Innehållsrik 'containing a great deal'. *Innehållsrik* brings forward the content of a concrete or abstract container. It can refer either to the size of the content or to its diversity, e.g. the OBJECTS *bibliotek* 'library', *fisksoppa* 'fish soup', *smörgåsbord* 'smorgasbord'. Some ABSTRACTIONS on the same theme are *halvtimme* 'half-hour', *rymdberättelse* 'space story'. There are two examples of organisations containing a great deal in the material, namely *institutionen* 'the institute' and *redaktion* 'editorial office', which are categorised as GROUPS.

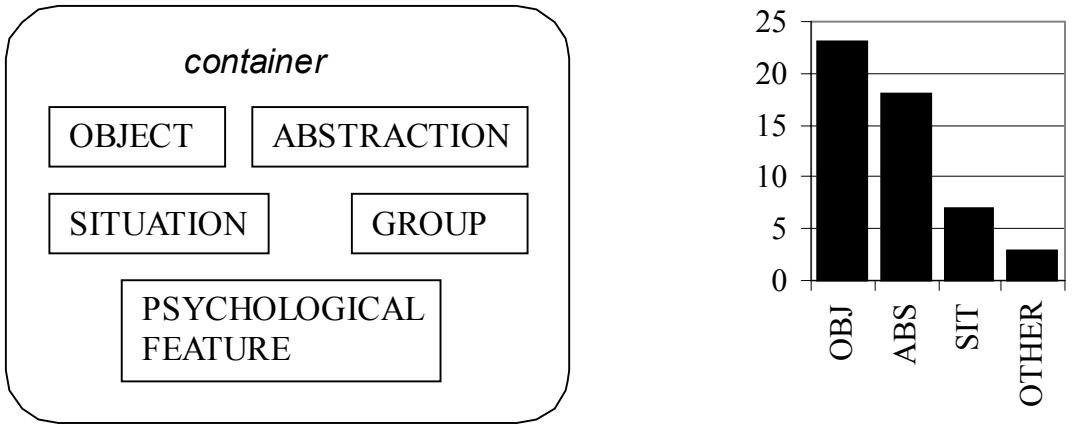


*Figure 30: The semantic range of **innehållsrik** 'containing a great deal'.*

Ofantlig 'enormous'. The meaning of *ofantlig* is a magnification of the core meaning of *stor*, i.e. the size of something is not only large, but very large. Apart from OBJECTS, the semantic range of *ofantlig* covers ABSTRACTIONS, e.g. *mängd* 'amount', SITUATIONS, e.g. *nedskärningar* 'reductions', CREATURES, e.g. *drake* 'dragon', MONETARY REPRESENTATIONS, e.g. *förmögenhet* 'fortune', and PSYCHOLOGICAL FEATURES, e.g. *känsla* 'feeling'.

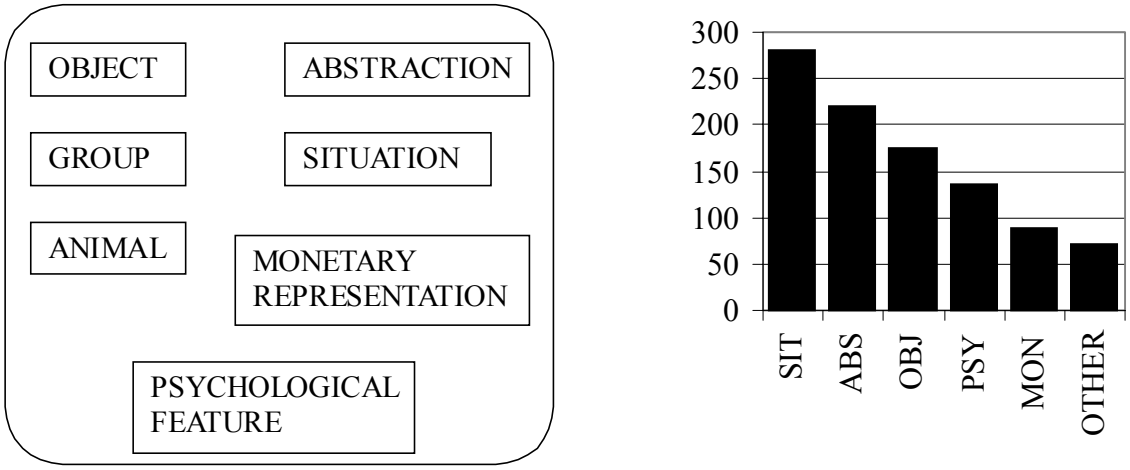


*Figure 31: The semantic range of **ofantlig** 'enormous'.*

Väldig 'huge'. *Väldig* is highly frequent and closely related in meaning to *ofantlig*; it says something about the dimension of the noun it modifies. Its semantic range is wide – nouns from most semantic categories can be modified by *väldig*; see the examples below.

OBJECTS: *hjul* 'wheel', *bål* 'fire'
ABSTRACTIONS: *smäll* 'bang', *fart* 'speed'
HUMANS: *matrona* 'matron'
SITUATIONS: *utveckling* 'development', *tittande* 'gazing'
GROUPS: *stjärnanhopning* 'cluster of stars', *kö* 'line (of people)'
PSYCHOLOGICAL FEATURES: *aptit* 'appetite', *smärta* 'pain'
MONETARY REPRESENTATIONS: *budgetunderskott* 'budget deficit', *rikedomar* 'riches'
ANIMALS: *tranor* 'cranes', *tjur* 'bull'
PHENOMENA: *åskväder* 'thunderstorm'

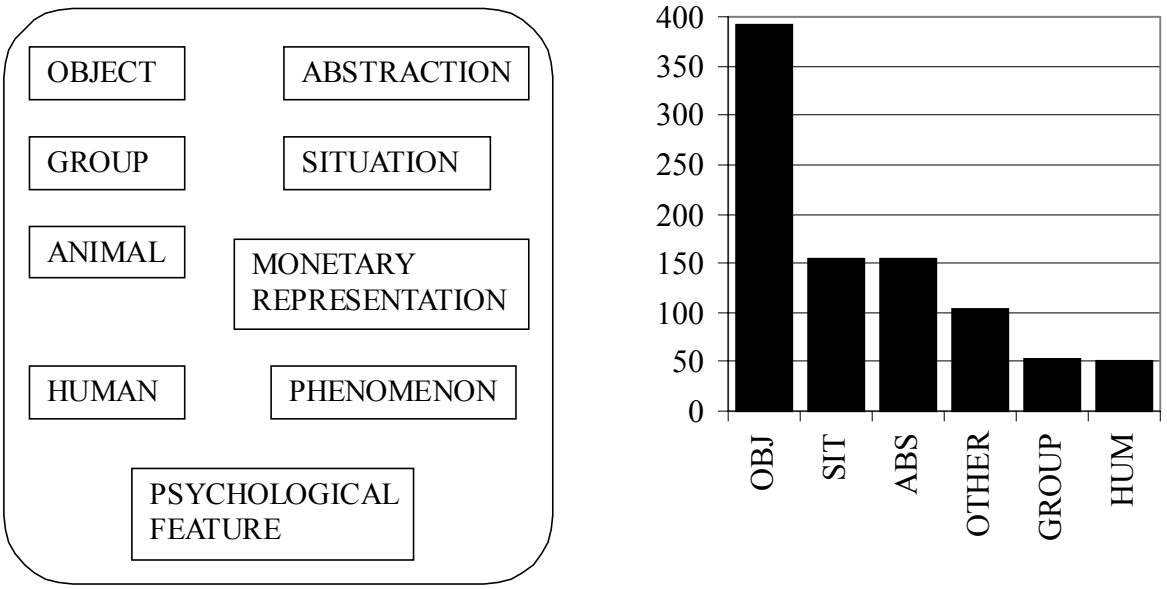


*Figure 32: The semantic range of **väldig** 'huge'.*

Riklig ‘plentiful’. *Riklig* has a more quantifying meaning than the words that modify dimensions; it indicates that there is a large amount of something. For example, *en riklig portion* ‘an abundant portion’ is a large amount of food. For SITUATIONS, the amount of activity is what is quantified, e.g. *den rikliga korrespondensen* ‘the plentiful correspondence’, *en riklig tillströmning* ‘a great influx’.

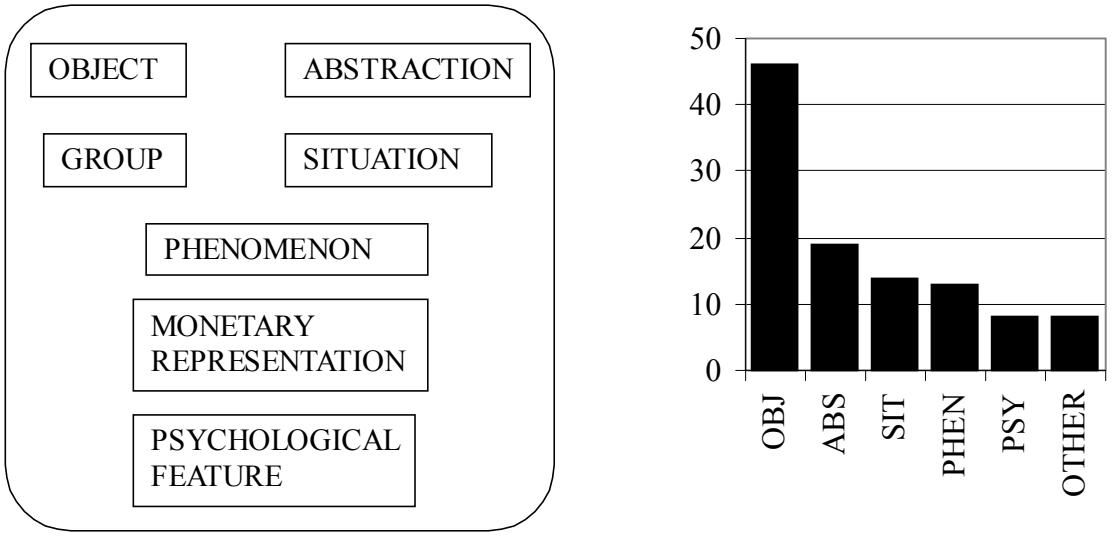


*Figure 33: The semantic range of **riklig** ‘plentiful’.*

Ädel ‘noble, magnanimous’. *Ädel* refers mainly to a non-dimensional quality of the modified noun. *Den ädle riddaren* ‘the noble knight’ is a *good man*, and in this sense he is also a *big man*. Other semantic categories than HUMANS that can be “good” are OBJECT, e.g. *en ädel kroknäsa* ‘a noble hook nose’, SITUATION, e.g. *ädla handlingar* ‘noble deeds’, ABSTRACTION, e.g. *ädelt modersmål* ‘noble mother tongue’, GROUP, e.g. *ädle släkt* ‘noble family’, and PSYCHOLOGICAL FEATURE, e.g. *ädla känslor* ‘noble feelings’.

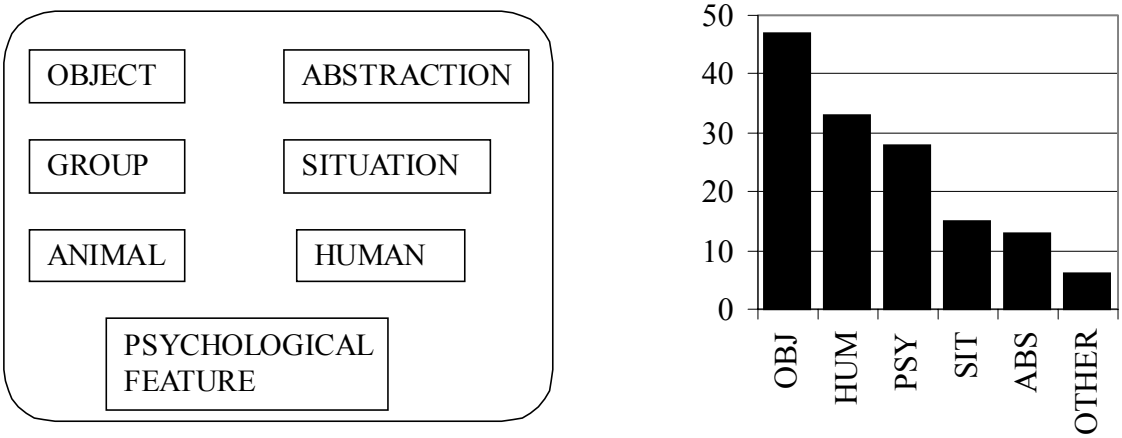


*Figure 34: The semantic range of **ädel** ‘noble, magnanimous’.*

Mäktig ‘powerful, immense’. *Mäktig* has several meanings. The most frequent type of noun modified is OBJECT; in this case *mäktig* says something about the size of the OBJECT, e.g. *en mäktig fornborg* ‘an immense ancient fortress’, *det mäktiga arbetsbordet* ‘the huge desk’. One specific meaning of *mäktig* that modifies only food OBJECTs is ‘heavy’. This does not relate to the quantity of food, but is used about types of food that cause satisfaction in small amounts. Nouns from the categories HUMAN, GROUP, and ORGANISATION can be modified by *mäktig* in the sense of ‘having great power’, e.g. *en mäktig president* ‘a powerful president’, *den mäktiga lobbygruppen* ‘the powerful lobby group’, *mäktig centralbank* ‘powerful central bank’.

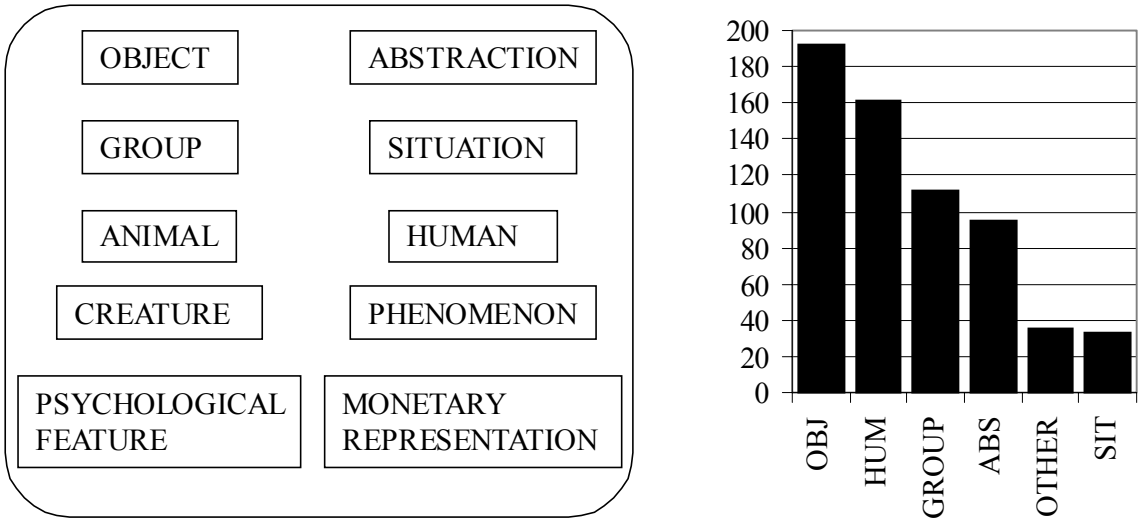


*Figure 35: The semantic range of **mäktig** ‘powerful, immense’.*

Förnäm ‘distinguished’. *Förnäm* is closely synonymous with *ädel* – it refers to a non-dimensional quality of the noun it modifies. For concrete concepts it implies that the OBJECT is of very high quality, e.g. *de förnämsta av Vasas skulpturer* ‘the most distinguished of Vasa’s sculptures’. Quite often, *förnäm* modifies HUMAN head nouns, indicating that the person is distinguished in some sense, e.g. *de förnäma fruarna* ‘the distinguished ladies’, *en i högsta grad förnäm linoleumsnittare* ‘a very distinguished linoleum carver’. For HUMANs *förnäm* can have a negative connotation, meaning ‘haughty, supercilious’, *de förnäma fruarna* is an example of this. An example of *förnäm* modifying an ABSTRACTION is *med förnäm min* ‘with a distinguished face’, which also has a negative connotation. In the instances where SITUATIONs are modified by *förnäm*, it has the same sense of ‘distinguished, good’, sometimes with a negative twitch to it, e.g. *förnäma bröllop* ‘distinguished weddings’.

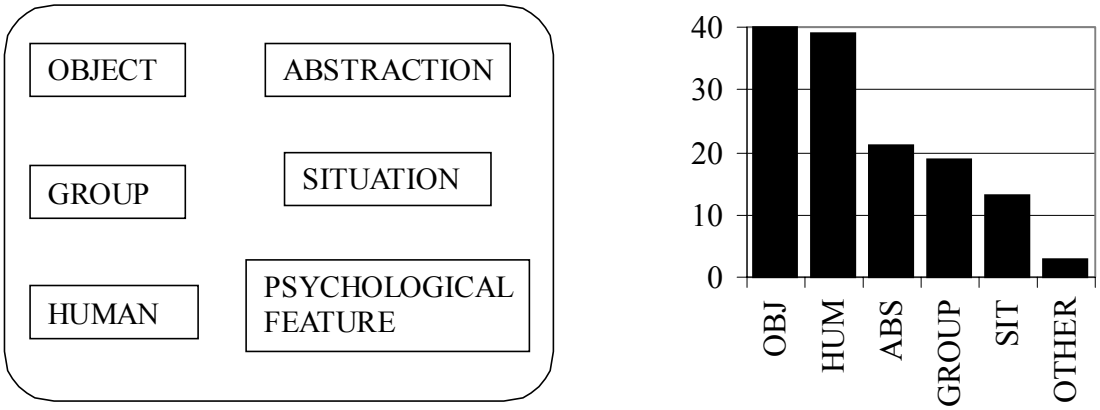


*Figure 36: The semantic range of **förnäm** 'distinguished'.*

Kolossal 'colossal'. *Kolossal* has almost equal parts of its semantic range in the three categories OBJECT, ABSTRACTION, and SITUATION. Examples of each category from the corpus are *kolossala orkidéer* 'colossal orchids', *ett kolossalt oljud* 'a loud noise', and *kolossala förändringar* 'enormous changes'. *Kolossal* resembles the leftmost words in Figure 27, i.e. *rymlig*, *vidsträckt*, *innehållsrik*, *ofantlig*, and *riklig*, in the sense that none of them is found to modify HUMANS. However, it is easy to accept both *kolossal* and *ofantlig* with HUMAN nouns, e.g. *en kolossal man* 'a colossal man' and *en ofantlig kvinna* 'a colossal woman'. The three leftmost words would seem less likely to combine with HUMANS, but such combinations could probably not be entirely ruled out.

Combinations with GROUP, PHENOMENON, PSYCHOLOGICAL FEATURE, and MONETARY REPRESENTATION are rare. Examples are:

GROUP: *En kolossal samling dubbletter* 'An enormous collection of duplicates'

PHENOMENON: *en kolossal tur* 'enormous luck'

PSYCHOLOGICAL FEATURE: *en kolossal nyfikenhet* 'enormous curiosity'

MONETARY REPRESENTATION: *kolossala belopp* 'colossal amounts'

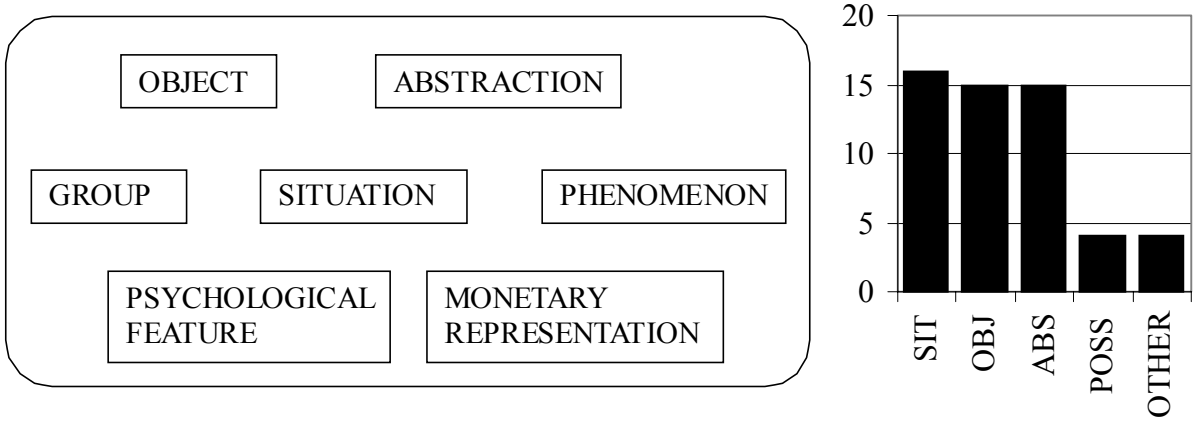


*Figure 37: The semantic range of **kolossal** 'colossal'.*

All adjectives that frequently modify OBJECTs can also modify ABSTRACTIONs. This is true for SITUATIONs and PSYCHOLOGICAL FEATUREs as well, though only one example of *rymlig* modifying a SITUATION is found in the material: *rymlig tystnad kring tingen* 'spacious silence around the things'. The major category HUMAN, however, is present only in the columns to the right: *väldig*, *ädel*, *mäktig*, and *förnäm* (and also for *stor*). There also seems to be a relation between the categories HUMAN and GROUP: if an adjective can modify HUMAN head nouns it is also likely to modify GROUPs. It is especially striking how *mäktig* and *förnäm*, which have a large portion of their head nouns in the category HUMAN, also have a considerable number of head nouns in the category GROUP. MONETARY REPRESENTATIONs, ANIMALs, PHENOMENA, and CREATUREs are so rare that it is hard to generalise from the data in this study.

We can also discern a pattern within the group. *Rymlig* and *vidsträckt* stand out with an extremely high portion of OBJECTs in their semantic range, between 70% and 90%. The second-most frequent semantic category modified by these two words is ABSTRACTION, and this is true for *innehållsrik*, *ofantlig*, *väldig*, and *riklig* as well.

The four rightmost words in Figure 27 – *ädel*, *mäktig*, *förnäm*, and *kolossal* – all have a significant number of HUMAN head nouns among the words they modify in the corpus. HUMAN is the second-most frequent category modified by them.

A closer look at the words in the four rightmost columns shows that the meaning of *ädel*, when modifying an OBJECT, is usually 'noble' or 'pure', e.g. *ädla stenar* 'precious stones', *ädel metall* 'precious metal', *ädla druvsorter* 'noble grapes'. Most of the OBJECTs that are modified by *mäktig* are actually modified in their concrete dimension, e.g. *ett mäktigt skrivbord* 'an immense desk', *den mäktiga katedralen* 'the immense cathedral'. *Förnäm* modifies OBJECTs in the meaning 'good', 'the result of good work', e.g. *den förnäma trävillan* 'the noble wooden house', *förnäma sniderier* 'superior carvings', *en förnäm gata* 'a distinguished street'.

An overview of the subgroups among the adjectives that frequently modify OBJECTs is presented below. There are three other semantic categories that are highly frequent in the semantic ranges of the words in this group: ABSTRACTION, HUMAN, and SITUATION.

Adjectives that primary modify OBJECTs and secondary modify		
ABSTRACTIONS	HUMANS	SITUATIONS
innehållsrik 'containing a great deal' *kolossal* 'colossal' *ofantlig* 'enormous' *riklig* 'plentiful' *rymlig* 'spacious' *vidsträckt* 'extensive, vast' *väldig* 'huge'	*förnäm* 'distinguished' *mäktig* 'powerful, immense' *ädel* 'noble, magnanimous'	*kolossal* 'colossal'

Table 28: Adjectives that frequently modify OBJECTS.

Kolossal appears in two of the subgroups above as well as in two primary groups in Table 26: ABSTRACTIONS and SITUATIONS. This reflects the fact that the three categories OBJECT, ABSTRACTION, and SITUATION represent nearly equal shares of its semantic range. *Förnäm* and *mäktig* are also found in the group of primary modifiers of HUMANS – se Section 8.2.5.4.4.

8.2.5.4.2 Adjectives that frequently modify ABSTRACTIONS

The adjectives that frequently modify ABSTRACTIONS form a group that is, in a way, more homogeneous than that constituted by the adjectives found to occur more frequently with OBJECTS. Of the words in the test set, eight prototypically modify ABSTRACTIONS; see Table 29.

ansenlig 'considerable'
avsevärd 'considerable'
diger 'voluminous'
dryg 'ample'
enorm 'enormous'
kolossal 'colossal'
oerhörd 'tremendous'
otrolig 'incredible'

Table 29: Words that frequently modify ABSTRACTIONS in Parole

Examples of ABSTRACTIONS found in the material are distances, time periods, sounds, and communication, e.g.:

en *ansenlig* bredd	'a considerable width'
en *avsevärd* sträcka	'a considerable distance'
ett *digert* sommarprogram	'a substantial "Summer" programme on the radio'
en *dryg* kilometer	'a good kilometre = slightly more than a kilometre'
med så *enorma* hastigheter…	'at such enormous speeds…'
kolossala mängder böcker	'enormous amounts of books'
en *oerhörd* nyhet	'tremendous news'
en *otrolig* styrka	'an incredible strength'.

Three of the words are also found in other groups: *enorm*, *kolossal*, and *otrolig*.

The meaning of the dimensional adjectives seems more consistent when the modified noun is an ABSTRACTION than it was found to be for the OBJECT noun heads. The words all have to do with quantity, a quantity larger than average or larger than expected. The distances, time periods, and sounds are possible to quantify somehow, while for the instances of communication the size cannot be measured, e.g. *ett digert sommarprogram* 'a substantial "Summer" programme on the radio', *en oerhörd nyhet* 'tremendous news'.

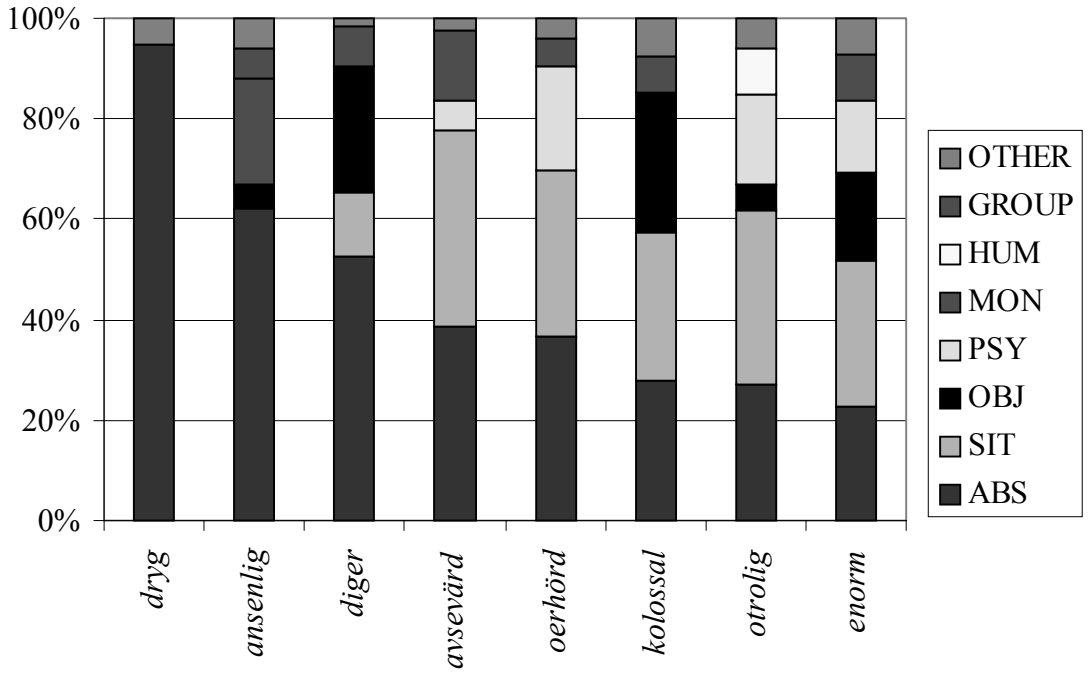


Figure 38: The semantic ranges of synonyms of ***stor*** *frequently modifying* ABSTRACTIONS *in Parole.*

The synonyms of *stor* that frequently modify ABSTRACTIONS are also found to modify OBJECTS, SITUATIONS, and PSYCHOLOGICAL FEATURES. MONETARY REPRESENTATIONS are modified by all words except *diger*. HUMAN head nouns are found for *drygt* and *oerhörd* but not for any other word in this group. All words except *diger* modify GROUPS.

The semantic ranges of the words will be further described below. They are presented in the same order as they appear in Figure 38.

Drygt 'ample'. Over 90% of the noun heads modified by *drygt* belong to the category ABSTRACTION, e.g. *en dryg timme* 'a good hour', *en dryg tredjedel* 'a good third'. It clearly has a quantifying function. The remaining 8% of instances of *drygt* modify SITUATIONS, HUMANS, PSYCHOLOGICAL FEATURES, and MONETARY REPRESENTATIONS; see the examples below.

SITUATION: *ett* ***drygt*** *arbete* 'a *heavy* task'

HUMAN: *en stor bardisk med **dryg** bartender* 'a big bar (equipped) with a *haughty* bartender'

PSYCHOLOGICAL FEATURE: *hans ... lagom **dryga**, lagom faktabemängda åsikter* 'his … opinions, with just the right amount of *haughtiness*[27] and a precisely adjusted sprinkling of facts'

MONETARY REPRESENTATION: *klubbens **dryga** skulder* 'the club's *heavy* debts'

In the examples found of *dryg* modifying HUMANS and PSYCHOLOGICAL FEATURES, it is not used in a dimensional sense, but rather equivalent to 'self-important'. MONETARY REPRESENTATION is the second-most common type of noun modified by *dryg* after ABSTRACTION. The semantic range of *dryg* is visualised in Figure 39.

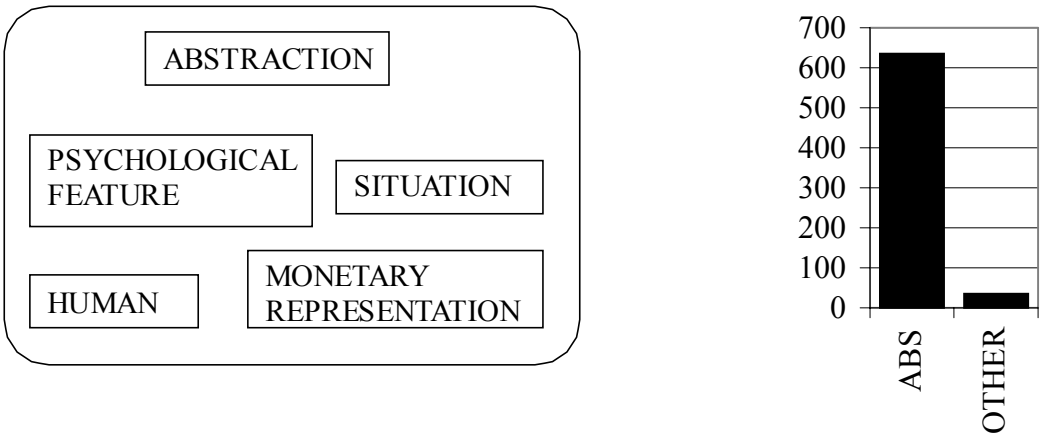


*Figure 39: The semantic range of **dryg** 'ample'.*

Ansenlig 'considerable'. *Ansenlig* has slightly over 60% of its head nouns in the category ABSTRACTIONS, e.g. *en ansenlig mängd* 'a considerable amount'. Of the remaining 40%, half of the head nouns are MONETARY REPRESENTATIONS, e.g. *ansenlig förmögenhet* 'considerable fortune'. *Ansenlig* has a quantifying function and will form a group together with *dryg*: adjectives that most frequently modify ABSTRACTIONS and second-most frequently MONETARY REPRESENTATIONS. There are also some examples of *ansenlig* modifying OBJECTS, SITUATIONS, PSYCHOLOGICAL FEATURES, and GROUPS in the material, e.g. *en ansenlig binge reselyrik* 'a considerable heap of travel poetry', *ett ansenligt välstånd* 'considerable wealth', *ditt ansenliga självförtroende* 'your considerable self-confidence', and *en ansenlig fiendeskara* 'a considerable group of enemies'. Figure 40 shows the semantic range of *ansenlig*.

[27] This is an example of a case where it is difficult to find an equivalent of the Swedish adjective within the word class of adjectives in English. The adjective *dryg* is here translated into a noun, 'haughtiness'.

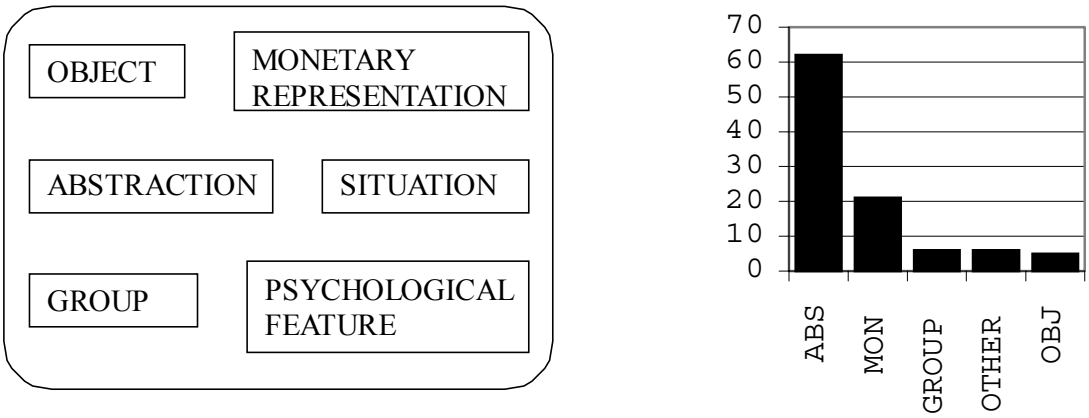


*Figure 40: The semantic range of **ansenlig** 'considerable'.*

Diger 'voluminous'. Though the majority of the occurrences of *diger* modify ABSTRACTIONS, e.g. *en diger antologi* 'an extensive anthology', OBJECTS occupy an important part of its semantic range. All of the examples of OBJECTS modified by *diger* border to the semantic category GROUPS, e.g. *julbord* 'Christmas buffé', *trave* 'pile', and *lunta* 'bundle'. SITUATIONS, PSYCHOLOGICAL FEATURES, and GROUPS are also represented in the material, e.g. *ett digert arbete* 'extensive work', *digra erfarenheter* 'a great deal of experience', and *sin redan digra medaljsamling* 'his already huge collection of medals', cf. Figure 41.

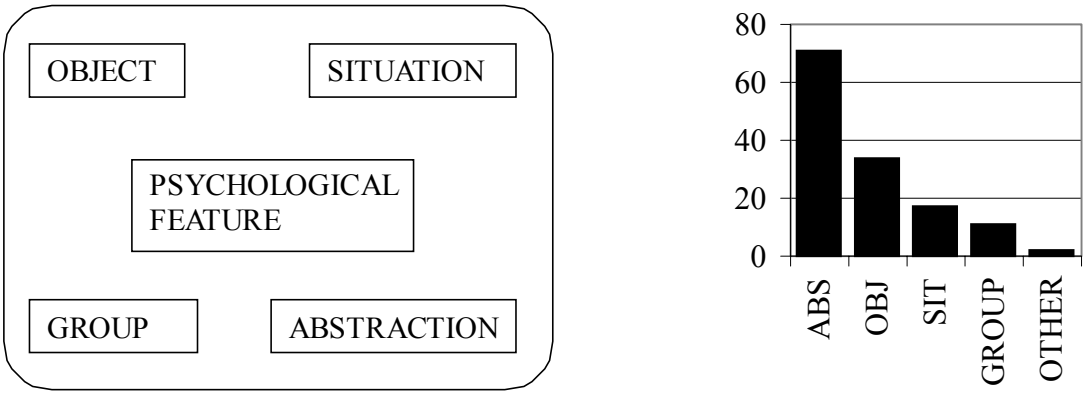


*Figure 41: The semantic range of **diger** 'voluminous'.*

Avsevärd 'considerable'. *Avsevärd* modifies OBJECTS, SITUATIONS, GROUPS, ABSTRACTIONS, PSYCHOLOGICAL FEATURES, and MONETARY REPRESENTATIONS, cf. Figure 42. SITUATIONS, e.g. *förbättring* 'improvement', *framgång* 'success', *investering* 'investment', and *konsekvenser* 'consequences', make up a considerable part of the semantic range of *avsevärd*. Therefore, *avsevärd* is a member not only of this group but also of the group of words often modifying SITUATIONS. It will be doubly categorised: as an ABSTRACTION-modifying adjective with SITUATIONS as its secondary category, and vice versa. Examples of the use of *avsevärd* are:

OBJECT: ***avsevärda*** *vapenlager* '*considerable* stores of arms'

SITUATION: *en* ***avsevärd*** *vårdförbättring* 'a *considerable* improvement of the health-care'

GROUP: *en* ***avsevärd*** *publik* 'a *big* audience'

ABSTRACTION: ***avsevärd*** *storlek* '*considerable* size'

PSYCHOLOGICAL FEATURE: ***avsevärd*** *nostalgi* '*considerable* nostalgia'

MONETARY REPRESENTATION: ***avsevärda*** *belopp* '*considerable* amounts'

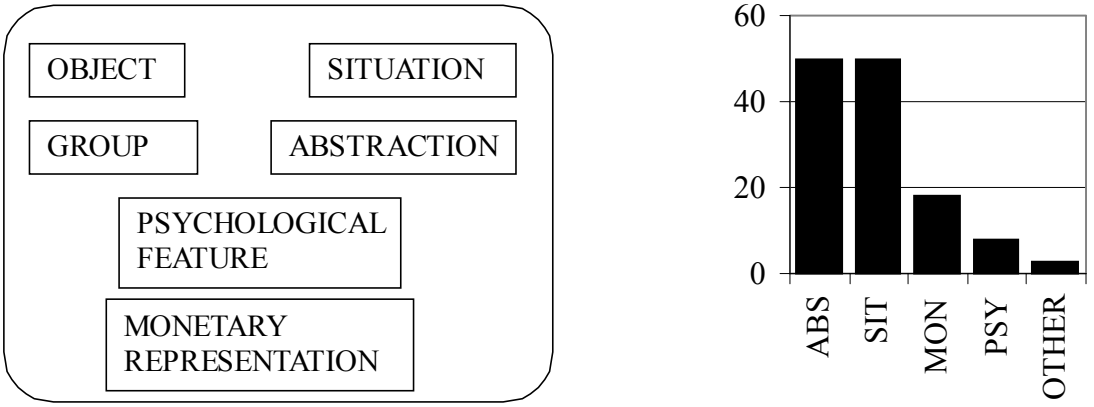


Figure 42: The semantic range of ***avsevärd*** *'considerable'.*

Oerhörd 'tremendous'. 35% of the nouns modified by *oerhörd* are of the type ABSTRACTION, e.g. *oerhörda skillnader* 'tremendous differences'. The semantic range of *oerhörd* resembles that of *avsevärd* in the large portion of SITUATIONS, e.g. *en oerhörd frihet* 'a tremendous freedom'. However, the semantic restrictions are less strict (cf.

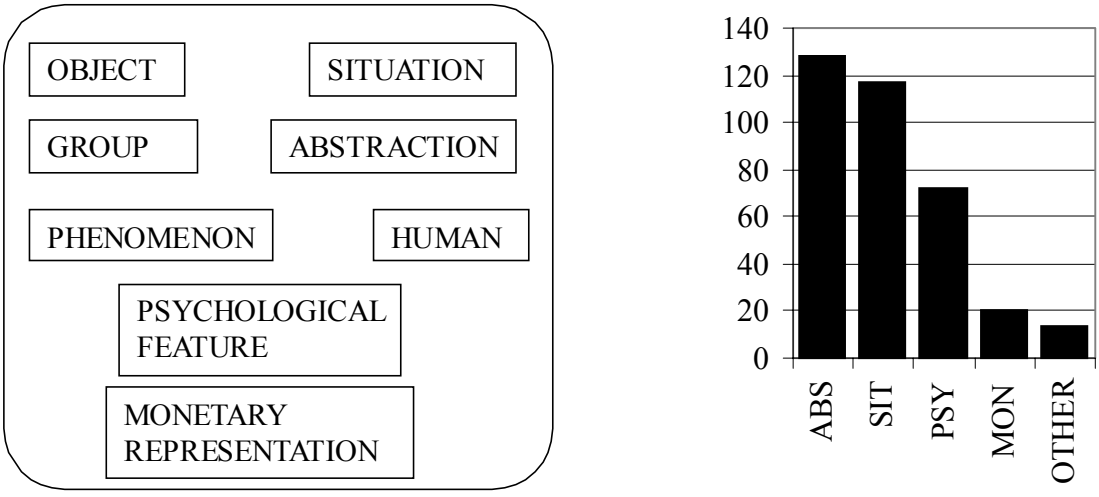


Figure 43). In addition, there are two rare semantic categories in the material: PHENOMENA, e.g. *det oerhörda gastrycket* 'the tremendous gas pressure', and HUMANS, e.g. *oerhörda snillen* 'tremendous geniuses'. Together with *avsevärd*, *oerhörd* forms a group of ABSTRACTION-modifying adjectives that second-most

often modify SITUATIONS. 20% of the nouns modified belong to the category PSYCHOLOGICAL FEATURE, e.g. *vilken oerhörd lyckokänsla* 'what a tremendous feeling of happiness'. There are some examples of OBJECTS and MONETARY REPRESENTATIONS: *en oerhörd tredimensionell arkipelag av miljarder galaxer* 'a tremendous three-dimensional archipelago of billions of galaxies' and *oerhörd vinst* 'huge profit'.

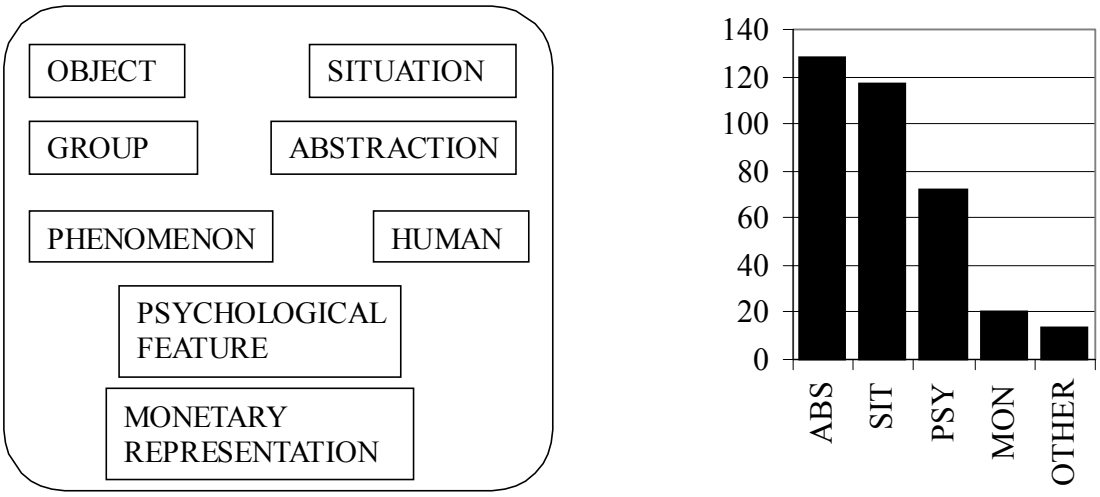


Figure 43: The semantic range of ***oerhörd*** *'tremendous'.*

The remaining words in the table below, *kolossal*, *enorm*, and *otrolig*, are described in the section on adjectives frequently modifying SITUATIONS.

Adjectives that frequently modify ABSTRACTIONS and second-most often modify		
OBJECTS	MONETARY REPRESENTATIONS	SITUATIONS
diger 'voluminous' *enorm* 'enormous' *kolossal* 'colossal'	*dryg* 'ample' *otrolig* 'incredible'	*ansenlig* 'considerable' *avsevärd* 'considerable' *enorm* 'enormous' *kolossal* 'colossal' *oerhörd* 'tremendous' *otrolig* 'incredible'

Table 30: Adjectives that frequently modify ABSTRACTIONS.

Some of the adjectives appear in several of the columns in Table 30. *Kolossal* and *enorm* both appear in the two groups whose members second-most frequently modify OBJECTS and SITUATIONS, respectively. *Otrolig* second-most often modifies both MONETARY REPRESENTATIONS and SITUATIONS. *Otrolig*, *avsevärd*, *oerhörd*, and *kolossal* are also found among the primarily SITUATION-modifying adjectives, and *kolossal* is furthermore found among the primarily OBJECT-modifying ones.

8.2.5.4.3 Adjectives that frequently modify SITUATIONS

Seven words from the test set are found in the group of words that frequently modify SITUATIONS: *avsevärd*, *enorm*, *grov*, *kolossal*, *kraftig*, *oerhörd*, and *otrolig*, cf. Table 31. Five of these words also appear in the group of adjectives frequently modifying ABSTRACTIONS, namely *avsevärd*, *enorm*, *kolossal* (which is also a primary OBJECT-modifier), *oerhörd*, and *otrolig*.

avsevärd 'considerable'
enorm 'enormous'
grov 'heavy'
kolossal 'colossal'
kraftig 'strong'
oerhörd 'tremendous'
otrolig 'incredible'

Table 31: Adjectives that frequently modify SITUATIONS in Parole.

The distribution of the semantic ranges of these seven words is shown in Figure 44.

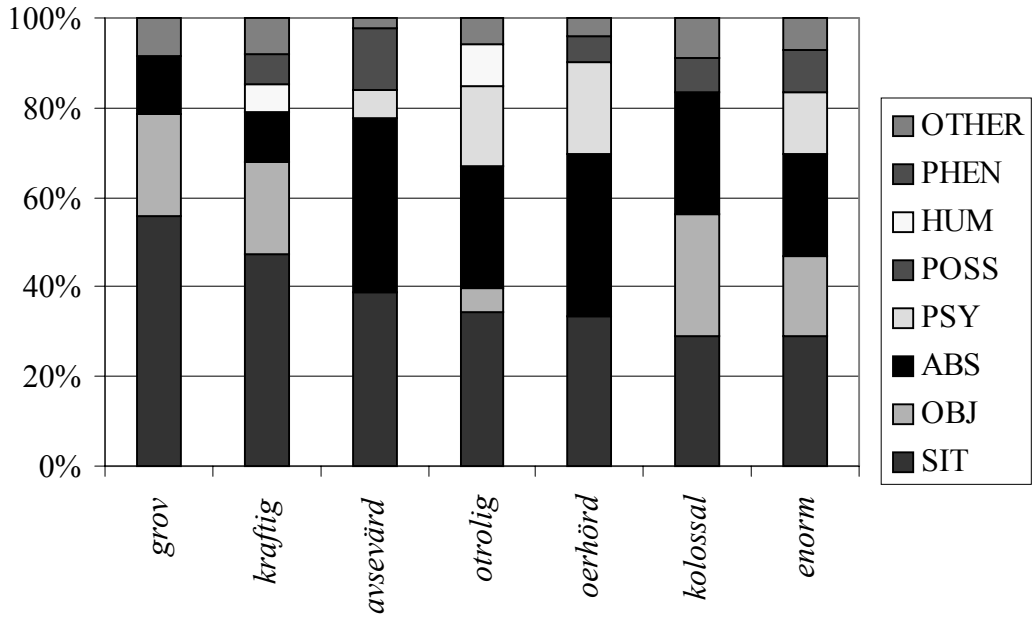
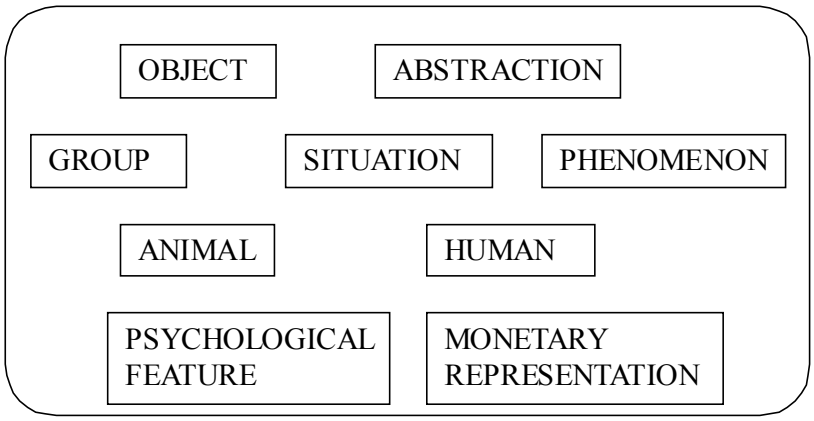


*Figure 44: The semantic ranges of synonyms of **stor** frequently modifying SITUATIONS.*

The adjectives that frequently modify SITUATIONS also often modify nouns from the categories OBJECT and ABSTRACTION. PSYCHOLOGICAL FEATURES are also represented for all the words, and so – more surprisingly – are MONETARY REPRESENTATIONS.

The four main categories, i.e. OBJECT, ABSTRACTION, SITUATION, and HUMAN, are represented in the semantic ranges of all of the words in this group. The semantic ranges here are generally very diverse; *enorm*, *grov*, *kraftig*, and

otrolig all have semantic ranges covering nine different categories, see Figure 44.



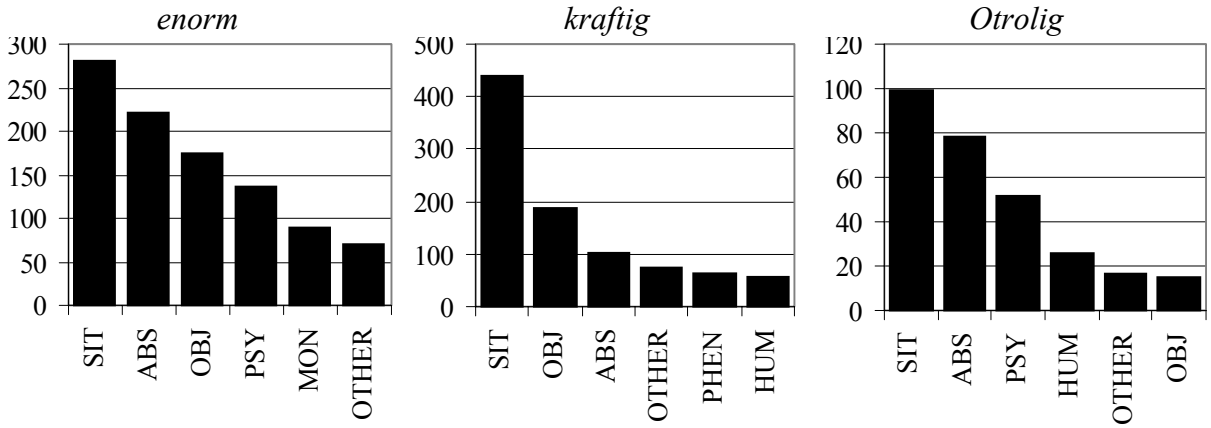


*Figure 45: The semantic ranges of **enorm** 'enormous', **grov** 'heavy', **kraftig** 'strong', and **otrolig** 'incredible'.*

Kolossal has a semantic range covering seven different categories, see Figure 46.

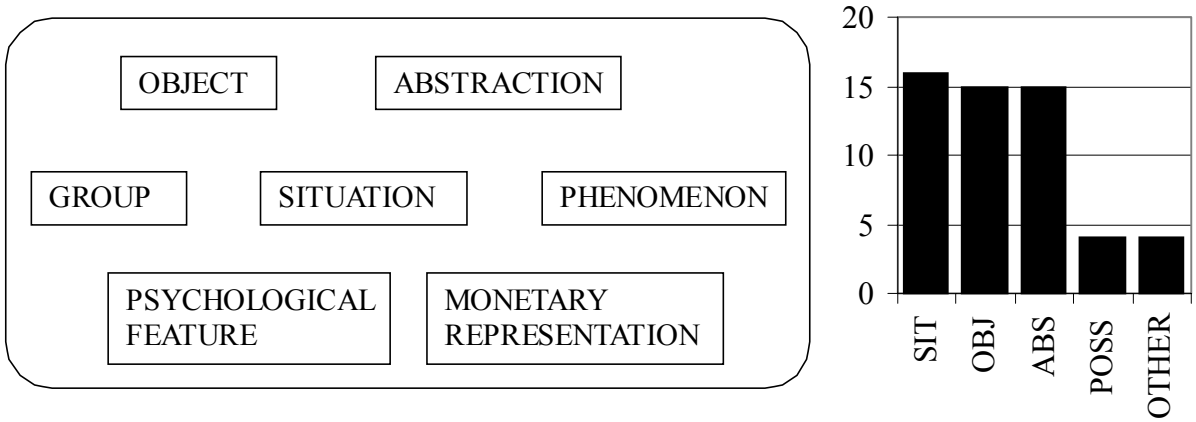


*Figure 46: The semantic range of **kolossal** 'colossal'.*

Oerhörd, *otrolig,* and *enorm* have significant numbers of PSYCHOLOGICAL FEATURES among the nouns they modify, e.g. *en otrolig känsla* 'an incredible feeling', *en enorm glädje* 'enormous joy'. This will not be reflected in the grouping in Table 32, but is nevertheless worth noting.

Avsevärd and *oerhörd* frequently modify SITUATIONS, but also modify a significant number of ABSTRACTIONS and so were discussed in an earlier section.

The second-most common semantic categories modified by the adjectives in this group are OBJECT and ABSTRACTION.

Adjectives that frequently modify SITUATIONS and second-most often modify	
OBJECTS	***ABSTRACTIONS***
enorm 'enormous' *grov* 'heavy' *kolossal* 'colossal' *kraftig* 'strong'	*avsevärd* 'considerable' *enorm* 'enormous' *kolossal* 'colossal' *otrolig* 'incredible' *oerhörd* 'tremendous'

Table 32: Adjectives that frequently modify SITUATIONS.

Enorm and *kolossal* are present in both subgroups here. They have also been mentioned in Sections 8.2.5.4.1 (*colossal* only) and 8.2.5.4.2.

8.2.5.4.4 Adjectives that frequently modify HUMAN head nouns

The rather special group of adjectives that frequently modify HUMAN head nouns consists of six words.

framstående 'prominent'
förnäm 'distinguished'
inflytelserik 'influential'
mäktig 'powerful, immense'
ryktbar 'famous'
storväxt 'tall'

Table 33: Adjectives that frequently modify HUMANS.

There are two aspects of the size of a HUMAN that can be modified, cf. the section on OBJECTS. One is the physical size of a person, e.g. *storväxt* 'tall' and *högrest*[28] 'tall'. The other is the more abstract side referring to the mental strength of a person, e.g. *en stark kvinna* 'a strong woman', or a person's power, e.g. *en inflytelserik man* 'an influential man', rather than the physical size. Some adjectives are used in both meanings, e.g. *stark*,[29] which when modifying female humans, e.g. *en stark kvinna* 'a strong woman', prototypically refers to her mental strength, while *en stark man* 'a strong man' says something about his

[28] *högrest* 'tall' is not frequent enough to be included in the test set.

[29] *stark* 'strong' was excluded from the test set; see above.

physical strength or his power. As regards the modification of HUMAN nouns, there seems to be an important dichotomy between abstract and physical properties.

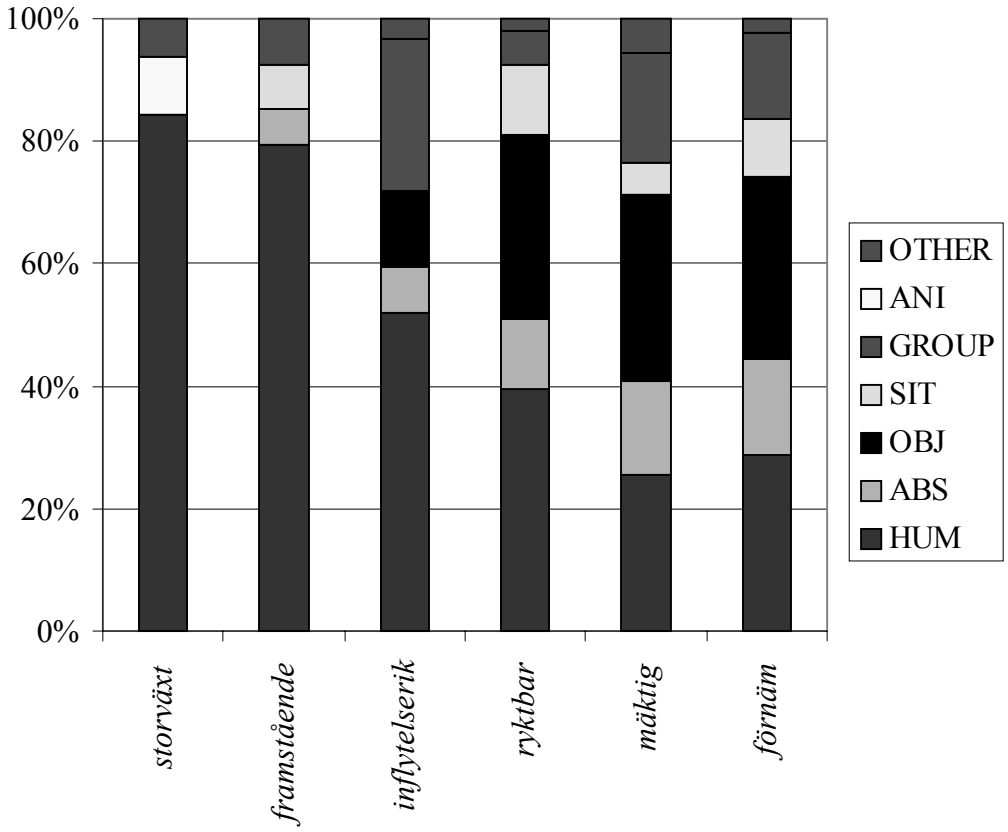


*Figure 47: The semantic ranges of synonyms of **stor** frequently modifying HUMANS.*

Storväxt 'tall' (literally 'large-grown'), which modifies only the physical properties of HUMAN head nouns, has a very large portion of its semantic range in the category HUMAN. The other categories modified by *storväxt* are ANIMAL, OBJECT, and ABSTRACTION, e.g. *en storväxt stridshingst* 'a big war-stallion', *en storväxt palm* 'a tall palm tree', *storväxt manlighet* 'tall masculinity'. However, unlike the words with an abstract profile in this group of words that most often modify HUMANS, *storväxt* does not modify SITUATIONS or GROUPS. The second-most common type of noun modified by *storväxt* is ANIMAL.

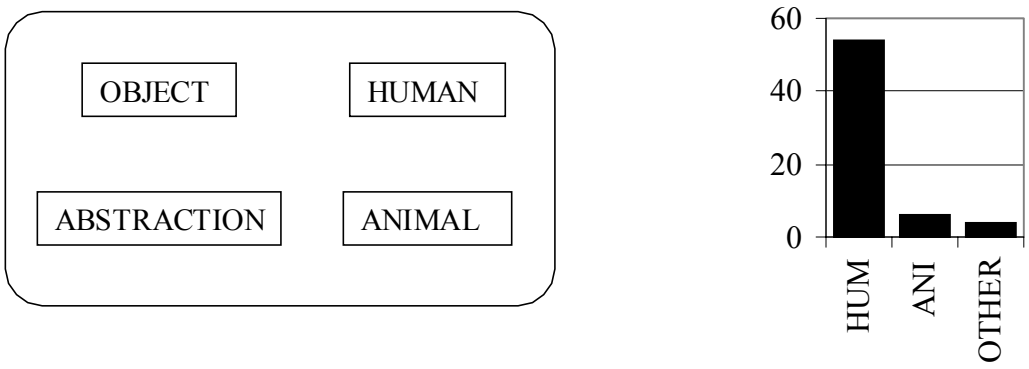


*Figure 48: The semantic range of **storväxt** 'tall'.*

The words in this group that modify only abstract properties are found to modify, apart from the major categories (HUMAN, ABSTRACTION, OBJECT, and SITUATION), nouns from the category GROUP, cf. Figure 49 – Figure 51. *Inflytelserik* 'influential' also modifies PSYCHOLOGICAL FEATURES, e.g. *inflytelserikt stöd* 'influential support', cf. Figure 49. A couple of examples of ANIMAL head nouns are found to be modified by *ryktbar* 'famous', e.g. *Ännu ryktbarare har glasögonormen blivit...* 'And even more famous is the Indian cobra…', cf. Figure 51.

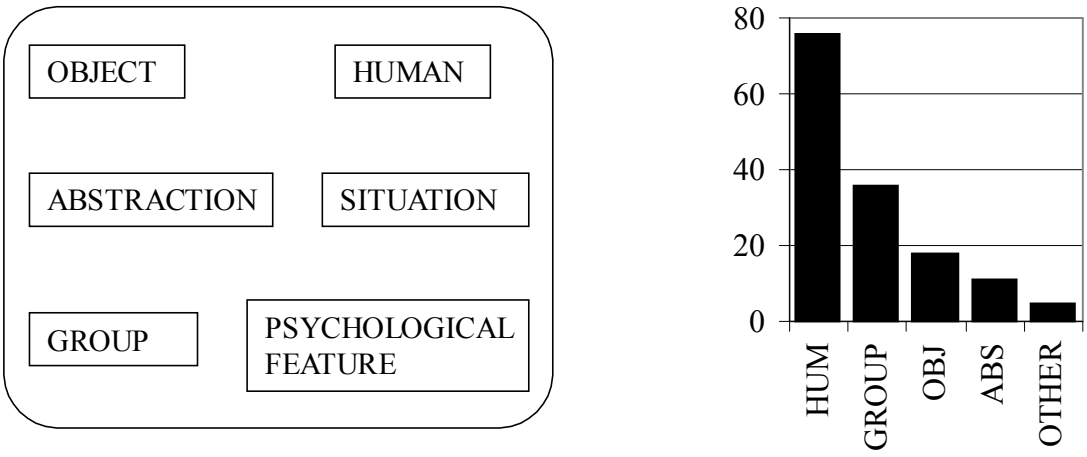


*Figure 49: The semantic range of **inflytelserik** 'influential'.*

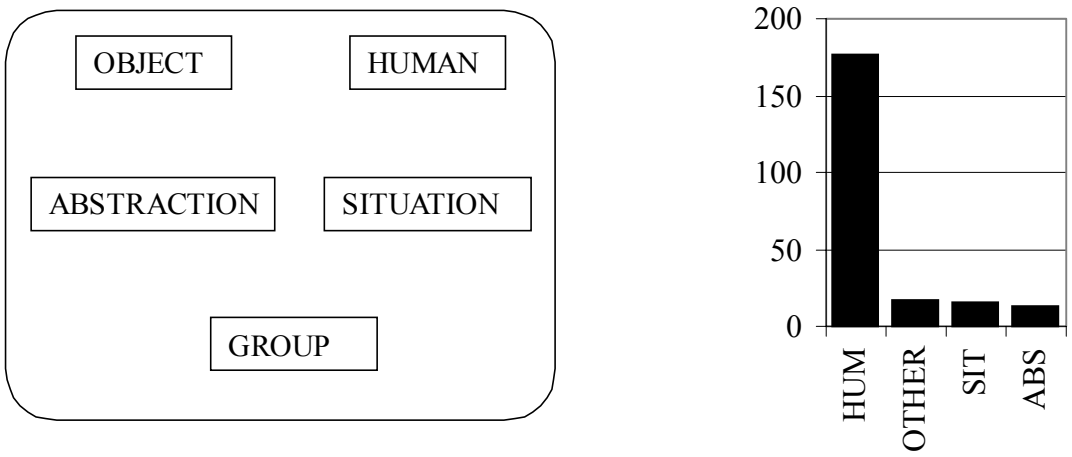


*Figure 50: The semantic range of **framstående** 'prominent'.*

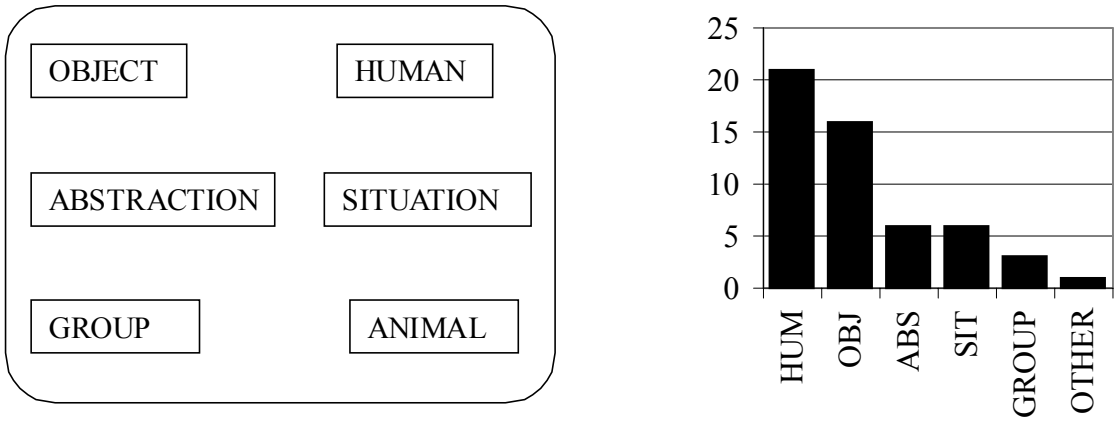


*Figure 51: The semantic range of **ryktbar** 'famous'.*

Grouping the words according to the frequency of the semantic categories modified by them, it turns out that the adjectives that frequently modify HUMAN head nouns also modify OBJECTS, GROUPS, SITUATIONS, and ANIMALS, cf. Table 34. In the descriptions of the semantic ranges of the adjectives above, it was noted that *storväxt* modifies the physical properties of its head-noun referent, whereas the other words in the group – *förnäm*, *mäktig*, *ryktbar*, *inflytelserik*, and *framstående* – modify abstract properties of their referents. The words are thus subclassified as *abstract* or *physical*.

Adjectives that frequently modify HUMANS			
Abstract features			***Physical features***
OBJECTS	***GROUPS***	***SITUATIONS***	***ANIMALS***
förnäm 'distinguished' *mäktig* 'powerful, immense' *ryktbar* 'famous'	*inflytelserik* 'influential'	*framstående* 'prominent'	*storväxt* 'tall'

Table 34: Adjectives that frequently modify HUMANS, grouped according to what type of property they modify and what semantic category they second-most often modify.

Förnäm and *mäktig* are also present in the group of primary OBJECT modifiers.

8.2.5.5 Summary

As shown above, the semantic ranges of the synonyms of *stor* can be divided into four main groups: words that primarily modify OBJECTS, ABSTRACTIONS, SITUATIONS, and HUMANS, respectively. These four main groups are then further divided according to what other semantic categories the synonyms contained in them frequently modify, as summarised in Table 35. Altogether, the test set is divided into eleven groups.

Primary semantic category	***Secondary semantic category***		***Adjective***
OBJECT	***ABSTRACTION***		*innehållsrik* *kolossal* *ofantlig* *riklig* *rymlig* *vidsträckt* *väldig*
	HUMAN		*förnäm* *kolossal* *mäktig* *ädel*
ABSTRACTION	***OBJECT***		*diger* *enorm* *kolossal*
	MONETARY REPRESENTATION		*ansenlig* *dryg*
	SITUATION		*avsevärd* *enorm* *kolossal* *oerhörd* *otrolig*
SITUATION	***OBJECT***		*kolossal* *enorm* *otrolig*
	ABSTRACTION		*avsevärd* *grov* *kolossal* *kraftig* *oerhörd*
HUMAN	***Physical***	***ANIMAL***	*storväxt*
	Abstract	***OBJECT***	*förnäm* *mäktig* *ryktbar*
		SITUATION	*framstående*
		GROUP	*inflytelserik*

*Table 35: Grouping of the synonyms of **stor** according to their semantic ranges.*

The results shown in Table 35 give an indication of how words can be organised according to the frequency of the semantic categories modified. We can discern several meanings for the various groups. In the bottom group, adjectives that primarily modify HUMANs, there is a clear distinction between the modification of concrete qualities, i.e. physical ones, such as by *storväxt*, and that of abstract qualities saying something about the importance of a person, by e.g. *förnäm*, *mäktig*, *ryktbar*, *framstående*, and *inflytelserik*. It has been noted earlier that OBJECTs can be modified either in the sense of concrete dimension or in a more abstract sense. In the group of words primarily modifying ABSTRACTIONs, two words also frequently modify MONETARY REPRESENTATIONs, namely *ansenlig* and *dryg*. These adjectives modify their referent as to quantity rather than as to concrete size. The quantity is countable when it concerns money; however, there are many cases of quantities that are not countable or measurable, e.g. *skillnad* 'difference', *allvar* 'seriousness', and *frihet* 'freedom'.

Grouping words according to the semantic categories they frequently modify gives a hint about how they cluster in synsets. But how are the synsets to be connected to the various meanings of *stor*? The next study will be an investigation of the various meanings of *stor* and their semantic ranges.

8.2.6 *A closer look at the polysemy of* stor

8.2.6.1 Introduction

Stor is a highly frequent word with many meanings. The entry for *stor* is one of the longest in the SAOB, and five different meanings are listed for *stor* in the NEO: (1) having a spatial extension above average, (2) having a non-spatial scope above average, (3) prominent (of HUMANs), (4) important (of SITUATIONs), and (5) valuable (of playing cards).[30] In this section, the semantic ranges for four meanings of *stor* will be investigated and described: *concrete dimension*, *countable quantity*, *uncountable quantity*, and *importance.*

In the studies above, the semantic ranges of *stor* and its synonyms have been studied from the viewpoint of the semantic type of the head words that they modify. There are, as previously mentioned, several possible meanings for the dimensional adjectives studied. The core meaning of *stor* is strictly dimensional: it says something about an object's extension in space. However, sometimes more abstract phenomena are modified as to dimension, and in these cases it is clear that what is intended is not extension in space but rather extension in some

[30] No examples of meaning (5) in the NEO were found in the corpora studied and this meaning will be disregarded in the further study.

abstract dimension. The following categories were used to classify the meanings of *stor* found in the material from Parole.

I Concrete dimension

II Countable quantity

III Uncountable quantity

IV Importance

The first meaning, *concrete dimension*, which can be viewed as the core meaning of *stor*, is the main meaning when *stor* modifies OBJECTS, e.g. *en stor boll* 'a large ball'. The second meaning, *countable quantity*, concerns size in non-spatial dimensions, e.g. *ett stort antal lägenheter* 'a large number of apartments', *de stora barnkullarna* 'the large cohorts of children', *ett stort avstånd* 'a large distance'. These three examples share one feature, the measurability or countability of the quantity, but there are many examples where this feature is not present: *det stora äventyret* 'the great adventure', *med stort allvar* 'with great seriousness', *ett stort ansvar* 'a big responsibility'. For these so-called mass nouns, some abstract dimension is unquantifiably larger than expected or average. These words belong to the category of *uncountable quantity*.

In many of the cases where dimensional adjectives are used together with HUMAN head nouns, the feature of being well-known is modified rather than the physical size of the person. The fourth category, *importance*, covers these examples, e.g. *den store arkitekten Frank Lloyd Wright* 'the prominent architect Frank Lloyd Wright', *antikens stora kvinnor* 'the great women of Antiquity'. This meaning of *stor* corresponds to meaning (3) in the NEO, that of *prominence*. The basic difference between meanings (3) and (4) – prominence and importance – concerns semantic range, i.e. the prominence meaning is used about HUMANS whereas the importance meaning is used about OBJECTS and SITUATIONS. However, I see no difference in the semantic content of the two meanings distinguished by the NEO, and so I will merge them in one meaning, namely *importance*. The importance meaning of *stor* is also found where *stor* modifies OBJECTS, ABSTRACTIONS, and SITUATIONS, e.g. *det stora Ryssland* 'prominent Russia',[31] *Piratens stora dag* 'Piraten's big day', and *kyrkoårets stora fest* 'the great feast of the ecclesiastical year'.

[31] In context, this example appears to refer to the important role played by Russia in international politics rather than to the country's large size.

As test data for the study of the polysemy of *stor*, I used the same set of sentences containing forms of *stor* as for the previous study, i.e. slightly over 1,000 sentences containing any form of the lemma *stor* in the Parole corpus, chosen at random.

8.2.6.2 When the head is not what is modified

Through the further study of the use of *stor*, it was found that in some cases what was modified was not really a feature of the head noun, but rather some other implicit element.

*... den **stora** exportören av spelare till elitserien...*

'… the *big* exporter of players to the first division…'

*... USA är Unicefs **största** bidragsgivare...*

'…the US is UNICEF's *largest* contributor…'

*... Kambodja och Mocambique […] är **stora** mottagare av svenskt minröjningsbistånd.*

'…Cambodia and Mozambique […] are *big* receivers of Swedish mine-disarmament aid.'

*...en av de **större** ungdomstidningarna...*

'… one of the *bigger* youth magazines …'

*...de **stora** genomfartsgatorna...*

'…the *large* through roads…'

In the first example, *den stora exportören* 'the big exporter' is not necessarily a big organisation; what is focused upon is the number of players it exports. In the second example, *Unicefs största bidragsgivare* 'UNICEF's largest contributor' what is big is not the contributor, but the contribution. The size of the contribution is what is modified in the third example too: *stora mottagare* 'big receivers'. In the next example, *de större ungdomstidningarna* 'the bigger youth magazines', the print run rather than the physical size of the magazine is modified. In the last example*, de stora genomfartsgatorna* 'the large through roads', the roads in question may be large, but what is more important is that they are heavily trafficked.

Åkermalm (1979:59ff) calls this type of construction, where an adjective modifies the first part of a composite expression, *förledsbestämning* 'first-constituent modification', e.g. *en personlig lägesprecisering* '*lit.* a personal

position-specification', *varm korvgubbe* '*lit.* hot dog-man' – intended meanings 'personal-position specification' and 'hot-dog man'. Modification of this type is also called *förledsanslutning* in the Scandinavian literature; for further reference, see Wellander 1954:236ff and 1939:580ff, Jespersen 1891:261ff, and Åkermalm 1965:106ff.

The cases where the referent is implicit, e.g. *stor exportör* 'big exporter', *stora mottagare* 'big receivers', cannot really be classified as first-constituent modification, but nevertheless represent a related phenomenon and will be treated in the same way.

The words in this group – which are syntactically modified by *stor* but where the semantic interpretation should be directed towards some other implicit or explicit noun – have been included in the study and categorised according to the noun implicitly modified. Twenty examples (about 2%) in the material studied turned out to display this feature, and in all cases it was explicitly noted in the database that the head of the noun phrase is not what is actually modified.

Metaphors also cause problems, e.g.

Därför reagerar jag mot att yrkesfisket framställs som den ***stora*** *busen.*

'That is why I react against professional fishing being portrayed as the *big* bad guy.'

Den stora busen 'the big bad guy' is an idiom, where the head noun, *busen* 'bad guy', is HUMAN, but in this case it is used metaphorically about an ACTIVITY, *yrkesfisket* 'professional fishing'. The target of the metaphor will not be taken into consideration in this classification system, even though it would be an interesting subject of research to make a deep analysis of adjective-containing metaphors and their targets.

8.2.6.3 Results

The graph below shows how the meanings of *stor* are distributed across the semantic categories of the head nouns.

It can be noted that the most frequent meaning of *stor* is *uncountable quantity*, followed by *countable quantity*. This is remarkable since the core meaning, *concrete dimension*, was expected to be the most frequent one.

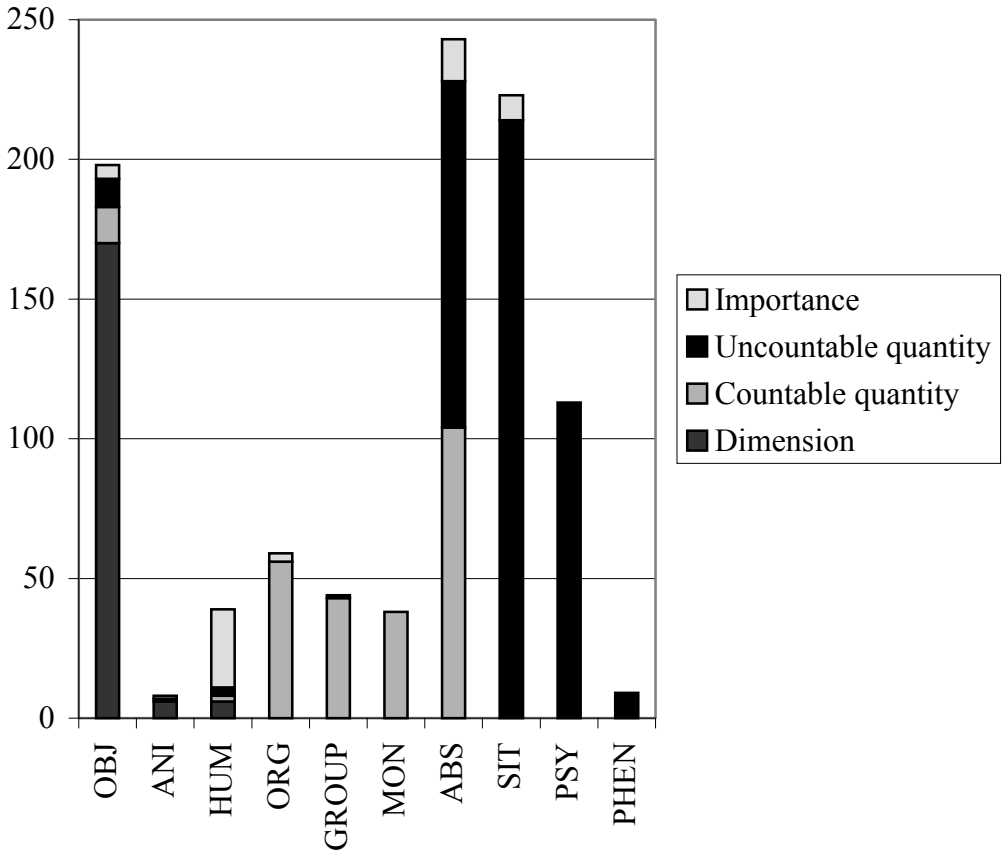


*Figure 52: Meanings of **stor** distributed across the semantic categories of the head nouns.*

The material divides nicely into three major groups. ANIMALS and OBJECTS are most often modified by *stor* in the *dimensional sense*. When modifying ORGANISATIONS and GROUPS, *stor* almost always has the sense of *countable quantity*. For MONETARY REPRESENTATIONS, *countable quantity* is the only meaning used in adjective-noun constructions with *stor*. When *stor* means 'larger than average in *uncountable quantity*' the noun heads are PHENOMENA, PSYCHOLOGICAL FEATURES, and SITUATIONS. A large portion of the ABSTRACTIONS, about half, are modified by *stor* in the sense of *uncountable quantity*. The other half of the ABSTRACTIONS modified by *stor* are modified in the sense of *countable quantity*; there are also a few cases of *importance*.

It is striking that three of the semantic categories of the head nouns are used with a single meaning of *stor*: PHENOMENA and PSYCHOLOGICAL FEATURES with *stor* in the sense of *uncountable quantity*, e.g. *den stora jordbävningen i Kobe* 'the great earthquake in Kobe', *med stort allvar* 'with great seriousness', and MONETARY REPRESENTATIONS with *stor* in the sense of *countable quantity*, e.g. *stor avgift* 'large fee', *stor vinst* 'large profit'.

The other semantic categories also show distinct patterns as to what property within the concept is most often modified. This is true for OBJECTS, ANIMALS, HUMANS, ORGANISATIONS, GROUPS, and SITUATIONS. The ABSTRACTIONS, however, fall into two major groups, one where *stor* modifies a countable quantity and one where it modifies an uncountable quantity.

The results can be qualitatively summarised as follows.

Semantic category	***Concrete dimension***	***Importance***	***Uncountable quantity***	***Countable quantity***
OBJ	X	X	X	X
HUM	X	X	X	X
ANI	X	X		
ABS		X	X	X
SIT		X	X	
ORG		X		X
GROUP			X	X
PSY			X	
PHEN			X	
MON				X

*Table 36: Semantic categories modified by the various meanings of **stor**.*

The core meaning of *stor* is the meaning with the heaviest restrictions on its semantic range – only nouns from three semantic categories are modified by *stor* in the *concrete-dimension* meaning: OBJECT, HUMAN, and ANIMAL. The extension to *importance* then allows another three semantic categories: ABSTRACTION, SITUATION, and ORGANISATION. The two rightmost columns show that the semantic range is very large in the quantitative meanings. In the cases where words from a semantic category cannot be modified by *stor* in the *uncountable-quantity* meaning, they can be modified by *stor* in the *countable-quantity* meaning, and vice versa – but this is mainly a reflection of the fact that some quantities are countable and others are not.

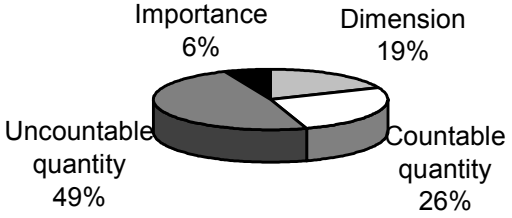


*Figure 53: The distribution of the meanings of **stor** in the SUC.*

It is striking and unexpected that the by far most frequent meaning of *stor* is that of *uncountable quantity*, cf. Figure 53. It is usually expected that the core meaning is also the meaning most frequently used. Is *concrete dimension* really

the core meaning of *stor*? Intuitively, I believe that this is the case, but the raw frequency data would seem to indicate that the core meaning is *uncountable quantity*.

8.2.7 Children's use of stor

The core meaning of a word is typically considered to be the meaning most frequently used as well as the first meaning to be acquired. If the core meaning of *stor* is *uncountable quantity*, children's use of *stor* should also primarily concern this meaning rather than that of *concrete dimension*. This is, in my view, highly unlikely, and therefore I undertook a study of children's early use of *stor*. The data below are based on a corpus of child language. *Stor* will be classified according to meaning, and the semantic range or ranges will be described as well.

Expressions concerning size are basic elements of language. Wierzbicka (1992, 1997) claims that there are certain classes of meaning to which human beings are innately tuned and for which they actively search. In her lists of universal semantic primitives, *big* and *small* are found under "attributes". Dimensional adjectives such as *big* and *small* are indeed words that children acquire early in their development, but according to Wierzbicka the child possesses a primitive form of protolinguistic representation of the concepts even before learning the words.

A study of children's use of *stor* was made using a spoken-language corpus (Richthoff 2000, Strömqvist et al. forthcoming). The data were collected from five Swedish children. Only correct pronunciations of *stor* were included in the study – early variants such as *ko* and *toj* were disregarded. The children's production of *stor* spans the age period of 22 months to 48 months.[32]

Bella:	30 months and 17 days – 41 months and 9 days
Harry:	33 months – 47 months and 20 days
Marcus:	22 months and 14 days – 33 months and 29 days
Tea:	29 months and 12 days – 47 months and 23 days

Of the 137 examples of *stor*, all are in the *concrete-dimension* meaning. Most of the modified nouns are OBJECTS; the rest are HUMANS or ANIMALS, cf. Figure 54.

[32] One of the children, Anton, does not produce any correct forms of *stor* during the early recordings and is thus not included.

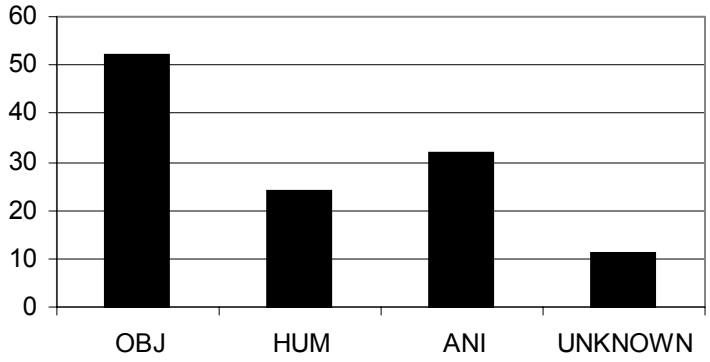


*Figure 54: Distribution of semantic categories modified by **stor** in a child-language corpus.*

Some examples of what the children produce are:

Bella, 36 months and 29 days: *Ja a-x en **stor** handduk.*
'I have a *big* towel'

Bella, 38 months and 19 days: *Han få-x ha en **stor** tallrik pojken.*
'He can have a *large* plate the boy'

Harry, 46 months and 23 days: *En **stor** val. En sån **stor** val som våran loge.*
'A *large* whale. Such a *large* whale as our barn'

Tea, 47 months and 23 days: *Nä, ja ä **stort** barn.*
'No, I am *big* child'

Only three types of nouns were found as head nouns of *stor* in the material: OBJECTS, HUMANS, and ANIMALS. In the examples categorised as "unknown", it was not possible to tell what the head noun was, like in the examples below.

*Hur lå **stor** ä den*
'How ? *big* is it'

*Vicken **stor**!*
'What a *big* one!'

The children's use of *stor* is exclusively in the sense of *concrete dimension*. Preliminary studies of child-directed adult speech also show that this is the predominant meaning of *stor* heard by children. Most of the examples of *stor* in the child-directed speech were in the *concrete-dimension* sense, e.g. *den stora väderkvarnen* 'the big windmill'. Only a few examples did not belong to the *concrete-dimension* type, but rather to that of *quantity*, and they all concerned the amount of food on the spoon when eating, e.g. *å en sån stor sked som kom in nu* 'and such a big spoon that came in now'.

It is quite obvious that the first meaning of *stor* acquired by the child is the one expressing *concrete dimension*. This is the only meaning of *stor* that children use and it is also the absolutely most frequent meaning in the input they receive from adults. But as we saw in the results in the previous section, the most common use of *stor* in Parole is not as a modifier of concrete OBJECTs but as a modifier of ABSTRACTIONs and SITUATIONs, expressing *uncountable quantity*. This may be due to the types of texts that constitute Parole, but it is likely that the quantity meanings of *stor* are more frequent in corpora of adult speech than in corpora of children's speech. At some point in the child's development, the core meaning is extended to the quantity meanings and to the meaning of *importance*. This will be discussed in the following section.

8.2.8 The polysemy of **stor** *in a cognitive semantic model*

The body is used as a central referent when describing size in many languages. Speakers of Jahai often use the body as "measuring rod" when expressing size. The size of various objects is often referred to by grasping a body part of relevant size and saying "this size" (Burenhult 2000).[33] In Indo-European languages it is common for units of measurement to be related to parts of the body, e.g. *tum* 'inch, *lit.* thumb', *fot* 'foot' (Swedish). These are interesting parallels to the view that dimensional attribution is a basic concept originating from the human body, with the human body as reference. The previous study of *stor* in child language clearly shows that the first meaning of *stor* acquired is the *dimensional* meaning. The question is how the core meaning of *stor*, *concrete dimension*, is transferred to the other domains: the meanings of *importance*, *countable quantity*, and *uncountable quantity*. The explanation that follows uses a cognitive semantic framework.

Lakoff & Johnson (1999:50) explain the extensions of the central meaning of *big*. In accordance with Wierzbicka, they use size as a basic sensorimotor domain. It is a constituent in one of the primary metaphors: *Important Is Big*.

Important Is Big

Subjective judgement: Importance

Sensorimotor Domain: Size

Example: "Tomorrow is a big day."

Primary Experience: As a child, finding that big things, e.g., parents, are important and can exert major forces on you and dominate your visual experience.

[33] Jahai belongs to the Aslian branch of the Mon-Khmer language family and is spoken in northern Peninsular Malaysia.

This metaphor explains the mapping from *stor* in the sense of *concrete dimension* to the meaning of *importance*.

Examples of importance:

*Piratens **stora** dag*

'Piraten's *big* day'

*de tre **stora** händelserna 1905*

'the three *big* events of 1905'

*den **stora** nyheten*

'the *big* news'

*Pär Lagerkvists **stora** roman Dvärgen*

'Pär Lagerkvist's *great* novel The Dwarf'

*1900-talslyrikens **stora** antikkännare, Rabbe Enckell*

the *great* classical scholar of 20th century lyrics, Rabbe Enckell

*en av landets **största** lokalrevyer*[34]

one of the country's *largest* local vaudevilles

The two remaining meanings both deal with quantities of ABSTRACTIONS. Countable quantities have an extension that can be measured in some way, e.g. *stora avgifter* 'large fees', *de stora barnkullarna* 'the large cohorts of children'. They share some features with the measuring-rod metaphor. Lakoff's metaphor *Ideas Are Objects* accounts for the transfer of meaning for both countable and uncountable quantities.

Uncountable quantities are viewed as the unmeasurable content of a container, e.g. *stort allvar* 'great seriousness', *stort besvär* 'great trouble'.

The reasoning above is summarised in Figure 55.

[34] It is difficult to tell exactly what is "large" here – the ensemble, the number of people having seen the show, the amount of money the production has cost, or the "cultural importance" of the vaudeville.

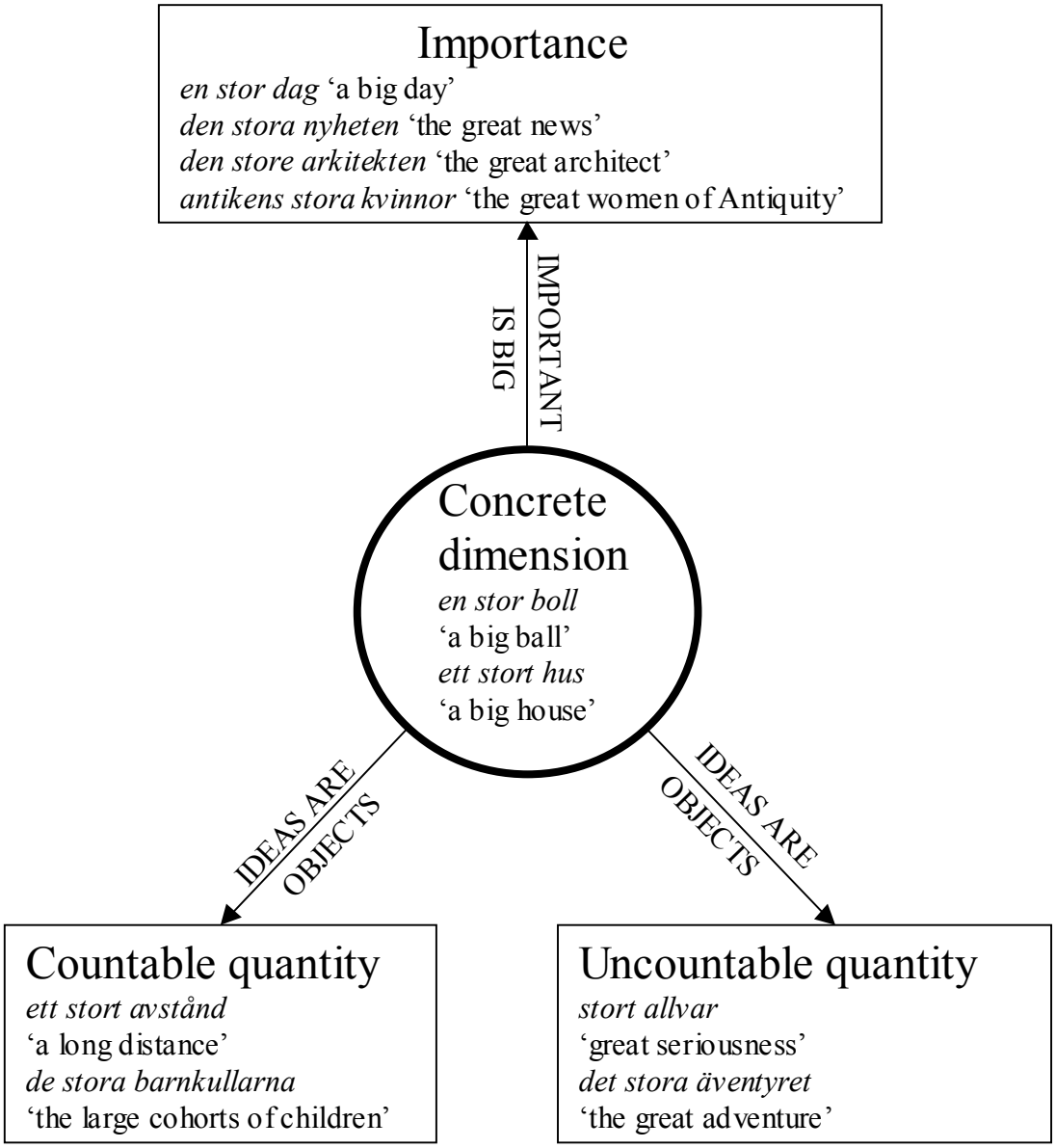


*Figure 55: Metaphoric extensions of **stor**.*

All the transitions share the property of transferring features from the concrete reality to a more abstract level. In the suggested model, the core meaning of *stor* learned by the small child is extended to the meanings of *importance*, *countable quantity*, and *uncountable quantity* through the metaphors *Important Is Big* and *Ideas Are Objects*.

8.3 Extending SWordNet: Connecting adjectives with nouns

Four meanings of *stor* have been distinguished in the studies above. The core meaning of *stor* is *concrete dimension*, but there will be separate entries to account for the each of the meanings of *stor* in SWordNet. It has been shown

that synonyms can be grouped according to semantic range and this will also be used when coding the entries in SWordNet.

Using the results from the study of the semantic ranges of the synonyms of *stor* (see Table 35), the relations between the synonyms can be graphed as in Figure 56.

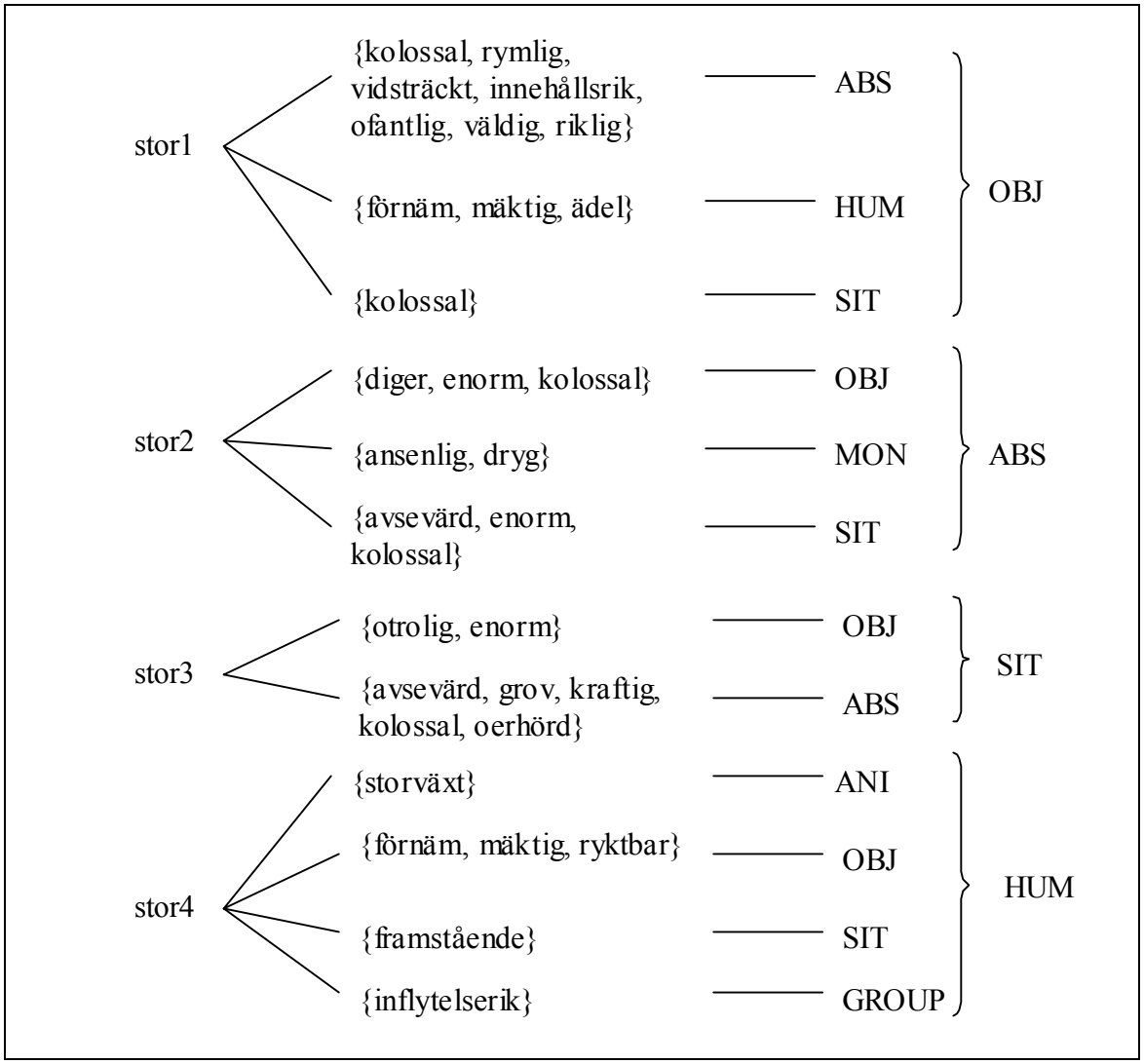


*Figure 56: The synonyms of **stor** organised in synsets around **stor**. The grouping is based on the results of the study of semantic range.*

The four meanings of *stor* are in turn linked to *liten* 'small' with antonym links, cf. Figure 4.

The organisation above is based on how often each adjective modifies nouns belonging to the different semantic categories. This is as far as semi-automatic methods can get us. However, there are several words that one would intuitively prefer to group differently.

Four basic meanings of *stor* were distinguished in Section 8.2: *concrete dimension*, *countable quantity*, *uncountable quantity*, and *importance*. I will let each of these meanings of *stor* be part of a direct-antonym pair heading a number of synsets.

Three of the synsets in Figure 56 are synonyms of *stor* in the *concrete-dimension* sense. The most obvious one is {storväxt}; the other two are {kolossal, rymlig, vidsträckt, innehållsrik, ofantlig, väldig, riklig} and {avsevärd, grov, kraftig, kolossal, oerhörd}. *Innehållsrik* and *riklig* rather modify quantities, and should not appear in this synset. The adjectives used in the *concrete-dimension* sense modify OBJECTs, HUMANs, and ANIMALs, cf. Table 36.

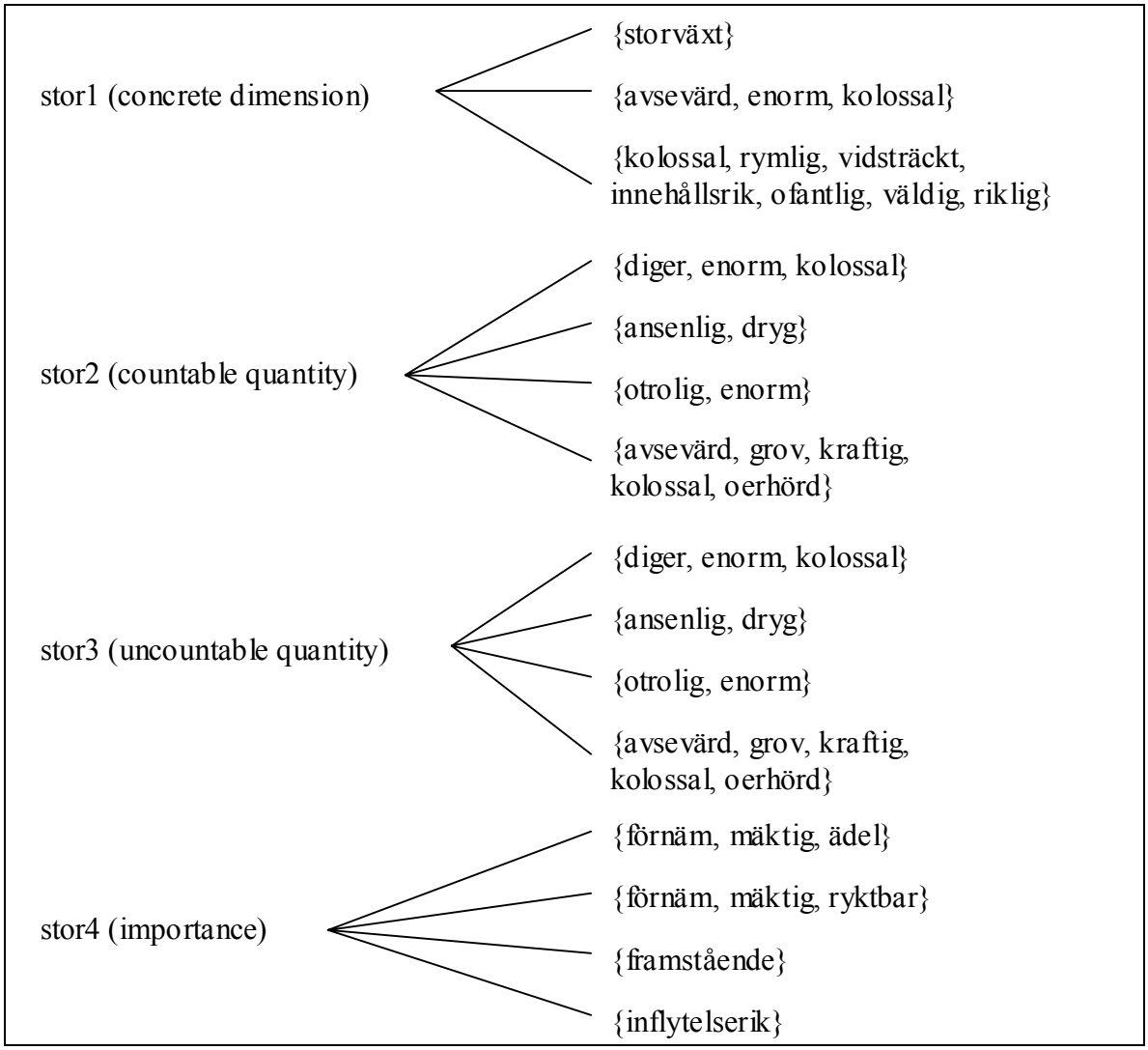


*Figure 57: The synonyms of **stor** organised in synsets around **stor**: a development of Figure 56.*

Another three synsets are synonymous with *stor* in the *importance* sense: {förnäm, mäktig, ryktbar}, {framstående}, and {inflytelserik}. These adjectives

are used with OBJECTS, HUMANS, ANIMALS, ABSTRACTIONS, SITUATIONS, and ORGANISATIONS, cf. Table 36.

The remaining synsets have to do with quantities: {diger, enorm, kolossal}, {ansenlig, dryg}, {avsevärd, enorm, kolossal}, {otrolig, enorm}, and {innehållsrik, riklig}. Whether the quantity is countable or not depends on the referent. These words modify nouns of all semantic categories except ANIMAL.

Each group is considered to have one meaning, and they are numbered 1 through 4: *stor1*, *stor2*, *stor3*, and *stor4*.

This is still a much simplified model; most of the adjectives really should occur in several synsets linked to several of the meanings of *stor*. However, I will not bring the analysis any further, since the point of the study was to show how frequency studies could facilitate the organisation of synonyms.

In a lexical database such as WordNet it is also possible to code collocational information as a relation between an adjective and a noun by connecting the lexical entry for an adjective to the lexical entries for the nouns it can modify. Guerreiro (2000) suggests the coding of such relations in the Portuguese WordNet, distinguishing two types of relations between adjectives and nouns: collocational expressions for which simple compositional processing fails are called "xenonymic", e.g. (Portuguese) *comentário azedo* 'acid remark', *sorriso estampo* 'wide smile', whereas expressions whose interpretation is predictable are called "philonymic", e.g. *peixe fresco* 'fresh fish'.

From each of the words in the structure, semantic-range links will be added to the noun hierarchies. For example *storväxt* 'tall', which has a limited semantic range, will have links to the top nodes for HUMANS, ABSTRACTIONS, OBJECTS, and ANIMALS, as shown in the figure below. This link can also be marked with information about a semantic pattern such as described by Warren, e.g. the link between *storväxt* and HUMAN can be marked FEATURE–WHOLE, following Warren's model. This type of implementation will facilitate the interpretation of the adjective-noun unit, pointing out what feature of the noun is actually brought forward by the adjective.

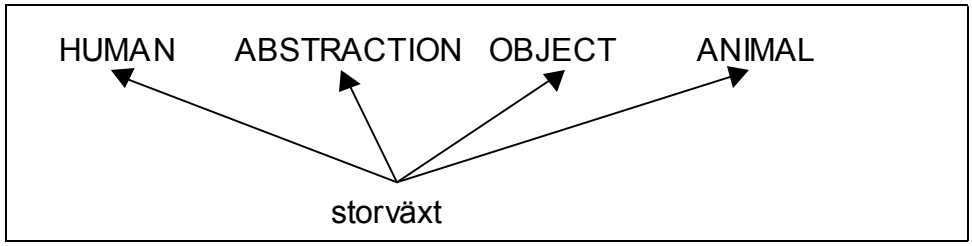


*Figure 58: Connection from the lexical entry for **storväxt** 'tall' to the noun hierarchies reflecting its semantic range.*

The links between the top nodes of the noun hierarchies and the words in the adjective net are recursively inherited to all the hyponyms of the top nodes.

The primary aim of the model described is to be useful for applications of language technology, e.g. for text interpretation and text generation. A system that makes use of the information about what properties the adjectives brings out in the noun has yet to be built. With human users of the electronic lexicon in mind, the set of relations between adjectives and nouns according to Warren (1984) has to be rephrased so that the information can be understood by non-linguists as well.

8.4 Chapter summary

The semantic range of an adjective is reflected in the types of nouns it can modify. The first study of the co-occurrence between adjectives and nouns focused on the scale *full—tom* 'full—empty', comparing the semantic ranges of the two words. My hypothesis was that the semantic ranges of the words diverge significantly and that this may explain why, unlike other direct antonyms, these two words do not co-occur sententially significantly often.

As expected, three different meanings of *full* were distinguished in the corpus study: the *container* meaning, the *rod* meaning, and a special case of the *container* meaning, *intoxication*. The semantic range of *tom* was less complicated, covering only the *container* meaning.

The semantic ranges of the words overlapped in five categories of the *container* meaning. It is obvious that only *full* can be used in the *rod* meaning. Moreover, it is not possible to use *tom* in the *intoxication* meaning. This means that the semantic ranges of the two words diverge considerably. Furthermore, it has been stated that both *full* and *tom* have more than one antonym. This is probably the reason why the frequency of sentential co-occurrence for the two words was not statistically significant.

If it is the case that the antonymy relation between two words is learned because these words co-occur sententially significantly more often than other word pairs on the same scale, the results of the collocational study show that the antonym relation between *full* and *tom* would then be difficult to acquire. The findings of this study show that the semantic ranges of the words are quite diverse; they overlap only in the *container* meaning.

Other words involved in the same semantic field as *full—tom* are *nykter* 'sober', *inget* 'nothing', and *fylld* 'filled'. An extensive study of the semantic ranges of these words and their synonyms, followed by a comparative study, should make it possible to distinguish them from each other and to identify the lexical relations between them.

The second part of the chapter is a study of *stor* and 28 synonyms of *stor*. The semantic ranges of the words were described based on data from the Parole corpus, and the words were then grouped according to what type or types of noun they frequently modify. Four meanings of *stor* were distinguished: *concrete dimension*, *importance*, *countable quantity,* and *uncountable quantity*. This grouping was used to organise the synonyms in the Swedish WordNet.

Usually, you expect the central meaning to be the meaning most frequently used. In this case, the central meaning of *stor*, *concrete dimension*, was expected to be predominant, but both *countable quantity* and *uncountable quantity* were more frequent in the material studied. Is *concrete dimension* not the prototypical meaning of *stor*?

It was considered highly likely that the first meaning of *stor* acquired is that of *concrete dimension*. A small study of four children's use of *stor* showed that this assumption is correct. In the age span of 22 months to 48 months, the children used *stor* exclusively in the *concrete-dimension* meaning. This indicates that the prototypical meaning of *stor* is indeed *concrete dimension*, though this is not the most common meaning found in corpora – at least not in the Parole corpus.

An attempt to explain the extensions of the central meaning of *stor* to the meanings *importance*, *countable quantity,* and *uncountable quantity* was presented within a cognitive semantic model. The metaphor *important is big* accounts for the extension from *concrete dimension* to *importance*, while the metaphor *ideas are objects* maps the central meaning to *countable* and *uncountable quantities*.

The results of the studies of *stor* and its synonyms are used to organise the adjectives in the Swedish WordNet; the grouping of the words according to what semantic categories they frequently modify is useful, but has to be interpreted with care. In addition, the semantic range of an adjective is coded in SWordNet by connecting it to the noun hierarchies it can modify.

Part IV: Conclusions

9 Conclusions

Corpora and methods to extract lexical information from corpora are central issues throughout this book. Antonymy and the semantic ranges of adjectives have been in focus. What types of lexical information can be extracted from corpora and how?

It has been suggested that the acquisition of antonym relations is cued by the high frequency of sentential co-occurrence of antonymous words. This idea is formulated in the co-occurrence hypothesis: "Two adjectives are learned as direct antonyms because they occur together in the same sentences more frequently than chance would allow". Empirical data in support of the co-occurrence hypothesis have previously been presented for English, and in this book the studies of co-occurrence of direct antonyms show that they co-occur more often than chance would predict in Swedish as well.

There are many reasons for co-occurrence and just as with direct antonyms, indirect antonyms co-occur more often than chance would predict. The co-occurrence hypothesis is reformulated as follows: "Two adjectives are learned as direct antonyms because they occur together in the same sentences more frequently than indirect antonyms on the same scale." A study of twelve different scales of Swedish antonyms supports the rewritten co-occurrence hypothesis. Direct antonyms co-occur sententially significantly more often than the indirect antonyms on the same scales. While the indirect antonyms co-occur overall 1.45 times more often than expected in the Swedish SUC corpus, the direct antonyms co-occur overall 3.12 times more often than expected in the same corpus.

Another cue for learning antonym relations is formulated in the substitutability hypothesis: "Two adjectives are learned as direct antonyms because they are interchangeable in most contexts, i.e. because any noun phrase that can be modified by one member of the pair can also be modified by the other." This has previously been proved for English and studies in this book show that direct antonyms co-occurring sententially appear most often in parallel contexts.

Frequency of antonym co-occurrence and parallelism of context are cues that may facilitate the acquisition of antonym relations between words. However, though the results of the studies support the co-occurrence hypothesis as well as the substitutability hypothesis, the type of data that they are based on is not common as input to the language learner. These data are written corpora, and

children are not generally the targets. A study of the co-occurrence patterns of direct and indirect antonyms in child-directed adult speech may provide more relevant evidence. The problem is the difficulty of finding enough material, and it is also problematic to choose a unit of study; sentences are not appropriate and the utterance is probably too narrow a unit.

I suggest that a more obvious cue to the antonym relations may be found in the prosodic features of the co-occurring antonyms in a spoken context. Preliminary studies show that co-occurring antonyms are marked with a focal accent in spoken language. The focal accent is a cue to the relation between the words and emphasises the contrast between the antonyms. It is probably a cue that is much easier for the language learner to identify than frequency-based information or syntactic constructions. But it has yet to be shown how antonyms are treated prosodically – this is a highly interesting topic for future research.

Words that co-occur sententially significantly often tend to be lexically related. When I sorted all co-occurring adjectives on rising p-value, I found that the direct antonyms appeared at the top of the list. Lists of adjective pairs that co-occur significantly often are useful tools for lexicographers; in a lexicon such as WordNet such lists can be used to identify a relation, whereupon the lexicographer can determine what type of lexical relation it is and code it in the net.

While the first empirical studies in this book focus on the relation between adjectives, the latter few deal with the relation between adjectives and nouns. The semantic ranges, i.e. the nouns that an adjective can modify, of a couple of antonym pairs that did not co-occur sententially as expected in the antonym studies were explored in detail.

The semantic ranges of *full* 'full' and *tom* 'empty' were described and compared. Three different meanings were distinguished for *full*: the *container* meaning, the *rod* meaning and a special case of the *container* meaning: *intoxication*. *Tom*, however, is used only in the *container* meaning. To a considerable extent, the semantic ranges of the words did not overlap. This indicates that *full* and *tom* are not direct antonyms, as suggested by Lundbladh (1988), or that there are several scales involved. *Full* is also in opposition to *nykter* 'sober' and *inget* 'nothing', while *fylld* 'filled' is often used to contrast with *tom*.

The last study concerns *stor* 'large' and its synonyms. Four meanings were distinguished: *concrete dimension*, *importance*, *countable quantity*, and *uncountable quantity*. Unexpectedly, the most frequent meaning was found not

to be *concrete dimension* – both *countable quantity* and *uncountable quantity* were more common in the material studied. However, a study of the early use of *stor* by children shows that (between 22 and 48 months of age) they use *stor* only in the *concrete-dimension* sense. Preliminary studies show that this is also the most frequent sense used by their parents when talking to the children.

An explanation of the semantic shift of *stor* from the central meaning *concrete dimension* to the meanings *importance*, *countable quantity*, and *uncountable quantity* was suggested within a cognitive semantic framework. The extension to cover the meaning *importance* is bridged with the metaphor *big is important*, while the extensions to the other two meanings, *countable* and *uncountable quantity*, are explained by the metaphor *ideas are objects*.

The semantic ranges of 28 synonyms of *stor* were described and the words were organised according to the semantic categories frequently modified by each adjective. The grouping based on semantic range was used as a basis to distinguish synsets and to code the words in the Swedish WordNet.

The study of semantic range is at present quite laborious owing to the lack of a large semantic lexicon. Tools for manual semantic tagging and for building semantic databases are described in the book: *Klassa* and *Para*. Once more material has been tagged, a larger semantic database will be available to facilitate further research. About 14,000 occurrences of *stor* or a synonym of *stor* have been classified. In each case, the referent has been distinguished as well as its semantic category, and for the occurrences of *stor*, the meaning has been coded. This material may also be useful for the study of other types of phenomena.

Many new research topics have been uncovered during the research presented in this book. The one that I find most intriguing concerns the acquisition of antonym relations, and the role of prosody in this matter. SWordNet has been improved by the addition of links from the adjective lexicon to the noun lexicon; this reflects the semantic range of each adjective and is useful information both to human users of the lexicon and to computer applications, such as translation tasks. But is it psychologically relevant? The Princeton WordNet was originally intended as a model of the mental lexicon, but how do we find out if it is a good model?

The research in this book has raised more questions than it has answered – as is the nature of research.

References

Aarts, J. & P. Calbert. 1979. *Metaphor and non-metaphor. The semantics of adjective-noun combinations*. Tübingen: Max Niemeyer Verlag.

Abelin, Å. 1996. A lexical decision experiment with onomatopoetic, sound symbolic and arbitrary words. *TMH-QPSR* 2/1996. Stockholm.

Åkermalm, Å. 1965. *Rubriksvenska och andra studier*. Falköping: Gleerups.

Åkermalm, Å. 1979. *Modern svenska: språk och stilfrågor*. Falköping: Liber läromedel.

Allan, K. 1977. Classifiers. *Language* 53, pp. 284–310.

Andersson, L.-G. 1984. Avhandlade adjektiv. *Nusvenska studier* 64, pp. 151–179.

Barnbrook, G. 1996. *Language and computers. A practical introduction to the computer analysis of language*. Edinburgh: Edinburgh University Press.

Bendz, G. 1965. Ordpar. *Svenska Humanistiska Förbundets skrifter* 74. Stockholm: P.A. Norstedt & Söners Förlag.

Berry-Rogghe, G. 1973. The computation of collocations and their relevance in lexical studies. A. Aitken, R. Baily, and N. Hamilton-Smith (eds.), *The Computer and Literary Studies*. Edinburgh: Edinburgh University Press.

Biber, D., S. Conrad & R. Reppen. 1998. *Corpus linguistics. Investigating language structure and use*. Cambridge: Cambridge University Press.

Bierwisch, M. & E. Lang (eds.) 1987. *Dimensional adjectives: Grammatical structure and conceptual interpretation*. Berlin: Springer Verlag.

Bolander, M. 1980. Predikativens funktion i svenskan. Om adjektiv som subjektiva predikativ. *Umeå studies in the humanities* 31. Ph.D. diss.

Bolinger. D. 1967. Adjectives in English: Attribution and predication. *Lingua* 18, pp. 1–34.

Burenhult, N. 2000. *Gestures in Jahai: a preliminary report from the field*. Dep. of Linguistics, Lund University. Unpub. field report.

Chan, E. 2000. *Co-occurrence of antonymous adjectives in the British National Corpus*. Dep. of computer & information science, University of Pennsylvania. Unpub. ms.

Charles, W. & G. Miller. 1989. Contexts of antonymous adjectives. *Applied Psycholinguistics* 10:3, pp. 357–375.

Charniak, E. 1993. *Statistical language learning*. Cambridge, MA: The MIT Press.

Clark, H. 1970. Word associations and linguistic theory. J. Lyons (ed.) *New horizons in linguistics*. Harmondsworth, UK: Penguin Books.

Climent, S., H. Rodriguez & J. Gonzalo. 1996. *Definition of the links and subsets for nouns of the EuroWordNet project*. Deliverable number D005 from http://www.hum.uva.nl:80/~ewn/docs.htm

Cruse, D. 1986. *Lexical semantics*. Cambridge: Cambridge University Press.

Cruse, D. 2000. *Meaning in language. An introduction to semantics and pragmatics*. Oxford: Oxford University Press.

Dahllöf, M. 1999. *Språklig betydelse*. Lund: Studentlitteratur.

Denny, J.P. 1976. What are noun classifiers good for? *Chicago Linguistics Society* 12, pp. 122–132.

Denny, J.P. 1979. The "extendedness" variable in classifier semantics: Universal features and cultural variation. M. Mathiot (ed.) *Ethnolinguistics: Boas*, *Sapir*, *and Whorf revisited*, pp. 97–119.

Deese, J. 1965. *The structure of associations in language and thought*. Baltimore: The Johns Hopkins Press.

Dixon, R.M.W. 1977. Where have all the adjectives gone? *Studies in Language*, 1.1:19–80.

Egan, R. 1968. Survey of the history of English synonymy. P. Gove (ed.) *Webster's new dictionary of synonyms*, pp. 5a–31a. Springfield, MA: Merriam-Webster.

Einarsson, J. & T. Hultman. 1984. God morgon pojkar och flickor. Om språk och kön i skolan. Stockholm: Liber.

Ejerhed, E., G. Källgren, O. Wennstedt & M. Åström. 1992. The linguistic annotation system of the Stockholm-Umeå corpus project. *Reports from the department of general linguistics* 33 (DGL-UUM-R-33). Umeå: University of Umeå.

Ejerhed, E. 1996. Design principles for a Swedish corpus annotation system. H. Feldweg & E. Hinrichs (eds.) *Lexicon und Text*, *Lexicographica* 73. Tübingen: Max Niemeyer Verlag.

Fellbaum, C. 1995. Co-occurrence and antonymy. *International Journal of Lexicography* 8:4, 281–303.

Fellbaum, C. 1998. Introduction. Fellbaum, C. (ed.), *WordNet: an electronic lexical database*. Cambridge, MA: The MIT Press.

Fellbaum, C., D. Gross & K. Miller. 1993. Adjectives in WordNet. *International Journal of Lexicography* 3(4), 1990, pp. 235–244. (Also referred to as *Five papers on WordNet*.)

Firth, J. 1957. A synopsis of linguistic theory, 1930–1955. *Studies in linguistic analysis*, pp. 1–32. Oxford: Blackwell.

Frawley, W. 1992. *Linguistic semantics*. Hillsdale, NJ: Lawrence Earlbaum Associates.

Fromkin, V. (ed.). 1973. *Speech errors as linguistic evidence.* The Hague: Mouton.

Fromkin, V. (ed.). 1980. *Errors in linguistic performance: slips of the tongue, ear, pen, and hand*. NY: NY Academic Press.

Givón, T. 1979. *On understanding grammar*. New York: Academic Press.

Givón, T. 1984. *Syntax: a functional-typological introduction*, Vol. 1. Amsterdam: John Benjamins Publishing Company.

Grefenstette, G. 1994. What is a word, What is a sentence? Problems of tokenisation. 3rd conference on computational lexicography and text research. Paper presented at *COMPLEX'*94, Budapest, July 7–10, 1994.

Griswold, R. & M. Griswold. 1983. *The Icon programming language*. Englewood Cliffs, NJ: Prentice-Hall.

Guerreiro, P. 2000. Improving lexical databases with collocational information: data from Portuguese. *LREC 2000 Proceedings*.

Hale, K. 1971. A note on a Walbiri tradition of antonymy. D. Steinberg & L. Jakobovits (eds.) *Semantics. An interdisciplinary reader in philosophy, linguistics and psychology*. Cambridge: Cambridge University Press.

Hellquist, E. 1948. *Svensk etymologisk ordbok*. Lund: Gleerup.

Holmqvist, K. & J. P[illegible]uciennik. 1996. Conceptualised deviations from expected normality: A Semantic comparison between lexical items ending in -ful and -less. *Nordic Journal of Linguistics*, 19, pp. 3–33.

Holtsberg, A. & C. Willners. Statistics for sentential co-occurrence. To appear as *Lund University, Dept. of Linguistics, Working papers* 48.

Hopper, P. & S. Thompson. 1984. The Discourse basis for lexical categories in Universal Grammar. *Language* 60:4, pp. 703–752.

Jespersen, O. 1891. Sproglige kuriosa. *Dania I*, pp. 261ff., reprinted in *Tanker og studier* 1932, pp. 159ff.

Jones, S. 1998. Approaching antonymy afresh. *Liverpool Working Papers in Applied Linguistics*, Vol. 4. No. 1. Department of English language and literature, University of Liverpool.

Justeson, J. & S. Katz. 1991. Co-occurrence of antonymous adjectives and their contexts. *Computational Linguistics* 17:1, pp. 1–19.

Justeson, J. & S. Katz. 1992. Redefining antonymy: the textual structure of a semantic relation. *Literary and Linguistic Computing* 7: pp. 176–184.

Kennedy, G. 1998. *An introduction to corpus linguistics*. London: Longman.

Kjellmer, G. 1984. Some thoughts on collocational distinctiveness. J. Aarts & W. Meijs (eds.) *Corpus linguistics. Recent developments in the use of computer corpora in English language research*. Amsterdam: Rodopi.

Lakoff, G. & M. Johnson. 1999. *Philosophy in the flesh*. New York: Basic Books.

Lakoff, G. 1992. The contemporary theory of metaphor. A. Ortony (ed.) *Metaphor and thought*. Cambridge: Cambridge University Press.

Lehrer, A. 1974. Semantic fields and lexical structure. *North-Holland linguistic series* 11. Amsterdam: North-Holland Publishing Company.

Li, C. & S. Thompson. 1981. *Mandarin Chinese: a functional reference grammar*. Berkeley: University of California Press.

Linell, P. 1982. Speech errors and grammatical planning of utterances. Evidence from Swedish. W. Koch, C. Platzack & G. Tottie (eds.) *Textstrategier i tal och skrift*. Stockholm: Almqvist & Wiksell.

Loman, B. 1956. Om relationen mellan ordföljd och betydelse hos framförställda attributiva bestämningar till substantiviska huvudord. *Arkiv för nordisk filologi* 71.

Lundbladh, C.-E. 1988. Adjektivets komparation i svenskan. En semantisk beskrivning. *Lundastudier i nordisk språkvetenskap* A 40. Lund: Lund University Press.

Lyons, J. 1977. *Semantics,* Vol. I & II. Cambridge: Cambridge University Press.

Malmgren, S.-G. 1990. *Adjektiviska funktioner i svenskan*. Göteborg: Acta Universitatis Gothoburgensis.

McEnery, A. & A. Wilson. 1996. *Corpus linguistics*. Edinburgh: Edinburgh University Press.

Miller, G. & C. Fellbaum. 1992. Semantic networks of English. B. Levin & S. Pinker (eds.) *Lexical & conceptual semantics*. Oxford: Blackwell.

Miller, G., R. Beckwith, C. Fellbaum, D. Gross & K. Miller. 1990. Introduction to WordNet: an on-line lexical database. *International Journal of Lexicography* 3(4), pp. 235–244. (Also referred to as *Five papers on WordNet*.)

Mourelatos, A. 1978. Events, Processes, and States. *Linguistics and Philosophy: an International Journal*, Vol. 2, pp. 415–434.

Muehleisen, V. 1997. *Antonymy and semantic range in English*. Evanston, IL: Northwestern University. Ph.D. diss.

Noreen, A. 1904. *Vårt språk*. Vol. 5. Lund.

Oakes, M.P. 1998. *Statistics for corpus linguistics*. Edinburgh: Edinburgh University Press.

Quirk, R., S. Greenbaum, G. Leech & J. Svartvik. 1985. *A grammar of contemporary English*. London: Longman.

Richthoff, U. 2000. *En svensk barnspråkskorpus. Uppbyggnad och analyser*. Licenciatavhandling i allmän språkvetenskap. Institutionen för lingvistik, Göteborgs Universitet.

Rusiecki, J. 1985. *Adjectives and comparison in English: a semantic study*. London: Longman.

Sigurd, B. 1972. Ord på gott och ont. *Språkvård*, 2, pp. 3–13.

Sigurd, B. 1976. Converse terms in Swedish. N. Enkvist & V. Kohonen (eds.) *Reports on text linguistics: Approaches to word order*. *Meddelanden från Stiftelsen för Åbo Akademi forskningsinstitut* 9. Åbo.

Sinclair, J. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Stroud, C. 1979. Kontrastiv lexikologi. K. Hyltenstam (ed.) *Svenska i invandrarperspektiv*. Lund: Gleerup.

Strömqvist, S., H. Ragnarsdóttir & U. Richthoff. Input and production in the acquisition of function words. To appear in B. Höhle & J. Weissenborn (eds.) *Approaches to bootstrapping: phonological, syntactic and neurophysiological aspects of early language acquisition. Language acquisition and language disorders* Vol. 2. Amsterdam: Benjamins.

Stubbs, M. 1995. Collocations and semantic profiles: on the cause of the trouble with quantitative studies. *Functions of Language*, 2(1), pp. 23–55.

Söderbergh, R. 1964. *Suffixet -mässig i svenskan. En historisk-semantisk ordbildningsstudie*. Stockholm: Almqvist & Wiksell.

Söderpalm, E. 1979. Speech errors in normal and pathological speech. *Travaux de l'Institut de linguistique de Lund* 14. Lund: CWK Gleerup.

Teleman, U. 1972. *Om svenska ord*. Lund: Gleerups.

Teleman, U., S. Hellberg & E. Andersson. 1999. *Svenska Akademiens grammatik*. Stockholm: Norstedts.

Thompson, S. 1988. A discourse approach to the cross-linguistic category 'adjective'. J. Hawkins (ed.) *Explaining language universals*. Oxford: Basil Blackwell.

Warren, B. 1978. Semantic patterns of noun-noun compounds. *Gothenburg Studies in English* 41. Göteborg: Acta Universitatis Gothoburgensis.

Warren, B. 1984. Classifying adjectives. *Gothenburg Studies in English* 56. Göteborg: Acta Universitatis Gothoburgensis.

Wellander, E. 1954. Varm korvgubbe – vandrare i dimma. *Festskrift till Elias Wessén* 15.4, pp. 236ff.

Wellander, E. *Riktig svenska: en handledning i svenska*. Stockholm: Norstedt.

Verkuyl, H. 1993. *A theory of aspectuality*. Cambridge: Cambridge University Press.

Viberg, Å. 1994. Vocabularies. I. Ahlgren & K. Hyltenstam (eds.) *Bilingualism in deaf education*. Hamburg: Signum.

Viberg, Å. 2000. Svenskt OrdNät – Lexikon på dator som modell för ordförrådet i hjärnan hos infödda talare och andraspråkstalare. H. Åhl (ed.*) Svenskan i tiden – verklighet och visioner*, pp. 287–305. Stockholm: HLS Förlag.

Wierzbicka, A. 1986. What is in a noun (or how do nouns differ in meaning from adjectives)? *Studies in language* 10, pp. 353–389.

Wierzbicka, A. 1992. *Semantics, culture and cognition.* Oxford: Oxford University Press.

Wierzbicka, A. 1997. *Understanding cultures through their key words*. Oxford: Oxford University Press.

Willners, C. 1997. Implementing WordNet for Swedish Adjectives. *Lund University, Dept. of Linguistics, Working Papers* 46, pp. 253–268.

Åkermalm, Å. *Sorted under A*.

Lexical resources

Svenska Akademiens ordbok (SAOB). 1893–2001. Lund: Gleerups. See also http://g3.spraakdata.gu.se/saob/, as of March 2001.

Nationalencyklopedins ordbok (NEO). 1995–1996. Höganäs: Bra Böcker.

Strömbergs synonymordbok. 1995. A. Strömberg, Angered: Strömbergs bokförlag.

Oxford advanced learner's dictionary of current English. 1995. J. Crowther (ed.). Oxford: Oxford University Press.

Corpora

The Brown corpus. ICAME, Norwegian computing centre for the humantities. See http://www.hd.uib.no/, as of March 2001.

British National Corpus (BNC). Oxford computing services. See http://info.ox.ac.uk/bnc/ as of March 2001.

The Stockholm-Umeå corpus (SUC 1.0). Dep. of Linguistics, Umeå University, and Dep. of Linguistics, Stockholm University. See http://www.ling.su.se/DaLi/Projects/SUC/Index.html as of March 2001.

The Parole corpus. Språkbanken, Göteborgs universitet. See http://spraakdata.gu.se/lb/parole/ as of March 2001.

Appendix: Analysed adjectives

djup	*grund*
djuplodande	flack
djupsinnig	lättvindig
dov	ytlig
dunkel	
hemlighetsfull	
inträngande	
invecklad	
mättad	
mörk	
obegriplig	
ofattlig	
profund	
svårfattlig	
underfundig	

gammal	*ung*
bedagad	aktuell
beprövad	fashionabel
erfaren	frisk
ålderstigen	fräsch
åldrad	färsk
åldrig	grön
	jungfrulig
	modern
	ny
	nymodig
	obegagnad
	obrukad
	oprövad
	originell
	ovanlig
	sen
	sällsam
	oerfaren
	överraskande

kort	*lång*
avmätt	ankellång
avvisande	avlägsen
flyktig	evinnerlig
fåordig	fjärran
kall	fotsid
knapp	förlängd
kortfattad	golvlång
korthuggen	högväxt
kortvarig	ihållande
kortvuxen	långdragen
kortväxt	långrandig
kylig	långvarig
liten	maxi
oansenlig	omständlig
ovänlig	reslig
småväxt	stor
snabb	utdragen
	utsträckt
	vid
	vidlyftig
	vidsträckt

långsam	*snabb*
bekväm	avmätt
dryg	avvisande
dröjande	flyktig
ledsam	fåordig
loj	kall
långrandig	knapp
långvarig	kort
maklig	kortfattad
sakta	korthuggen
senfärdig	kortvarig
släpig	kortvuxen
sölig	kortväxt
tråkig	kylig
trög	liten
tungrodd	oansenlig
utdragen	ovänlig
	småväxt

hög	*låg*
aktningsbjudande	dov
betydande	dvärglik
bjärt	dämpad
blossande	förkrympt
brinnande	gemen
dominerande	grov
drogad	kort
dryg	liggande
fin	liten
förnäm	lumpen
genomskinlig	låglänt
gräll	lågmäld
gäll	lågsinnad
högdragen	lågt
högfärdig	lågtstående
högljudd	mild
högröstad	mindervärdig
högt	nedrig
högvuxen	oansenlig
högväxt	obetydlig
imponerande	ofin
imposant	ohyfsad
klar	otydlig
kraftig	platt
livlig	plump
ljudlig	primitiv
lång	reducerad
majestätisk	ringa
mallig	ringaktad
molnfri	sakta
mäktig	sekunda
nackstyv	simpel
nedlåtande	småväxt
påtänd	stilla
påverkad	svag
ren	tarvlig
reslig	tyst
skarp	undanskjuten
snorkig	underordnad
stark	usel
stor	vardaglig
sublim	vulgär
tydlig	
upphöjd	
upprymd	
uppsatt	
vördnadsbjudande	
ädel	
ärad	

ljus	*mörk*
blond	djup
fin	djuplodande
fördelaktig	djupsinnig
genomskinlig	dov
glad	dunkel
glansfull	hemlighetsfull
glänsande	inträngande
glättig	invecklad
god	mättad
harmonisk	obegriplig
hoppfull	ofattlig
klar	profund
lovande	svårfattlig
lycklig	underfundig
lysande	
lätt	
munter	
optimistisk	
präktig	
skinande	
snygg	
solig	
solklar	
transparent	
trevlig	
upplivande	
upplyst	
uppmuntrande	
vit	
öppen	

lätt	*svår*
enkel	allvarlig
finlemmad	allvarsam
hårfin	ansträngande
klen	bekymmersam
knapp	benig
liten	besvärlig
mager	betungande
obetydlig	bitter
ringa	brydsam
skrumpnad	elakartad
slank	farlig
smal	fatal
smäcker	förtretlig
smärt	grav
sparsam	intrikat
spenslig	invecklad
spinkig	kinkig
späd	komplicerad
trång	kritisk
tunn	krånglig
utmärglad	kvistig
	kännbar
	mödosam
	osäker
	penibel
	pinsam
	plågsam
	prekär
	påkostande
	smärtsam
	snårig
	svårfattlig
	svårlöst
	svårtydd
	trasslig
	trist
	vansklig
	ömtålig

tjock	*smal*
fet	fin
bred	genomskinlig
frodig	gles
fyllig	innehållsfattig
grov	klar
grumlig	luftig
grötig	lätt
knubbig	lättflytande
korpulent	mager
kraftig	skör
mastig	smäcker
mullig	smärt
omfångsrik	spinkig
plufsig	spröd
rundnätt	transparent
simmig	tunn
stinn	utspädd
svullen	vattning
trögflytande	ytlig
tät	
vidlyftig	
voluminös	
välgödd	
yppig	

bred	*smal*
grov	enkel
kraftig	finlemmad
omfångsrik	hårfin
platt	klen
rymlig	knapp
tjock	liten
utbredd	lätt
utsträckt	mager
vid	obetydlig
vidlyftig	ringa
vidsträckt	skrumpnad
öppen	slank
	smäcker
	smärt
	sparsam
	spenslig
	spinkig
	späd
	trång
	tunn
	utmärglad

varm	*kall*
echaufferad	avmätt
eldig	avvisande
glödande	flyktig
het	fåordig
hjärtlig	knapp
innerlig	kort
kokande	kortfattad
kylslagen	korthuggen
livfull	kortvarig
ljummen	kortvuxen
svettig	kortväxt
tempererad	kylig
upphettad	liten
varmblodig	oansenlig
	ovänlig
	småväxt
	snabb

tung	*lätt*
bastant	alkoholsvag
besvärlig	enkel
dyster	eterisk
däst	flyktig
förkrossande	gladlynt
massiv	klar
mödosam	laber
ohanterlig	ledig
svår	lindrig
svårsmält	legär
tyngande	luftig
vägande	lättfattlig
åbäkig	lättsinnig
	lös
	löslig
	mild
	obehindrad
	obetydlig
	obetänksam
	okomplicerad
	ostadig
	ringa
	rörlig
	sakta
	sangvinisk
	snabb
	sorglös
	svag
	tanklös
	tunn
	utspädd
	vig

stor	*liten*
ansenlig	betydelselös
avsevärd	diminutiv
berömd	dvärglik
betydande	futtig
betydlig	fåtalig
bred	föraktlig
diger	förkrympt
dryg	försumbar
enorm	gles
framstående	klen
fullvuxen	knapp
förnäm	kort
grov	lumpen
himmelsvid	låg
högrest	marginell
inflytelserik	mikroskopisk
innehållsrik	minimal
kolossal	nätt
kraftig	oansenlig
liten	obetydlig
lång	ringa
muskulös	skral
mycken	småaktig
mäktig	småväxt
märklig	snäv
oerhörd	sparsam
ofantlig	späd
omfattande	svag
omfångsrik	torftig
omåttlig	trång
otrolig	urvuxen
riklig	
ryktbar	
rymlig	
stark	
storsinnad	
storsint	
storväxt	
utmärkt	
utsträckt	
vidsträckt	
voluminös	
väldig	
ädel	

TRAVAUX DE L'INSTITUT DE LINGUISTIQUE DE LUND

FONDÉS PAR BERTIL MALMBERG. PUBLIÉS PAR GÖSTA BRUCE ET ÅKE VIBERG

1 *Carl-Gustaf Söderberg*. A Typological Study on the Phonetic Structure of English Words with an Instrumental-Phonetic Excursus on English Stress. 1959.
2 *Peter S. Green*. Consonant-Vowel Transitions. A Spectrographic Study. 1959.
3 *Kerstin Hadding-Koch*. Acoustico-Phonetic Studies in the Intonation of Southern Swedish. 1961.
4 *Börje Segerbäck*. La réalisation d'une opposition de tonèmes dans des dissyllabes chuchotés. Étude de phonétique expérimentale. 1966.
5 *Velta Ruke-Dravina*. Mehrsprachigkeit im Vorschulalter. 1967.
6 *Eva Gårding*. Internal Juncture in Swedish. 1967.
7 *Folke Strenger*. Les voyelles nasales françaises. 1969.
8 *Edward Carney*. Hiss Transitions and their Perception. 1970.
9 *Faith Ann Johansson*. Immigrant Swedish Phonology. 1973.
10 *Robert Bannert*. Mittelbairische Phonologie auf akustischer und perzeptorischer Grundlage. 1976.
11 *Eva Gårding*. The Scandinavian Word Accents. 1977.
12 *Gösta Bruce*. Swedish Word Accents in Sentence Perspective. 1977.
13 *Eva Gårding, Gösta Bruce, Robert Bannert* (eds.). Nordic Prosody. 1978.
14 *Ewa Söderpalm*. Speech Errors in Normal and Pathological Speech. 1979.
15 *Kerstin Nauclér*. Perspectives on Misspellings. 1980.
16 *Per Lindblad*. Svenskans sje- och tjeljud (Some Swedish sibilants). 1980.
17 *Eva Magnusson*. The Phonology of Language Disordered Children. 1983.
18 *Jan-Olof Svantesson*. Kammu Phonology and Morphology. 1983.
19 *Ulrika Nettelbladt*. Developmental Studies of Dysphonology in Children. 1983.
20 *Gisela Håkansson*. Teacher Talk. How Teachers Modify their Speech when Addressing Learners of Swedish as a Second Language. 1987.
21 *Paul Touati*. Structures prosodiques du suédois et du français. Profils temporels et configurations tonales. 1987.
22 *Antonis Botinis*. Stress and Prosodic Structure in Greek. A Phonological, Acoustic, Physiological and Perceptual Study. 1989.
23 *Karina Vamling*. Complementation in Georgian. 1989.
24 *David House*. Tonal Perception in Speech. 1990.
25 *Emilio Rivano Fischer*. Topology and Dynamics of Interactions - with Special Reference to Spanish and Mapudungu. 1991.
26 *Magnus Olsson*. Hungarian Phonology and Morphology. 1992.
27 *Yasuko Nagano-Madsen*. Mora and Prosodic Coordination. A Phonetic Study of Japanese, Eskimo and Yoruba. 1992.
28 *Barbara Gawronska*. An MT Oriented Model of Aspect and Article Semantics. 1993.
29 *Bengt Sigurd* (ed.). Computerized Grammars for Analysis and Machine Translation. 1994.
30 *Arthur Holmer*. A Parametric Grammar of Seediq. 1996.
31 *Ingmarie Mellenius*. The Acquisition of Nominal Compounding in Swedish. 1997.
32 *Christina Thornell.* The Sango Language and Its Lexicon (Sêndâ-yângâ tî Sängö). 1997.
33 *Duncan Markham.* Phonetic Imitation, Accent, and the Learner. 1997.
34 *Christer Johansson.* A View from Language. Growth of Language in Individuals and Populations. 1997.
35 *Marianne Gullberg.* Gesture as a Communication Strategy in Second Langauge Discourse. A Study of Learners of French and Swedish. 1998.
36 *Mechtild Tronnier*. Nasals and Nasalisation in Speech Production. With Special Emphasis on Methodology and Osaka Japanese. 1998.
37 *Ann Lindvall*. Transitivity in Discourse. A Comparison of Greek, Polish and Swedish. 1998.
38 *Kirsten Haastrup & Åke Viberg* (eds.). Perspectives on Lexical Acquisition in a Second Language. 1998.
39 *Arthur Holmer, Jan-Olof Svantesson & Åke Viberg* (eds). Proceedings of the 18th Scandinavian Conference of Linguistics. 2001.
40 *Caroline Willners.* Antonyms in Context. A Corpus-based Semantic Analysis of Swedish Descriptive Adjectives. 2001.