

Project ref. no. LE3-4239

Project title *SCARRIE Scandinavian Proof-reading Tools*

Deliverable number DEL 5.3.3

Deliverable title *Multi-word Expressions for Swedish*

Number of pages 30

WP/Task responsible *Anna Sågvall Hein, Department of Linguistics, Uppsala University, Box 513, S-751 20 Uppsala, Sweden
Email: anna@ling.uu.se*

Author(s) *Olga Wedbjer Rambell*

EC Project Officer *Ray Hudson, Antonio Sanfilippo*

Keywords *multi-word expressions, idioms, multi-word units, Swedish*

Abstract Multi-word expressions have been collected primarily to be used by the grammar checker. The focus has been on the syntactical behaviour of the phrases.

Multi-word expressions containing words that only appear in certain phrases have also been identified and incorporated in the dictionary.

An outline is given of how to proceed with the work on the multi-word expressions dictionary.

Executive Summary

Multi-word expressions have been collected mainly from a core vocabulary of Swedish defined for a lexical database at the Department of Linguistics, Uppsala University. These phrases have been complemented by multi-word expressions from other sources such as Swedish grammar books.

The selection criteria have focused on the syntactical functions of the multi-word expressions. A phrase has been inserted in the idiom dictionary if its syntactical function cannot be easily derived from the phrase type, if it is not formed according to regular rules, or if it contain a multi-word unit. A multi-word unit is a word that only can occur in a very limited number of phrases.

In total, 508 multi-word expressions have been incorporated in the idiom dictionary after the initial selection phase. A fourth of these are multi-word prepositions, a fifth are adverbials, and another fifth are prepositional phrases. The rest consists of multi-word pronouns and subordinating conjunctions. Further investigations ought to be carried out. For instance, frequent noun phrases functioning as adverbials ought to be identified.

Multi-word Expressions for Swedish

The Initial Selection Phase

Olga Wedbjer Rambell

Department of Linguistics, Uppsala University

SCARRIE

DEL 5.3.3

29 May 1998

Contents

1	Introduction	1
2	Objectives	1
3	Method and Material	2
3.1	Sources	2
3.2	Selection Criteria	3
3.3	Multi-word Units	3
3.4	N-grams and Frequency Information	4
3.5	Grammatical Codes	4
4	Adjective Phrases	4
5	Adverbials	4
5.1	Adverb Phrases	4
5.2	Noun Phrases	5
5.3	Prepositional Phrases	5
5.4	Various Phrase Types	5
6	Prepositional Phrases	5
7	Noun Phrases	6
8	Pronouns	6
9	Prepositions	6
10	Conjunctions	6
10.1	Coordinating Conjunctions	6
10.2	Subordinating Conjunctions	7
11	Summary and Future Work	7
	Literature	9

Appendix A: The Idiom Dictionary

Appendix B: Multi-word Units

Appendix C: Key words

1 Introduction

The idiom dictionary will fill multiple purposes in the Scarrie program, since it will be used both for spell checking and grammar checking. It contains multi-word expressions with or without foreseen errors. It also handles phrases containing words that only appear in certain phrases (so called multi-word units).

In the initial selection phase, the multi-word expressions to be included in the Scarrie idiom dictionary should either incorporate a multi-word unit or have a syntactic function which cannot be derived from the combination of the words in the phrase. Such multi-word expressions have been collected and assigned a grammatical code.

This report gives an account of the objectives (chapter 2) and the material and the methods used to establish which phrases to include in the Scarrie idiom dictionary (chapter 3). In chapters 4–10, different function types are discussed in more detail. The report ends with a summary and an outline of future work with multi-word expressions within the Scarrie project.

2 Objectives

The objective of the deliverable presented in this report was to collect multi-word expressions primarily to be used for grammar checking purposes. The focus is thus on the syntactic behaviour of the phrases. During the collecting process, phrases with multi-word units have also been identified and incorporated in the idiom dictionary.

The multi-word expressions do not need to be idioms in a strict sense (see Teleman 1994 p. 212 for a definition and a discussion of idioms) to be included in the idiom dictionary (`idiom.txt`). On the other hand, for efficiency reasons, it is important to restrict the size of the idiom dictionary, why frequency information cannot be the primary basis for deciding whether a phrase ought to be included in the dictionary or not.

For spell checking purposes, it is important to identify words that can only appear in a limited number of phrases. If such a multi-word unit has been erroneously used in another context, the spell checker should recognise the error since the word is not a member of the main word form dictionary. If the word, on the other hand, is used correctly, the phrase should be approved by the program. (The multi-word units are stored in `idiom.idf`.)

For parsing purposes, multi-word expressions must be identified whose functions cannot be easily derived from the words in the phrase. Examples of such phrases are multi-word prepositions. Noun phrases do not usually function as adverbials, but some phrases might due to their semantic content. Frequent phrases of this type ought to be identified and included in the dictionary. Prepositional phrases are of less interest since they normally function as adverbials. However, phrases that are joined in a less regular way may be included in the idiom dictionary. To establish the relevance of certain expressions for the Scarrie project, frequency information has to be obtained for the phrases from the archive material delivered by the Swedish newspapers *Upsala Nya Tidning* and *Svenska Dagbladet*. This work lies, however, outside the scope of the initial selection phase.

Each phrase in the idiom dictionary is given a morphosyntactic code to be used by the grammar checker. The classification of the phrases focuses on the functions that the expressions may have, and not on the phrase type or on the internal structure of the phrase.

3 Method and Material

3.1 Sources

Multi-word expressions were collected mainly from a Swedish core vocabulary defined for a lexical database at the Department of Linguistics, Uppsala University, presented in Wikholm (1994). This core vocabulary included nearly 6.000 multi-word expressions classified according to phrase type, internal structure, and syntactic function. In a first stage, phrases containing variable elements were excluded together with discontinuous expressions leaving 4.685 phrases. These expressions were sorted after phrase type (e.g. NP, PP) and function type (that is, word category, phrase category or functional category, e.g. preposition, noun phrase, adverbial). The number of phrases in each combination of phrase type and function type is shown in table 3.1.

Table 3.1 Multi-word expressions from Wikholm (1994)

Phrase type	Function type	No of phrases
AP	AP	20
AdvP	advl	242
NP	advl	1294
	NP	6
	subj, obj, compl	4
	subj, obj, compl, advl	993
	subj, compl	4
PronP	pron	12
PP	advl, attr, compl	409
	advl	1434
	prep	176
ConjP	coord. conj	16
Sub.conjP	subord. conj	5
AttP	subord. conj	50
OmP	subord. conj	3
SomP	subord. conj	17
TOT		4685

Comments to table 3.1:

AttP, *OmP* and *SomP* are phrases ending with *att*, *om* and *som*, respectively.

Other sources used in collecting functional phrases were grammar books (Ljung & Ohlander 1982, Teleman 1974, Thorell 1973) and the error corpora database (Wedbjer Rambell et al. 1998). In total, 208 multi-word expressions were gathered from these sources and divided into different function types only (see table 3.2).

Table 3.2 Multi-word expressions from other sources

Function type	No of phrases
advl	90
pron	79
prep	18
coord. conj	7
subord. conj	14
TOT	208

3.2 Selection Criteria

All multi-word expressions are not equally interesting. For instance, adjective phrases functioning as adjective phrases can be excluded since they do not fulfil one of the following selection criteria: The phrase

- a) contains a multi-word unit,
- b) has a syntactic function that cannot be easily derived from the phrase type, or
- c) is not formed according to regular rules.

The phrases in the material have been manually inspected according to the selection criteria. Being an idiom in the strict sense does not automatically qualify the phrase to the idiom dictionary. The fact that a phrase is frequent does not necessarily mean that it ought to be in the idiom dictionary. A large idiom dictionary will slow down the performance of the spell checker, why a restrictive approach has been taken in selecting the multi-word expressions to be included in the dictionary.

3.3 Multi-word Units

A multi-word unit can only appear in a very limited number of phrases, in many cases in one phrase. These words should be moved from the main word form dictionary (if present there) to a special file (*idiom.idf*). The phrase(s) in which the word may occur should be listed in the idiom dictionary (*idiom.txt*).

Some of the multi-word units are ambiguous, belonging to more than one word class. For example, *våras* may be a noun or a verb. As a noun, it is a multi-word unit, while as a verb, it is not. For the time being, it is not totally clear how to handle these units. They ought to be removed from the main dictionary but probably not inserted in the file *idiom.idf*. In the example, *våras* as a noun should be removed from the main dictionary while it remains as a verb.

There are occasions where a word is a potential multi-word unit. That is, it may exist outside the multi-word expression, but is less likely to do so. For these words, it has to be established whether they should be treated as true multi-word units or not by comparing the number of occurrences of the word with frequency of the multi-word expression of which the word is a unit.

3.4 N-grams and Frequency Information

N-grams and frequency information about them have been formed from the text corpus for generating dictionaries (see Dahlqvist 1998 for a presentation of the text corpus). Sentence boundaries have not been crossed when making the n-grams. Punctuation marks such as comma and semicolon have not been removed since two words separated by a comma or a semicolon have a less tight connection to each other than two words only a space token apart. In addition, the distinction between headings and plain text has been up-held, because the style in headings and in plain text may differ. For instance, the copula is often left out in headings. When necessary, frequency information about multi-word expressions can be collected from these n-gram files.

The n-grams files were, however, not totally compiled when the work with the multi-word expressions was carried out. As a result, some work has to be done to improve the idiom dictionary. Some of this work is mentioned in the following sections discussing the different phrase types. An outline of future work is presented in the last chapter.

3.5 Grammatical Codes

Each multi-word expression has been assigned a code that comprises morphosyntactic information. Four categories have been added that are only used for multi-word expressions. QU is assigned to quantifiers expressing measurement such as *ett antal* and QM to quantifiers like *en grupp* and *ett parti* expressing a group of something (Thorell 1973:56f).

AD is assigned to phrases that function as adverbials. In addition, the phrase category PP is assigned to prepositional phrases that may fill any function an ordinary prepositional phrase may fill. All other codes are the same as for the words in the word form dictionary, with the same categories and feature values.

4 Adjective Phrases

The adjective phrases in the core vocabulary consist of coordinated adjectives, such as *bättre eller sämre*. None of them contain a multi-word unit. To coordinate adjectives is a regular way to form an adjective phrase. Therefore, the adjective phrases in the core vocabulary are irrelevant for the Scarrie idiom dictionary.

5 Adverbials

5.1 Adverb Phrases

Many adverb phrases in the core vocabulary are quite lexicalised, in that they are pronounced as one unit and not as two separate words. However, they are regularly formed and should not be included for parsing efficiency reasons. Frequency information has to be taken into consideration before including the adverb phrases in the idiom dictionary.

Two phrases in the core vocabulary may contain multi-word units. Frequency information can tell if the words are to be considered as such. If so, the phrases have to be added to the idiom dictionary.

5.2 Noun Phrases

Noun phrases functioning as adverbials need to be identified. Since this type of noun phrases are formed by productive rules, only the most frequent ones can be included for efficiency reasons. Frequency information from the n-grams files needs to be gathered for the collected noun phrases. In addition, other frequent noun phrases must have to be identified. This can be done with the help of key words, that is, nouns that expresses, for instance, a temporal content and thus can be the head word of a noun phrase functioning as an adverbial. Key words can be identified by assigning values to the variable elements in noun phrases in the core vocabulary (e.g. the variable *veckodagar* can take the values *måndag*, ..., *söndag*). Note that the values can occur in an inflected form, and that the head noun may be followed by an adverb or by a prepositional phrase. A list of possible key words is found in appendix C.

5.3 Prepositional Phrases

Prepositional phrases function usually as adverbials. There is therefore no reason to include prepositional phrases in the idiom dictionary due to syntactic behaviour. However, there are two cases in which a prepositional phrase should be inserted in the idiom dictionary. Firstly, if the phrase contains a multi-word unit. Two prepositional phrases containing multi-word units have been identified. Secondly, if it is formed in an irregular way.

A prepositional phrase does normally consist of a preposition and a noun phrase in the definite form. If the noun phrase consist of a noun in the singular and in the indefinite form, the prepositional phrase is considered to be irregularly formed, thus satisfying the selection criterion c) above (e.g. *i morgon*). If the complement in the prepositional phrase is an adjective phrase, the PP is also to be considered as irregularly formed (e.g. *på gott och ont*). This view on regularity in forming prepositional phrases is dependent on how the grammar checker functions, and can be revised if found appropriate.

Most of the prepositional phrases collected from the core vocabulary do not fulfil the selection criteria. The prepositional phrases are thus formed in a productive, regular way, and should not be included in the idiom dictionary. There are, however, prepositional phrases in the core vocabulary that contain variable elements that might fulfil the third selection criterion. The variable elements are the same as for the noun phrases, in particular the names of the days (e.g. *på fredag*). In total, 44 prepositional phrases have been taken from the core vocabulary and inserted into the idiom dictionary.

5.4 Various Phrase Types

Multi-word expressions that can function as adverbials were gathered from other sources than the core vocabulary. A list of 90 phrases was set up from which 56 have been selected to the idiom dictionary and considered to function as adverbials only.

6 Prepositional Phrases

Ordinary prepositional phrases do not only function as adverbials but also as attributes and complements. Prepositional phrases given the grammatical code AD function only as adverbials, while those assigned PP may fill other functions as well. The same selection criteria are used here as above. From the core vocabulary, 99 phrases were inserted in the idiom dictionary because of their irregular form or because of multi-word units. The prepositional phrases included so far do all consist of two words: a preposition and a noun.

There are 289 phrases consisting of three words or more that ought to be checked against the n-grams files, since they have properties that make them quite lexicalised but they may be produced in a regular, productive way.

From the multi-word expressions collected from other sources than the core vocabulary, some of the 90 phrases mentioned in section 5.4 are prepositional phrases that can function as adverbials, attributes and complements. Due to their fulfilling the selection criteria, 6 phrases were incorporated in the idiom dictionary and therefore assigned PP as the grammatical code.

7 Noun Phrases

Noun phrases do normally fill the syntactic functions of subject, object, and complement. The phrases from the core vocabulary do not contain any multi-word unit. For these reasons, they are of no interest for the purposes of the idiom dictionary. The same is true for the noun phrases functioning as noun phrases, which consist of coordinated nouns.

8 Pronouns

The pronoun phrases in the core vocabulary consist of coordinated pronouns, such as *den eller den* and *en och en*. Multi-word pronouns have been added from other sources, e.g. demonstrative pronouns (*den där*), and quantifiers (*ett par*). All but one of the coordinated pronouns from the core vocabulary are excluded from the idiom dictionary for the time being, while all other multi-word pronouns have been included with a grammatical code attached to each of them, separating them into the same pronoun categories as single word pronouns plus the phrase categories QU and QM for the multi-word quantifiers.

9 Prepositions

Multi-word prepositions are phrases that function as prepositions, such as *med hänsyn till* and *mitt emot*. 133 multi-word prepositions have been taken from the core vocabulary: All 102 phrases consisting of a preposition + a noun + a preposition have been included in the idiom dictionary together with all 8 phrases consisting of coordinated prepositions. In addition, 3 phrases with other internal structure function as prepositions.

The rest of the multi-word prepositions in the core vocabulary consist of an adverb + a preposition. These 63 phrases have been excluded in the first selection phase, since it is not obvious that these expressions really function as prepositions. The phrases may be included later on.

18 multi-word prepositions were collected from other sources than the core vocabulary. All of these have been included in the idiom dictionary.

10 Conjunctions

10.1 Coordinating Conjunctions

The multi-word coordinating conjunctions in the core vocabulary consist of a coordinating conjunction + an adverb. These phrases cannot be perceived as idioms and will not be part

of the idiom dictionary. The coordinating conjunctions collected from the grammar books have, on the other hand, been included in the idiom dictionary.

10.2 Subordinating Conjunctions

All multi-word subordinating conjunctions in the core vocabulary have been included except for the 5 phrases denoted *Subj.conjP* in table 3.1, which consist of a conjunction + an adverb. 70 multi-word subordinating conjunctions were thus collected from the core vocabulary. In addition, all subordinating conjunctions from the grammar books have been included in the idiom dictionary.

11 Summary and Future Work

In a first selection phase, 507 multi-word expressions have been included in the idiom dictionary: 327 multi-word expressions have been collected from the core vocabulary, and 180 phrases from other sources. The distribution among the different syntactical functions are shown in table 11.1. A full account of the idiom dictionary is given in appendix A.

Table 11.1 Multi-word expressions and their syntactic functions in the idiom dictionary

Function type	No of phrases
AP	0
AD	100
PP	105
NP	0
Pron	80
PR	131
CN	7
SN	84
TOT	507

24 multi-word units have been identified. In addition, 12 such units ought to be further investigated whether to be included in the multi-word unit file or not. The multi-word units and their phrases are listed in appendix B.

So far, syntactical function has been the main selection criteria. In the next phase, frequency information ought to be taken into consideration in identifying other multi-word expressions to be included in the idiom dictionary. The selection criteria will then need to be refined.

The future work with the multi-word expressions will include not only work with common word but will also focus on identifying multi-word proper names and their multi-word units. The n-grams files, in which words at sentence beginnings can be removed due to the sentence boundary markers, will be processed and information retrieved about sequences of words with initial capital letters. Frequency information about identified proper name

phrases can then be compared with the frequencies of the in-going names to find multi-word units. For instance, *Personvagnar* might perhaps only occur in the proper name *Volvo Personvagnar*. Proper names that are potential multi-word units have already been identified to some extent during the work with the word form dictionary, and have to be checked against the n-grams of proper names.

Abbreviations must be inserted in the dictionary. This work needs to be carried out in close cooperation with the newspapers, since there are different norms regarding how to use abbreviations. Special attention will be paid to non-valid abbreviations that have occurred in the archive materials, so that they can be corrected according to the newspapers' standards.

Foreseen errors in multi-word expressions ought to be identified not only for abbreviation but also for common words and proper nouns. This task ought to be carried out in cooperation with the newspapers. The error corpora database can be used as a source of information.

Future work with multi-word expressions may thus involve:

- frequency information on phrases collected in the first selection phase but not incorporated in the idiom dictionary (e.g. adverb phrases)
- frequency information for identified potential multi-word units
- searching the n-grams files after key words to identify frequent noun phrases that may function as adverbials
- searching the n-grams files for frequent proper names and their multi-word units
- searching the n-grams files for other frequent multi-word expressions
- foreseen errors in multi-word expressions (in cooperation with the newspapers)
- valid and non-valid abbreviations (in cooperation with the newspapers)

Literature

- Dahlqvist, Bengt (1998): *A Swedish Text Corpus for Generating Dictionaries*. SCARRIE DEL 3.1.3, Department of Linguistics, Uppsala University.
- Ljung, Magnus & Ohlander, Sölve (1982): *Allmän grammatik*. Stockholm.
- Teleman, Ulf (1974): *Manual för grammatisk beskrivning av talad och skriven svenska*. Lund.
- Thorell, Olof (1973): *Svensk grammatik*. Stockholm.
- Wedbjer Rambell et al. (1998): *An Error Database of Swedish*. SCARRIE DEL 2.1.3.2, Department of Linguistics, Uppsala University.
- Wikholm, Eva (1994): Svensk kärnvokabulär i MULDA: Enords- och flerordsenheter. In: *Reports from Uppsala University Linguistics, RUUL vol. 26*. P. 45–73. Uppsala.

Appendix A

The Idiom Dictionary

i grund och botten	: AD
i lugn och ro	: AD
i tur och ordning	: AD
med nöd och näppe	: AD
på ort och ställe	: AD
på sätt och vis	: AD
på vinst och förlust	: AD
till höger och vänster	: AD
till punkt och pricka	: AD
i mångt och mycket	: AD
på gott och ont	: AD
per minut	: AD
från morgon till kväll	: AD
i natt	: AD
per natt	: AD
i morgon	: AD
per dygn	: AD
per vecka	: AD
i sinom tid	: AD
på tid	: AD
i nord	: AD
i norr	: AD
i syd	: AD
i söder	: AD
i väst	: AD
i väster	: AD
i öst	: AD
i öster	: AD
mot nord	: AD
mot norr	: AD
mot syd	: AD
mot söder	: AD
mot väst	: AD
mot väster	: AD

mot öst	: AD
mot öster	: AD
åt nord	: AD
åt norr	: AD
åt syd	: AD
åt söder	: AD
åt väst	: AD
åt väster	: AD
åt öst	: AD
åt öster	: AD
allt som oftast	: AD
dörr i dörr	: AD
et cetera	: AD
framför allt	: AD
från topp till tå	: AD
för övrigt	: AD
för alltid	: AD
för fullt	: AD
för närvarande	: AD
hand i hand	: AD
helt enkelt	: AD
hur pass	: AD
i andanom	: AD
i tid	: AD
i kväll	: AD
i sommar	: AD
i övermorgon	: AD
i fjol	: AD
i förrgår	: AD
i förfjol	: AD
i går	: AD
i dag	: AD
i år	: AD
i morse	: AD
i julas	: AD
i påskas	: AD
i måndags	: AD
i tisdags	: AD
i onsdags	: AD
i torsdags	: AD

i fredags	: AD
i lördags	: AD
i söndags	: AD
i övrigt	: AD
i vintras	: AD
i höstas	: AD
i våras	: AD
i somras	: AD
ingen annanstans	: AD
ingen vart	: AD
med detsamma	: AD
med flera	: AD
med mera	: AD
någon annanstans	: AD
någon vart	: AD
rent av	: AD
slag i slag	: AD
strängt taget	: AD
till godo	: AD
till sist	: AD
till skogs	: AD
till övers	: AD
trots allt	: AD
vice versa	: AD
vägg i vägg	: AD
över huvud	: AD
i fråga	: PP
i hästväg	: PP
i matväg	: PP
i sänder	: PP
på vippen	: PP
till exempel	: PP
i allmänhet	: PP
på allvar	: PP
utan anledning	: PP
utan anmärkning	: PP
utan avbrott	: PP
på avstånd	: PP
på begäran	: PP
efter behag	: PP

efter behov	: PP
i beredskap	: PP
av betydelse	: PP
i bredd	: PP
i början	: PP
från början	: PP
i detalj	: PP
i efterhand	: PP
med eftertryck	: PP
med ens	: PP
till exempel	: PP
i fall	: PP
med fog	: PP
med framgång	: PP
till fullo	: PP
ur funktion	: PP
i funktion	: PP
i följd	: PP
i förbifarten	: PP
med fördel	: PP
på förhand	: PP
mot förmodan	: PP
efter förmåga	: PP
i förstone	: PP
på försök	: PP
i förtid	: PP
i förväg	: PP
över förväntan	: PP
i gengäld	: PP
i genomsnitt	: PP
utan gräns	: PP
för hand	: PP
efter hand	: PP
till hands	: PP
i huvudsak	: PP
till höger	: PP
vid jämförelse	: PP
för jämnan	: PP
i klump	: PP
under kontroll	: PP

i korthet	: PP
över lag	: PP
utan like	: PP
med lätthet	: PP
på måfå	: PP
i mängd	: PP
i nödfall	: PP
till nöds	: PP
med nödvändighet	: PP
utan omsvep	: PP
i onödan	: PP
i oordning	: PP
i oändlighet	: PP
vid pass	: PP
per person	: PP
enligt praxis	: PP
i princip	: PP
i regel	: PP
utan reservation	: PP
utom räckhåll	: PP
inom räckhåll	: PP
med rätta	: PP
i sammandrag	: PP
i samverkan	: PP
på sikt	: PP
till skänks	: PP
med skärpa	: PP
till slut	: PP
i sträck	: PP
per styck	: PP
utan svårighet	: PP
till synes	: PP
inom synhåll	: PP
i synnerhet	: PP
vid tillfälle	: PP
till trots	: PP
utan tvekan	: PP
utan tvivel	: PP
på tvären	: PP
i undantagsfall	: PP

i underkant	: PP
utan uppehåll	: PP
enligt uppgift	: PP
till vardags	: PP
i vardagslag	: PP
av vikt	: PP
till vänster	: PP
per år	: PP
till äventyrs	: PP
i överflöd	: PP
i överkant	: PP
var och en	: PDTDUSB
den här	: PDDDUSB
den där	: PDDDUSB
det här	: PDDDNSB
det där	: PDDDNSB
de här	: PDDDXPB
de där	: PDDDXPB
sådan där	: PDCIUSB
sådan här	: PDCIUSB
sån där	: PDCIUSB
sån här	: PDCIUSB
sådant där	: PDCINSB
sådant här	: PDCINSB
sånt där	: PDCINSB
sånt här	: PDCINSB
sådana där	: PDCIXPB
sådana här	: PDCIXPB
såna där	: PDCIXPB
såna här	: PDCIXPB
allt vad	: PVXXB
något som	: PVXXB
vad för	: PIXXBF
vad för en	: PIUSBX
vad för ett	: PINSBX
vad för någon	: PIUSBX
vad för något	: PINSBX
vad för några	: PIXPBX
en del	: QU
en handfull	: QU

en hel del	: QU
en hop	: QU
en massa	: QU
en myckenhet	: QU
en mängd	: QU
en rad	: QU
en smula	: QU
ett antal	: QU
ett flertal	: QU
ett fåtal	: QU
ett par	: QU
gott om	: QU
massor av	: QU
nog med	: QU
några få	: QU
några stycken	: QU
ont om	: QU
rikligt med	: QU
tillräckligt med	: QU
en grupp	: QM
en samling	: QM
en serie	: QM
ett lag	: QM
ett parti	: QM
ett stort uppbåd	: QM
vart och ett	: PDTDNSB
vars och ens	: PDTDUSG
ingen som helst	: PDTIUSB
inget som helst	: PDTINSB
inga som helst	: PDTIXPB
varje som helst	: PDTIXSB
vilken som helst	: PDTIUSB
var för sig	: PDTIXXB
vem helst	: PIUSBX
vem som helst	: PDTIUSB
vilken som	: PDTIUSB
vilket som	: PDTINSB
vilken som helst	: PDTIUSB
vilket som helst	: PDTINSB
vad helst	: PIXXBX

vad som helst	: PDTIXXB
hurdan som helst	: PDTIUSB
före detta	: PDCIXXB
litet av varje	: PDQIXXB
lite av varje	: PDQXXB
så kallad	: PDCIUSB
så kallat	: PDCINSB
så kallade	: PDCIXPB
var sin	: PDQIUSB
var sitt	: PDQINSB
var sina	: PDQIXPB
i analogi med	: PR
vid anblicken av	: PR
med anledning av	: PR
i anslutning till	: PR
med avseende på	: PR
i avvaktan på	: PR
mot bakgrund av	: PR
på basis av	: PR
med beaktande av	: PR
på bekostnad av	: PR
i betraktande av	: PR
i brist på	: PR
i början av	: PR
i början på	: PR
i centrum av	: PR
i centrum för	: PR
i egenskap av	: PR
i enlighet med	: PR
i form av	: PR
i fråga om	: PR
till följd av	: PR
i förbindelse med	: PR
till fördel för	: PR
i förening med	: PR
i förhållande till	: PR
till förmån för	: PR
på förslag av	: PR
till försvar för	: PR
under förutsättning av	: PR

till glädje för	: PR
på grund av	: PR
till grund för	: PR
på grundval av	: PR
med hjälp av	: PR
i hopp om	: PR
i händelse av	: PR
av hänsyn till	: PR
med hänsyn till	: PR
med hänvisning till	: PR
under hänvisning till	: PR
till höger om	: PR
i höjd med	: PR
under inflytande av	: PR
på initiativ av	: PR
på jakt efter	: PR
i jämförelse med	: PR
i klass med	: PR
i kombination med	: PR
i konkurrens med	: PR
i kontrast till	: PR
i kraft av	: PR
med ledning av	: PR
under ledning av	: PR
i likhet med	: PR
i linje med	: PR
i ljuset av	: PR
inom loppet av	: PR
under loppet av	: PR
i mitten av	: PR
i mitten på	: PR
vid mitten av	: PR
i motsats till	: PR
i motsättning till	: PR
efter måttet av	: PR
i namn av	: PR
i nivå med	: PR
i närheten av	: PR
i proportion till	: PR
inom ramen för	: PR

i relation till	: PR
i riktning mot	: PR
inom räckhåll för	: PR
i samarbete med	: PR
i samband med	: PR
i samförstånd med	: PR
i samråd med	: PR
i samverkan med	: PR
med sikte på	: PR
under sken av	: PR
till skillnad från	: PR
i skydd av	: PR
i skymundan för	: PR
i slutet av	: PR
i slutet på	: PR
vid slutet av	: PR
i spetsen för	: PR
i stil med	: PR
i stället för	: PR
med stöd av	: PR
till stöd för	: PR
på tal om	: PR
med tanke på	: PR
med tillhjälp av	: PR
med undantag för	: PR
med undantag av	: PR
i utbyte mot	: PR
på väg mot	: PR
på väg till	: PR
till vänster om	: PR
i väntan på	: PR
i överensstämmelse med	: PR
med eller utan	: PR
från och med	: PR
i och för	: PR
i och med	: PR
till och med	: PR
före och efter	: PR
inom och utom	: PR
med och utan	: PR

så när som på	: PR
vad gäller	: PR
tack vare	: PR
vid tiden för	: PR
i och med	: PR
från och med	: PR
till och med	: PR
i och för	: PR
i avseende på	: PR
beroende på	: PR
bortsett från	: PR
i avsknad av	: PR
i mån av	: PR
jämfört med	: PR
med hänsyn till	: PR
när det gäller	: PR
oberoende av	: PR
under loppet av	: PR
utan tanke på	: PR
vad beträffar	: PR
vad angår	: PR
vad avser	: PR
för övrigt	: CN
icke desto mindre	: CN
på grund härav	: CN
och för övrigt	: CN
vare sig	: CN
utan tvärtom	: CN
snarare än	: CN
i synnerhet som	: SNO
efter hand som	: SNO
så mycket mera som	: SNO
så mycket mindre som	: SNO
allra helst som	: SNO
så gott som	: SNO
så ofta som	: SNO
så snart som	: SNO
allt efter som	: SNO
bäst som	: SNO
helst som	: SNO

just som	: SNO
precis som	: SNO
samtidigt som	: SNO
så som	: SNO
särskilt som	: SNO
såtilllvida som	: SNO
i avvaktan på att	: SNO
i färd med att	: SNO
i hopp om att	: SNO
i samband med att	: SNO
i stället för att	: SNO
i väntan på att	: SNO
med hänsyn till att	: SNO
med hänvisning till att	: SNO
med tanke på att	: SNO
på grund av att	: SNO
under förutsättning av att	: SNO
av den orsaken att	: SNO
för den händelse att	: SNO
i den bemärkelse att	: SNO
med den skillnaden att	: SNO
på det villkor att	: SNO
ur den synpunkten att	: SNO
i all synnerhet som	: SNO
på samma sätt som	: SNO
i avsikt att	: SNO
i avsikten att	: SNO
i fall att	: SNO
i försök att	: SNO
i försöket att	: SNO
i försöken att	: SNO
i konsten att	: SNO
i syfte att	: SNO
under förutsättning att	: SNO
på villkor att	: SNO
efter det att	: SNO
i det att	: SNO
på det att	: SNO
under det att	: SNO
till dess att	: SNO

efter att	: SNO
för att	: SNO
genom att	: SNO
trots att	: SNO
utan att	: SNO
utom att	: SNO
i och med att	: SNO
därför att	: SNO
därigenom att	: SNO
fast att	: SNO
ifall att	: SNO
så att	: SNO
tills att	: SNO
låt vara att	: SNO
antag att	: SNO
förutsatt att	: SNO
helst om	: SNO
som om	: SNO
även om	: SNO
efter det	: SNO
i det	: SNO
rätt som	: SNO
så fort	: SNO
så fort som	: SNO
så länge	: SNO
så länge som	: SNO
så ofta	: SNO
så snart	: SNO
till dess	: SNO
under det	: SNO
förutsatt att	: SNO
i den mån	: SNO
är det så att	: SNO

Appendix B

Multi-word Units

If no information is given, the multi-word units are nouns.

Identified multi-word units

sinom (pron)	i sinom tid	: AD
urminnes (adj)	sedan urminnes tider	: AD
fullo	till fullo	: PP
förbifarten	i förbifarten	: PP
förstone	i förstone	: PP
förväg	i förväg	: PP
gengäld	i gengäld	: PP
jämnan	för jämnan	: PP
måfå	på måfå	: PP
onödan	i onödan	: PP
tvären	på tvären	: PP
vardagslag	i vardagslag	: PP
andanom	i andanom	: AD
hästväg	i hästväg	: PP
fjol	i fjol	: AD
förfjol	i förfjol	: AD
julas	i julas	: AD
påskas	i påskas	: AD
somras	i somras	: AD
höstas	i höstas	: AD
vintras	i vintras	: AD
vippen	på vippen	: PP
godo	till godo	: AD
övers	till övers	: AD

Potential multi-word units

omsider (adv)	sent omsider	: AD
---------------	--------------	------

småningom (adv)	så småningom	: AD
förtid	i förtid	: PP
hands	till hands	: PP
nöds	till nöds	: PP
skänks	till skänks	: PP
synes	till synes	: PP
vardags	till vardags	: PP
äventyrs	till äventyrs	: PP
matväg	i matväg	: PP
annanstans	ingen annanstans	: AD
	någon annanstans	: AD
vart	ingen vart	: AD
	någon vart	: AD

Appendix C

Key words

Frequent noun phrases that function as adverbials may be identified through searching after key words in the n-grams files. This appendix contain suggestions of such key words, which have been collected from the core vocabulary (Wikholm 1994).

minut
minuten
minuter
minuterna
sekund
sekunden
sekunder
sekunderna
timme
timmar
timmarna
morgontimme
morgontimmar
morgontimmarna
natt
natten
nätter
nätterna
dygn
dygnet
dygneten
helg
helgen
helger
helgerna
dag
dagen
dagar

dagarna
kväll
kvällen
kvällar
kvällarna
vecka
veckan
veckor
veckorna
månad
månaden
månader
månaderna
år
året
åren
århundrade
vår
våren
vårar
vårarna
sommar
sommaren
somrarna
höst
hösten
höstar
höstarna
vinter
vintern
vintrarna
tid
tiden
tider
tidpunkt
tidpunkten

tidpunkter
framtid
framtiden
gång
gången
gångar
veckodag
veckodagen
veckodagar
veckodagarna
månad

centimeter
decimeter
kilometer
meter
millimeter
mil
hektar
kubikcentimeter
kubikdecimeter
kubikkilometer
kubikmeter
kubikmillimeter
kvadratcentimeter
kvadratdecimeter
kvadratkilometer
kvadratmeter
kvadratmillimeter
centigram
decigram
dussin
gram
kilo
kilogram
milligram

ton
grad
grader

Bibliography and References

- Dahlqvist, Bengt (1998): *A Swedish Text Corpus for Generating Dictionaries*. SCARRIE DEL 3.1.3, Department of Linguistics, Uppsala University.
- Ljung, Magnus & Ohlander, Sölve (1982): *Allmän grammatik*. Stockholm.
- Teleman, Ulf (1974): *Manual för grammatisk beskrivning av talad och skriven svenska*. Lund.
- Thorell, Olof (1973): *Svensk grammatik*. Stockholm.
- Wedbjer Rambell et al. (1998): *An Error Database of Swedish*. SCARRIE DEL 2.1.3.2, Department of Linguistics, Uppsala University.
- Wikholm, Eva (1994): Svensk kärnvokabulär i MULDA: Enords- och flerordsenheter. In: *Reports from Uppsala University Linguistics, RUUL vol. 26*. P. 45–73. Uppsala.