

From Subdomains and Parameters to Collocational Patterns: On the Analysis of Swedish Medical Collocations

Emma Sköldberg
Maria Toporowska Gronostaj
Göteborg University

This paper presents a study on Swedish collocations in an electronic medical lexicon, currently under construction at the University of Gothenburg, Department of Swedish Language. There are two strands discussed in the paper. The first one is about a knowledge-based, onomasiological, approach to detecting and analysing medical collocations and their patterning. The second one deals with the representation of these collocations in both a general lexicon module and a collocational lexicon module. In the latter module, there are some advanced search options made available which enable selective access to the content of the lexicon. It is assumed that the onomasiological approach to the analysis of medical collocations complements the semasiological one and that the fusion of the two paves the way for a more consistent and exhaustive description of medical collocations and their patterns.

1. Introduction

A rapidly growing consumption of health care is a characteristic feature of today's society. Laypersons are expected to be informed health consumers, actively responding to medical information supplied by the Internet and mass media. Their medical knowledge and the means of expressing it are growing concurrently with ongoing medical research. Furthermore, in contacts between medical professionals and laypersons, nuanced communication is a prerequisite for eliminating potential misinterpretations. There is no doubt that command of collocations plays a relevant role in such communication. This is one of the reasons why medical collocations deserve a more exhaustive and systematic presentation in general-purpose as well as medical lexicons.

The main purpose of this paper is to share some insights into the detection and description of medical collocations in a Swedish electronic lexicon, henceforth referred to as MedColl lexicon. The MedColl lexicon is conceived of as a potential module in a multifunctional medical workbench MedLex+, currently under construction at the University of Gothenburg, Department of Swedish Language. (For more details about MedLex+, see Kokkinakis and Toporowska Gronostaj in *Euralex proceedings 2008*.) The MedColl lexicon is meant to be descriptive and multifunctional, for both reception and production, having native and non-native laypersons and health care personnel as its potential users. To meet this challenge, a broad spectrum of both semasiological and onomasiological information on medical terms and their collocations will be provided in the lexicon as well as a wide choice of search options. The focus in our paper is on medical collocations with a nominal base.

We assume that a knowledge-oriented approach to lexicons, in line with the one proposed by Martin (2006), can promote a more consistent and exhaustive description of collocations. For this reason we view Swedish medical collocations from an onomasiological perspective. The collocations analyzed pertain to the medical subdomains DISEASES, DIAGNOSTIC TESTS and MEDICAL EXAMINATIONS as well as TREATMENTS. By increasing the number of subdomains investigated, as compared to Sköldberg and Toporowska Gronostaj (forthcoming), the feasibility of the methodology chosen is further verified.

It is also worth noting that compiling a medical collocational lexicon for Swedish may also contribute to meta-lexicographic research on specialized languages. The collocational lexicon

may also help to streamline computational procedures aiming at the automatic retrieval of collocations from corpora and supporting both machine and human translation tasks in the medical field.

This paper is organized as follows: the background part sketched in the next section is followed by a presentation of the material in section 3. The methodology and analytical model employed are presented in section 4. Modes of representation of collocations in the MedColl lexicon, including search functions, are dealt with in section 5. A short summary is given in the last section.

2. Background

2.1. *Basic characteristics of collocations*

Research on collocations is extensive in conventional as well as computational lexicography and the term *collocation* is used in widely divergent senses (see Hausmann 2004 and Siepmann 2005). The questions posed often concern the basic properties of collocations in general language, including the issues concerning the boundaries between collocations and idioms, on the one hand, and collocations and free combinations, on the other. Our interpretation of the term has to a considerable extent been inspired by Hausmann (1985), Malmgren (2003), Bartsch (2004), Heid (2004), Svensén (2004), and Martin (2006). The most salient properties of collocations are listed in (1) to (5) below, together with examples taken from the medical domain.

A typical collocation is a recurrent combination of two or more words with an internal semantic and syntactic structure displaying lexical binding, e.g. *kraftig infektion* (heavy infection) and *ta tempen* (take one's temperature).

A typical collocation usually consists of a base and a collocate. The base is regarded as a kind of semantic head, while the collocate functions as a support or modifying word. In the collocation *ta tempen* (take one's temperature), *tempen* (temperature) is the base and the de-lexicalized verb *ta* (*take*) is the collocate. Since it is often the base that determines the lexical choice of the collocate, a relation of directionality can be traced in many collocations, in particular when the collocate is a delexicalized verb. However, there are collocations in which the directionality is less easy to detect, namely when the verb contributes in a more significant way to the meaning of a collocation (e.g. *ordinera läkemedel*, prescribe drugs).

A typical collocation is to a considerable extent transparent and therefore fairly easy to understand, but more difficult to produce. Nevertheless, the degree of transparency can vary due to such factors as, for instance, polysemy or metaphoric use of an element in a collocation. But the mere fact that the elements in a collocation keep company helps to dissolve the potential ambiguity of the items. The polysemy case can be illustrated with the noun *förband* having two main meanings, namely, "bandage" and "military unit". Whenever the word *förband* is a part of the collocation *lägga förband* (to put on a bandage), the meaning of the base is disambiguated. A similar phenomenon can be observed in collocations like *tung andning* (lit. heavy breath), where the collocate *tung* (heavy) is used metaphorically.

A typical collocation captures recurrent semantic relations between the base and the collocate. The most frequent relations are described in terms of lexical functions (Mel'čuk 1998). For instance, the lexical function Magn, which captures a degree relation, appears in collocations like *hög feber* (high temperature/fever), *djup depression* (deep depression) and *svår huvudvärk* (bad/severe headache).

A typical collocation may include implicit cultural and encyclopaedic information which needs to be decoded in order to fully understand the intended meaning. For example, full comprehension of the meaning of the collocation *fri abort* (lit. free abortion) presupposes some knowledge of the social system and its regulations. However, some expert knowledge is required for decoding collocations like *medicinsk abort* (medical abortion) and *kirurgisk abort* (surgical abortion).

Furthermore, collocations are usually classified according to their syntactic structure. Some syntactic patterns recurrent in Swedish, with examples from medical language, are listed below:

ADJ + NOUN	<i>molande värk</i> (dull pain)
VERB + NOUN	<i>ställa diagnos</i> (make a diagnosis)
VERB + PP	<i>svara på medicinen</i> (respond to the medicine)
VERB + ADV	<i>andas djupt</i> (breathe deeply)
ADV + ADJ	<i>djupt medvetslös</i> (deeply unconscious)
NOUN + VERB	<i>febern stiger</i> (the temperature rises)

In several studies on collocations, it has been pointed out that collocations can occur in parallel syntactic structures, such as *djupt medvetslös* (ADV + ADJ) (deeply unconscious) and *djup medvetslöshet* (ADJ + NOUN) (deep unconsciousness).

2.2. Towards a systematisation of collocations in medical language

The above, somewhat rough characteristics of typical collocations have been stipulated in the first place for describing collocations in general language. However, as follows from the examples given in 2.1., the arguments are well suited for detecting collocations with medical content. Thus, the question is whether there are systematic differences between the collocations in general language and those in medical language and, if so, what the nature of these differences is and how they are manifested. To answer this question, one needs to take a look at different language levels, namely lexical, syntactic and conceptual (cf. Martin 1992).

As far as the lexical level is concerned, medical language obviously shows a higher proportion of medically related words in collocations than general language does. On the syntactic level, however, there are no conspicuous structural differences between these types of language. A look at the conceptual level reveals some differences of interest for this study. These differences can be seen against the background of a frame-based model as presented by Martin (2006), advocating a knowledge-oriented approach of relevance not only for the analysis of collocations but also for dictionary construction. The conflation of onomasiological and lexico-semantic dimensions contributes to the classification of collocations with regard to the degree of binding between the elements, in particular on the conceptual level. Martin distinguishes between *token bound* and *type bound collocations*, on the one hand, and *non-bound combinations*, on the other, differing as to the degree of lexical binding between the elements (Martin 2006: 289). Token bound collocations are characterised by strong lexical binding (e.g. *sprängande huvudvärk*, splitting headache), while type bound collocations display a tendency to weaker lexical binding (e.g. *akut leukemi*, acute leukaemia, *akut sjukdom*, acute disease). Since the degree of binding within particular collocation types shows some variation, the boundaries between the types are not very sharp and a certain overlap can be observed.

According to Martin, the frame model promotes systematization of the description of meanings as well as collocations. This observation applies to both general and specialized languages. However, the advantages of using frames stand out more clearly for the medical language investigated as its repository of concepts is strictly hierarchically structured and categorized according to the onomasiological relations displayed. The onomasiological relations, being part of medical collocations, are discussed in more detail in section 4.

3. Material

The point of departure for our study has been approximately 150 nouns referring to diseases, different types of diagnostic tests and medical examinations as well as treatments. These nouns are entries in both general-purpose language dictionaries, e.g. *Nationalencyklopedins ordbok* (1995-96), and medical dictionaries, e.g. *Medicinsk terminologi* (Lindskog 2004). We have studied the collocational behaviour of the nouns in both a general text corpus, Språkbanken (The Swedish Language Bank), containing 100 million Swedish words (<http://spraakbanken.gu.se>) and in a medical text corpus, MedLex Corpus, of 25 million Swedish words. As auxiliary tools to detect and extract collocations we have used KWIC concordances and a statistical program based on mutual information. The extracted list of collocations has been subjected to a critical evaluation based on the characteristics of collocations outlined in 2.1. The issue of drawing a distinction between collocations and other

types of multiword expressions has not been equally straightforward, as the word combinations examined matched the characteristics postulated with varying degrees of accuracy.

Even though medical collocations are the main subject of this paper, we would like to make a few comments on multiword terms with regard to their position in the microstructure of the medical lexicon. By a term, we mean a one-word or multiword form with a well-defined concept used recurrently in a specialized language (cf. Bergenholtz et al. 1997). While some multiword terms are fully legitimate entries in the medical lexicon, others are listed as collocations. Thus, multiword expressions like *gula febern* (yellow fever) and *Downs syndrom* (Down's syndrome) can be regarded as terms having entry status in the lexicon. Multiword terms like *gula febern* do not meet the requirements imposed on collocations in 2.1. as the relation between the syntactic head and its modifier is entirely of a morpho-syntactic nature. The lack of a proper semantic relation between the component items, *gula* and *febern*, turns this word combination into a simplex concept. A similar non-collocational status can be claimed for medical eponyms like *Down's syndrome*, since the lexico-semantic relation between the syntactic head and the modifier is not fully transparent. Thus, such eponyms also act as simplex concepts.

4. Parameter-based analysis of medical collocations

A general assumption underlying the approach presented below is that a fusion of semasiological and onomasiological approaches can in a relevant way contribute to a more exhaustive and systematic detection and description of collocations. Since the semasiological approach is prevalent in today's lexicographic practice, some arguments for including an onomasiological perspective as a part of our analytical model are given below:

The analysis and description of collocations in a sublanguage require not only linguistic knowledge but also some field-related knowledge, in our case of medicine.

Knowledge of how a medical field is ontologically structured into subdomains (e.g. anatomy, diseases, treatments, etc.) seems to provide some insights and guidance as to the types of recurrent lexical meanings expressed in a particular subdomain and on their lexical manifestations. Information provided by Medical subject headings thesaurus (MESH, <http://mesh.kib.ki.se>) has proved useful for the study undertaken.

Thematically viewed data promote the selection of a relevant set of recurrent, semantic parameters underlying the medical collocations examined.

The main purpose of the analysis outlined in the remainder of this section is to account for some thematic collocational patterns appearing in medical subdomains such as DISEASES, DIAGNOSTIC TESTS and MEDICAL EXAMINATIONS as well as TREATMENTS. The subdomain DISEASES is described more thoroughly in order to clarify the methodology underlying the analysis. The label DISEASES, as used throughout this paper, covers a wide spectrum of medical conditions termed either as diseases or illnesses.

4.1. Subdomain DISEASES

For each examined subdomain we have attempted to propose a set of knowledge-based parameters capturing aspects of the recurrent meanings of particular collocational sets. For example, the subdomain DISEASES is analysed in terms of parameters like (with English equivalents) Falling ill, Status, Stage, Disease course, Type, Cure and Recovery. Thus, it is assumed here that each specific set of parameters proposed for a particular medical subdomain builds a pattern. To verify and evaluate the relevance of these parameters, we have examined a number of collocations within each subdomain. The collocations have been chosen to cover a wide spectrum of lexicalized meanings. For instance, for the subdomain DISEASES, we have taken a look at terms which stand for diseases that differ considerably as to medical type, symptoms and gravity, namely *depression*, *diabetes*, *cancer* and *cold*. Collocations with these terms are described in Sköldbberg and Toporowska Gronostaj (forthcoming). In Table 1 we provide examples and illustrate the analytical procedure with two of these terms, namely *depression* and *cold*. In all the cases examined, the nouns indicating diseases form the base in the collocations.

Parameters	depression (depression)	förkylning (cold)
Falling ill	få en depression (get a d.) drabbas av (en) depression (be stricken with d.) gå in i en depression (go into (a) d.)	få en förkylning (get/catch a c.) drabbas av en förkylning (be stricken with a c.) dra på sig en förkylning (be stricken with a c.) ådra sig en förkylning (be stricken with a c.)
Status	ha (en) depression (have a d.) lida av (en) depression (suffer from d.)	ha en förkylning (have a c.)
Stage	djup depression (deep d.) svår depression (severe d.) lätt(are) depression (light/minor d.) mild depression (mild d.)	elak förkylning (nasty c.) kraftig förkylning (bad c.) lätt förkylning (slight c.)
Disease course	långdragen depression (protracted d.) kronisk depression (chronic d.) akut depression (acute d.)	långdragen förkylning (persistent c.) långvarig förkylning (persistent c.) envis förkylning (stubborn c.)
Type	årstidsbunden depression (seasonal d.) bipolär depression (bipolar d.)	
Cure	bota depression (cure d.) behandla depression (treat d.)	bota förkylning (cure a c.) kurera en förkylning (nurse a c.)
Recovery	komma ur en depression (come out of a d.) ta sig ur en depression (get over a d.)	

Table 1. Parameter-based classification of collocations with the nouns *depression* (depression, d.) and *förkylning* (cold, c.).

From the above examples, it follows that there are collocates which are typical of only one of the diseases. For instance, the collocates *gå in i* (go into) and *årstidsbunden* (seasonal) occur almost exclusively with the mentioned noun as a base, and so *gå in i depression* and *årstidsbunden depression* can be considered as token bound collocations. Other collocates are more recurrent and somewhat less restrictive as to choice of base, e.g. *få* (get) and *drabbas av* (be stricken with, contract), and so tend to occur in type bound collocations.

The empty fields in the table indicate lack of collocations and reflect a certain asymmetry on the conceptual level in the system. The asymmetry may result either from a factual incompatibility between a parameter and a disease type (e.g. the parameter Recovery in combination with

incurable diseases), or from a preference for using free combinations rather than collocations. It is worth noting that this asymmetric distribution of collocations reveals potential difficulties in encoding situations, as the distribution deviates from the assumed general pattern.

Finally, it can be noted that some of the parameters in Table 1, e.g. Falling ill and Stage, are analogous to lexical functions presented by Mel'čuk.

4.2. Subdomain *DIAGNOSTIC TESTS and MEDICAL EXAMINATIONS*

Situations relating to different types of diagnostic tests and examinations make up another subdomain. These situations involve interaction between health care personnel and patients. This interaction can be viewed from two perspectives, an agentive and a non-agentive one. These perspectives are considered here as criterial for distinguishing the parameters Performing and its converse Undergoing. Their lexical manifestations are illustrated in Table 2 with *blood test* and *ultrasound examination* as bases of the collocations.

Parameters	<i>blodprov</i> (blood test)	ultraljudsundersökning (ultrasound examination)
Performing	ta (ett) blodprov (take/do a b.) göra ett blodprov (make/do a b.)	göra en ultraljudsundersökning (lit. make an u.) utföra en ultraljudsundersökning (carry out an u.)
Undergoing	lämna (ett) blodprov (submit (a) b. lit. leave (a) b.) göra ett blodprov (make/do a b.)	göra en ultraljudsundersökning (lit. make an u.) genomgå en ultraljudsundersökning (undergo an u.)
Type	venöst blodprov (venous b.) kapillärt blodprov (capillary b.)	vaginal ultraljudsundersökning (vaginal u.) transrektal ultraljudsundersökning (transrectal u.)

Table 2. Parameter-based classification of collocations with the nouns *blodprov* (blood test, b.) and *ultraljudsundersökning* (ultrasound examination, u.).

From the Swedish examples in Table 2 follows that the collocate *göra* (lit. *make*) can be used both with agentive and non-agentive subjects. To be able to decode the sense of ambiguous collocations like *göra en ultraljudsundersökning* (lit. *make an ultrasound examination*), one has to look for clues in the context. Thus, the lexicon should provide some information on possible clues for decoding, but also give some information on encoding strategies. Similar comments could be included concerning the choice of collocates with regard to the complexity of the examination procedure. While simple procedures often employ light verb collocates, like *ta* (take) and *göra* (make), the more complex ones involve verbs like *utföra* (perform) and *genomgå* (undergo).

4.3. Subdomain *TREATMENTS*

The subdomain TREATMENTS shares some characteristics with the subdomain DIAGNOSTIC TESTS and MEDICAL EXAMINATIONS, due to interaction between medical professionals and patients. Thus, parameters like Performing and Undergoing turn up here as well, together with Type. The parameters Purpose, Method, Course, Point in time and Prognosis are new, as compared to the previously discussed subdomains and their parameters.

Parameters	operation	<i>strålbehandling</i> (radiotherapy)
Performing	<i>göra en operation</i> (lit. make an o.) <i>genomföra en operation</i> (carry out an o.)	ge strålbehandling (give r.)
Undergoing	<i>genomgå en operation</i> (undergo an o.)	få strålbehandling (get r.) <i>genomgå strålbehandling</i> (undergo r.)
Type	<i>kosmetisk operation</i> (cosmetic o.) <i>fryskirurgisk operation</i> (cryosurgical o.)	adjuvant strålbehandling (adjuvant r.)
Purpose	<i>explorativ operation</i> (exploratory o.) <i>preventiv operation</i> (preventive o.) <i>kurativ operation</i> (curative o.)	palliativ strålbehandling (palliative r.)
Method	<i>öppen operation</i> (open o.) <i>radikal operation</i> (radical o.) <i>fryskirurgisk operation</i> (cryosurgical o.)	inre/intern strålbehandling (internal r.) yttre/extern strålbehandling (external r.)
Course	<i>akut operation</i> (acute o.)	
Point in Time	<i>tidig operation</i> (early o.)	preoperativ strålbehandling (preoperative r.) intraoperativ strålbehandling (intraoperative r.) postoperativ strålbehandling (postoperative r.)
Prognosis	<i>lyckad operation</i> (successful o.)	

Table 3. Parameter-based classification of collocations with the nouns *operation* (operation, o. surgery, s.) and *strålbehandling* (radiotherapy, r.).

The parameters listed in Table 3 can serve as preliminary guidelines for further exploration of the subdomain TREATMENTS, which, besides different surgical treatments and radiotherapy, also includes drug therapy, rehabilitation, etc.

All the collocations considered so far have been interpreted as instantiations of one parameter, except *fryskirurgisk operation* (cryosurgical o.), listed in Table 3 under the parameters Type and Method. The double affiliation of this collocation follows from the complexity of the meaning associated with the compound collocate *cryosurgical*. The example shows that the proposed model can deal with the description of collocations displaying complex parameters.

It is a characteristic feature of the sets of collocations considered in this section that they include both noun and verb phrases. The noun phrases show a wide register ranging from colloquial collocations (e.g. *lyckad operation*, successful operation) to terminological ones (e.g. *bipolär depression*, bipolar depression). The terminological collocations, often—but not exclusively—grouped under the parameter Type, are those which presuppose a certain degree of medical knowledge, since often both the collocate and the base bear medical meanings. In contrast to such phrases, the verb phrases might seem somewhat easier to understand, as the verbal collocate is usually familiar; thus, the base alone determines the difficulty level of the collocation.

To conclude, in this section we have outlined a methodology for the analysis and description of medical collocations. The examples presented in Tables 1 to 3 corroborate the significance of the conceptual level for a systematic survey of medical collocations with regard to their particular subdomains. The parameters introduced help to formalize the relevant aspects of the meaning of collocations. However, it should be noted that neither the list of proposed parameters nor the list of collocations presented in this paper is exhaustive, but hopefully they may serve as guiding principles in a more extensive study. In the following section, we will

propose some practical solutions to how this information can be represented and searched for in the electronic medical lexicon.

5. Representation of collocations in the MedColl lexicon

The electronic MedColl lexicon is planned to comprise two linked modules, a general-purpose lexicon module and a collocational lexicon module. The modular structure of the lexicon and its content are aimed at meeting the following main objectives: (1) to provide a comprehensive and harmonised lexical resource with information on words and their context; (1) to meet a variety of search needs in a user-friendly and efficient way; and (3) to bridge the gap between general-purpose and specialized language dictionaries.

In both modules, the vocabulary used in the medical field will be focused on. The collocation module is designed as a complement to the general-purpose lexicon, but the two can be used separately. The words functioning as base in medical collocations will have the status of primary entries in the both lexicon modules. Thus their lexical specification will be more detailed, compared to the supplementary entries formed by collocates. Direct access to collocations will also be provided if a collocation is entered as a search string. The main reason for this, at first glance, somewhat redundant representation of entries is to keep different search options open for users with widely divergent search needs. Both those less familiar with the exact form of a collocation and those in command of it, will be adequately and effectively supported.

Below, we give an account of the information types and search functions to be included in the two lexicon modules, with no claims whatsoever concerning how the information will be visualised for lookups in the final electronic version of the lexicon.

5.1. *The general-purpose lexicon module with a medical profile*

The lexical data presented in the general-purpose lexicon module will be a subset of the data recorded in the lexicon module in the multifunctional workbench MedLex+. The following types of information will be provided: entry word, pronunciation (with preference for audio pronunciation), word class, inflectional forms, meaning indicated by a guide word and a definition including definition comment. By guide word, we mean a concise wording which captures the overall meaning of a particular sense. The role of guide words is to highlight the meaning(s) of a word and act as a short cut to the sense searched. For instance, the entry *cancer* would have the senses DISEASE and TUMOUR highlighted. Moreover, the guidewords are supposed to support onomasiological searches, which in this particular case would result in the lists of terms involving DISEASES or TUMOURS.

Furthermore, the entry will also provide information on compounds containing the entry word, on semantically related words, possible synonyms, antonyms, hyperonyms and hyponyms, if any, as well as information about the most frequent collocations with the entry word. Examples and some information on morphologically related words, focused on derivation, will complete the specification of the lexical part of the entry.

The example below illustrates the above-mentioned structure and content for the entry *depression* in the general module. The Swedish version is followed by a version in English.

lemma: depression

uttal: (inspelat uttal)

ordklass: substantiv

böjning: depression, depressionen, depressioner, etc.

betydelseindikator: SJUKDOM

definition: sjukligt tillstånd med svår (långvarig) nedstämdhet

definitionstillägg: ofta omfattande ångest, handlingsförlamning och livströtthet

sammansättningar: utmattningsdepression, förlösningsdepression m.fl.

semantiskt besläktade ord: antonym *mani*

kollokationer:

ADJ + depression djup depression, svår depression; lätt(are) depression |
akut depression, långdragen depression, kronisk depression |
årstidsbunden depression

VERB + depression drabbas av (en), gå in i en depression |
ha en depression, lida av en depression |
ta sig ur en depression

exempel: Hon lider av en djup depression som orsakar sömnlöshet

morfologiskt relaterade ord: v. deppa; adj. deprimerad, deprimerande, deppig, depressiv

länk: till kollokationsmodul

lemma: depression

pronunciation: (audiopronunciation)

word class: *noun*

inflection: depression, depressions

guide word: ILLNESS

definition: a medical condition (with long periods) of severe despondency

definition comment: often including anxiety, inability to act and tiredness of life

compounds: fatigue depression, postnatal depression, etc.

semantically related words: antonymy *mania*

collocations:

ADJ + depression deep depression, severe depression; mild depression |
acute depression, protracted depression, chronic depression | seasonal depression

VERB + depression be stricken with depression, go into (a) depression |
have a depression, suffer from a depression |

come out of a depression

example: She suffers from a deep depression causing insomnia

morphologically related words: adj. *depressed, depressing, depressive*

link: to the collocational module

As shown above, the collocations are sorted according to their grammatical structure. In the next step, they are separated with a vertical line, which groups them according to underlying parameters: Falling ill, Status, Recovery etc.

Words in this module are linked to corresponding words in the collocation module, where some additional information is provided as well as some more search possibilities.

5.2. The collocation module with a medical profile

As already hinted, the collocation module can be used either as an independent module or as a complement to the general-purpose lexicon module, outlined in section 5.1. In this module, the focus is on more exhaustive description of collocations with regard to their lexico-semantic and onomasiological properties. The exhaustive description supports active sharing of the meta-lexicographic knowledge with the lexicon users. Hopefully, this knowledge may promote a more profound understanding of the lexical system by the users of the lexicon and give them information beyond the scope covered by traditional printed dictionaries. To achieve this objective, we provide a possibility to make more advanced, onomasiological and semantic searches on collocations listed in the lexicon, based on the meta-categories used for the description and classification of the medical collocations (see section 5.3. for more details).

Let us first take a look at a lemma-based search, which is the default case, and the information types presented. The presentation, in the form of running text, provides information on lemma, word class, inflection, guide word and definition. This information is followed by a list of collocations, often more exhaustive than the corresponding information in the general module, as well as a link to the lemma in the general-purpose lexicon module. The microstructure of an entry is given below, with *depression* as an example. For reasons of space, only content in Swedish is provided.

lemma: depression

pronunciation: (inspelat uttal)

word class: substantiv

inflection: depression, depressionen, depressioner, etc.

guide word: SJUKDOM

definition: sjukligt tillstånd med svår (långvarig) nedstämdhet

collocations:

ADJ + depression djup depression, svår depression, allvarlig depression; mild depression, lätt(are) depression |
 akut depression, långdragen depression, kronisk depression |
 årstidsbunden depression, årstidsrelaterad depression |
 bipolär depression | postnatal depression | egentlig depression

v + depression få depression, drabbas av depression, gå in i en depression | ha depression, lida av depression | behandla en depression, medicinera mot depression, bota en depression | ta sig ur en depression, komma ur en depression

link: till allmänspråklig modul

The listed collocations are structured grammatically and semantically according to the same principles as in the general module, but here they also function as links. A click on a specific collocation will open a pop-up window showing more information on that particular collocation. In the pop-up window, the user is supplied with information on the parameter(s) underlying this collocation, the collocation's relative frequency, its register with respect to domain and style. Furthermore, authentic examples from corpora are provided as well as comments, if required. In the comment field some further information on the collocation's meaning and usage can be given. For example, if one chooses to activate the link to the collocation *svår depression*, the following information could be made available to the lexicon user:

svår depression

parameter: Grad (Stage)

frekvens: **** (= mycket frekvent) (very frequent)

bruklighet: allmänspråk (general language)

exempel: *Vid en medelsvår eller svår depression väljer läkaren ofta att kombinera samtalsterapi med en medicinsk behandling.* (In cases of moderate or severe depression the doctor often tends to combine conversational therapy with medical treatment.)

kommentar: Indelning i *lätt*, *måttlig* och *svår depression* bygger på en skattningsskala utifrån förekommande symtom. Det är betydelsefullt att gradera depressionen eftersom patientens förutsättningar att bli fullt frisk ökar om behandlingen anpassas efter depressionens svårighetsgrad. (Classification into *mild depression*, *moderate depression* and *severe depression* is based on the estimation scale according to the symptoms shown. It is important to distinguish these stages as the patient's chances of recovery will increase if the treatment is adopted to the degree of depression.)

5.3. Search options in the collocation module

In the preceding two subsections, we have hinted at the lemma-based search and the selective collocation search, providing more lexico-semantic and encyclopaedic information on collocations, served in a pop-up window activated by the user. These search options can be complemented with some more advanced, semantically oriented search possibilities, as outlined below.

Simple semantic search with a guide word or parameter as search criterion. For instance, a global search using the guide word DISEASE would result in a list over names of diseases occurring in collocations. A global search using the parameter Grade lists adjectives occurring in that function, such as *svår* (severe), *lätt* (mild), *långt framskriden* (advanced), etc.

Mixed semantic search in which a guide word and/or parameter is combined with a concrete word. This search can be helpful if one, for example, wants to know what diseases or treatments can combine with *acute* as a modifier (e.g. *acute DISEASE*, *acute TREATMENT*). The former

search would yield *akut depression* (acute depression), *akut pankreatit* (acute pancreatitis) and *akut lymfatisk leukemi* (acute lymphatic leukaemia), among other collocations. It is also possible to find out with this search option whether and how a certain parameter can be manifested by a certain word. For instance, the parameter Grade in combination with the noun *depression* would result in *djup depression* (deep depression), *svår depression* (severe depression), *lätt(are) depression* (mild depression), etc.

Pattern search provides information on the types of semantic patterns instantiated in the collocation module. Such search makes it possible to find out, for instance, what parameters combine with DISEASE. The resulting patterns include such as Falling ill+DISEASE, Stage+DISEASE, Disease Course+DISEASE, Cure+DISEASE and Recovery+DISEASE.

Pattern-based search is designed to find all collocations in the lexicon that match a chosen pattern. For example, the pattern Falling ill+DISEASE results in *få en depression* (get a depression), *gå in i en depression* (go into a depression), *få cancer* (get cancer), *utveckla cancer* (develop cancer), etc.

The extended inventory of search options opens up multiple possibilities of accessing information on medical collocations, if the onomasiological and lexico-semantic information is made an inherent part of the lexicon. The meta-lexicographic layer offers an efficient tool for making implicit knowledge explicit, structuring it by means of patterns. These patterns support the lexicon users in their searches, lexicographers in checking the consistency of the lexical data they work on, and computational linguists in tasks such as automatic retrieval of collocations from corpora and text understanding.

6. Summary

In this paper we have presented ongoing work on medical collocations and their representation in the MedColl lexicon, a potential module in MedLex+, currently under construction at the University of Gothenburg, Department of Swedish Language. The analytical model applied has shown that integrated lexical and encyclopaedic knowledge promotes identification of parameters relevant for particular medical subdomains, and also that these parameters capture underlying collocational patterns. The collocational patterns are not only a useful tool for lexicographers in their efforts to make collocational lexicons exhaustive; they also provide efficient shortcuts for smart search functions in the electronic MedColl lexicon. The explicit access to lexical and encyclopaedic knowledge also provides a valuable opportunity for lexicon users to get more insights into meta-lexicography and to make them realize that such knowledge may promote information searches. We are fully aware that the collocational patterns presented here are restricted to the subjects discussed. Nonetheless, we assume that the methodology proposed could be applied to other medical subdomains and also to studies of collocations in other specialized languages.

References

- Bartsch, S. (2004). *Structural and Functional Properties of Collocations in English. A corpus study of lexical and pragmatic constraints on lexical co-occurrence*. Tübingen: Gunter Narr Verlag.
- Bergenholtz, H. et al. (1997). *Nordisk leksikografisk ordbok*. Oslo: Universitetsforlaget.
- Hausmann, F. J. (1985). "Kollokationen im deutschen Wörterbuch. Ein Beitrag zur Theorie des lexikographischen Beispiels". In Bergenholtz, H.; Mugdan, J. (eds.) *Lexikographie und Grammatik. Akten des Essener Kolloquiums zur Grammatik im Wörterbuch* 28.–30.6. Tübingen: Niemeyer. 226–257.
- Hausmann, F. J. (2004). "Was sind eigentlich Kollokationen?". In Steyer, K. (ed.) *Wortverbindungen—mehr oder weniger fest. (Institut für Deutsche Sprache, Jahrbuch 2003)*. Berlin, New York: Walter de Gruyter. 309-334.
- Heid, U. (2004). "On the presentation of collocations in monolingual dictionaries". In Williams, G.; Vessier, S. (eds.) *Proceedings of the Eleventh EURALEX International Congress, Lorient, France July 6-10, 2004*. Lorient: Université de Bretagne-Sud. 729–738.
- Kokkinakis, D.; Toporowska Gronostaj, M. (2008). "MedLex+: An Integrated Corpus-Lexicon Medical Workbench for Swedish". In *Euralex proceedings 2008*. Barcelona.
- Lindskog, B. (2004). *Medicinsk terminologi*. Stockholm: Norstedts Akademiska Förlag.
- Malmgren, S. G. (2003). "Begå eller ta självmord? Om svenska kollokationer och deras förändringsbenägenhet". In *Språk and Stil. Tidskrift för svensk språkforskning* 13. 123-168.
- Martin, W. (1992). "Remarks on Collocations in Sublanguages". In *Terminologie et Traduction* 2-3. 157–164.
- Martin, W. (2006). "Frame-based Lexicons and Making of Dictionaries". In Corino, E.; Mareello C.; Onesti, C. (eds.) *Proceedings of the XII Euralex International Congress, Torino, Italia, September 6–9, 2006*. Alessandria. 281-293.
- Medical subject headings thesaurus*, MESH [on-line]. <http://mesh.kib.ki.se>. [Access date: 28 March 2008].
- Mel'čuk, I. (1998). "Collocations and Lexical Functions". In Cowie, A. P. (ed.) *Phraseology: Theory, Analysis, and Applications*. Oxford: Clarendon Press. 23-53.
- Nationalencyklopedins ordbok*. Höganäs: Språkdata, Göteborgs universitet, Bra Böcker; 1995-96.
- Siepmann, D. (2005). "Collocation, Colligation and Encoding Dictionaries. Part I. Lexicological Aspects". *International Journal of Lexicography* 18 (4). 409-443.
- Sköldberg, E.; Toporowska Gronostaj, M. (forthcoming). "Modell för beskrivning av kollokationer i ett medicinskt lexicon (MedLex)". *Nordiska Studier i Lexikografi* 9. Akureyri, Maj 2007.
- Språkbanken [on-line]. <http://sprakbanken.gu.se> [Access date: 28 March 2008].
- Svensén, B. (2004). *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. [2 uppl.]. Stockholm: Norstedts Akademiska Förlag.