To appear in: G.Thompson and S.Hunston, (eds.) *System and Corpus: Exploring Connections.* London: Equinox

How to handle lexical semantics in SFL: a corpus study of purposes for using size adjectives

Serge Sharoff*

Fakultät für Linguistik und Literaturwissenschaft, Universität Bielefeld,

Postfach 10 01 31, D-33501 Bielefeld, Germany,

tel: +49-521-1065323; fax: +49-521-1066447

e-mail: serge.sharoff@uni-bielefeld.de

Abstract

In the systemic framework, lexis has received little attention in comparison to grammar. The paper compares existing systemic approaches to lexis and outlines a model for describing lexical items from the viewpoint of realization of communicative intentions by lexical items using the systemic network of lexical choices. As an example of the approach, the paper attempts to describe contrastive lexical semantics for a set of words, namely, size adjectives in English, German and Russian. Unlike several other treatments, which focused mostly on physical properties of objects designated by size adjectives, for instance, works by Bierwisch and Tucker, the proposed model accounts for basic purposes with which size adjectives are used. First, the investigation goes beyond physical properties of objects, since size adjectives are used for many different purposes, including specification of intensity, number of elements, importance, etc (statistically, they are more frequent than spatial uses). Second, size adjectives are chosen by the speaker to achieve a rhetorical impact on the hearer, so the investigation goes beyond specifying ideational properties alone. The paper considers the use of the systemic network for studying translation equivalence between uses of size adjectives in the three languages by comparing combinations of features in the network for respective translations.

1. Introduction

In the systemic framework, lexis has received little attention in comparison to grammar. In the beginning of "Introduction to Functional Grammar" (IFG), Halliday writes:

> in order to make explicit the fact that syntax and vocabulary are part of the same level in the code, it is useful to refer to it comprehensively as 'lexicogrammar', but it becomes cumbersome to use this term all the time, and the shorter term [grammar] usually suffices. (p. xiv)

After this remark he continues with exploring topics related to the grammar in more narrow sense and pays much less attention to lexical items.[1] The same is true for the work on the NIGEL lexicogrammar (Matthiessen, 1996) and many other studies.

* The research presented in the paper has been supported by the Alexander von Humboldt Foundation, Germany. I'm grateful to John Bateman, Dafydd Gibbon, Peter Hellwig, Susan Hunston, Irina Kobozeva, and Ekaterina Rakhilina for comments and discussions.

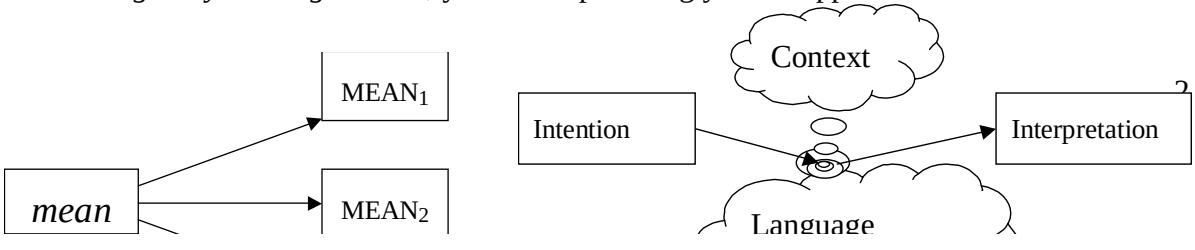
The formal model: meaning as concept	The functional model: meaning as use

Figure 1. The difference between formal and functional approaches to lexical semantics

One comprehensive account of the English lexicon, which is functional and cognate to systemic linguistics, is the Collins Cobuild English Dictionary (CCED) and the pattern grammar approach (Hunston, Francis, 1999) which followed from it. However, descriptions in a human-oriented dictionary lack formal mechanisms for dealing with meanings of lexical items in a computationally tractable way, i.e. such descriptions cannot be directly used in computational applications for language understanding and generation or machine translation.

Some other scholars have described the behavior of lexical items using systemic networks as the formal mechanism. This includes the study of lexis as *most delicate* grammar (Hasan, 1987), the study of lexical options for adjectives *within* the lexicogrammatical network (Tucker, 1998) and the polysystemic model, treating lexis as a *separate* stratum (Wanner, 1997). These approaches address some properties of the behavior of lexical items, but remain restricted in certain respects. Hasan (1987) treats few lexical items, like *strew, spill,* etc, which are also quite rarely used. Tucker's model covers a wide range of adjectives and includes a description of size adjectives, which provided the starting point for our investigation, but it suffers from cursory treatment of the complete range of uses of size adjectives, like *a broad coalition, a little girl,* or *high interest rates*. Wanner's model uses a separate stratum of lexis.[2] In the context of machine translation or generation, it provides resources for converting language-independent situation specifications into lexicalized semantic specifications, while the stratum of grammar provides resources for converting lexicalized specifications into strings. For instance, the situation specification BATH(X), where X is the actor, can be converted to two lexicalized specifications, one with the non-directed process *bath*, and one with the directed process *take* and the goal *bath*, each of which can be used by the grammar to produce *X bathed* vs. *X took a bath*. However, the separation of the lexicon and the grammar can lead to reduplication of grammar-induced properties in the lexical network and vice versa.

However, there is one criticism applicable to these three approaches at once: even though they belong to systemic-functional linguistics, they are not functional enough, in the sense that they do not relate uses of lexical items to their communicative functions. In this respect they differ from definitions in CCED, which use special presentational devices for describing the relationship between form and function; for instance, one type of using *big* is explained in it as: "If you call someone *a big bully* or *a big coward*, you are emphasizing your disapproval of them".



English (23): *big, brief, broad, deep, fine, great, high, huge, large, little, long, low, major, minor, narrow, short, slight, small, tall, thick, thin, tiny, wide*

German (21): *breit, dick, dünn, eng, fein, gering, groß, hoch, klein, knapp, kurz, lang, mäßig, nieder, niedrig, riesig, schmal, tief, weit, wenig*

Russian (22): *boljshoj, dlinnyj, dolgij, glubokij, gromandyj, korotkij, krupnyj, malenjkij, malyj, melkij, menjshij, neboljshoj, nichtozhnyj, nizkij, ogromnyj, shirokij, tesnyj, tolstyj, tonkij, uzkij, velikij, vysokij*

Table 1. Most frequent size adjectives in English, German and Russian

The relationship between lexical form and function can be clarified by viewing the opposition between attitudes of formal and functional linguistics to the subject of their study. Formal linguistics treats language as a set of rules for building well-formed structures, while functional linguistics treats it as a system of resources for enabling interaction between the two parties. In the case of lexical semantics, the first approach assumes that a word *has* a meaning, which is defined as a dictionary sense and can be represented as a concept in some knowledge representation language (the left side of Figure 1). The second approach assumes that the meaning of a word is a function of its use in purposeful communication. The speaker uses resources of language for expressing communicative intentions according to the context of situation. Physical signals, which are realizations of intentions, are interpreted by the listener because, in the case of successful communication, the set of language resources and the context of situation are to a significant extent shared by the two parties (the right side of Figure 1). In this view, the task of a study in lexical semantics is to describe how lexical resources contribute to the realization of communicative intentions. For a longer discussion of differences between the two approaches see (Sharoff, 2002a).

The paper presents an attempt to study uses of size adjectives in English, German and Russian in their relationship to communicative intentions of the speaker. The paper starts with a list of size adjectives to be studied and problems with their descriptions in existing studies. Then, I will present the systemic network for describing uses of the size adjectives in texts and the trinocular view on the lexical network: "from around", "from above" and "from below". Finally, I will discuss types of translation equivalence and differences between translations, when uses are described by means of the lexical network.

2. The scope of the study

2.1. The list of size adjectives

The objective of the study reported here is similar to the objective of IFG to construct a grammar “that would make it possible to say sensible and useful things about any text, spoken or written, in modern English” (Halliday, 1985: xv). Similarly, my aim was to describe the most frequent uses of the most frequent size adjectives in English, German and Russian. The study considers all the size adjectives that belong to a frequency list that would cover more than 75% of word uses in a representative corpus. The threshold of 75% is also used in CCED as the definition for the core lexicon (such entries are marked with 5 or 4 diamonds in CCED). The current estimation is that about 2000 words in a language are within the 75% threshold.

The English set is based on frequency data from the word list compiled by Adam Kilgariff on the basis of the British National Corpus (BNC). The German data are based on the frequency from the Münster corpus (XLEX). Since no exact information on words with 75% coverage is available for German, the selection used words which frequency is more than 40 ipm (instances per million words), since the frequencies of respective items in the English and Russian lists are roughly similar. The Russian data are taken from a frequency list that is based on the reference corpus of Russian (Sharoff, 2003).

Selection of words for the lists is not as trivial as it seems. On the one hand, the fact that an adjective, when it is used in a text, denotes a size of an object is a fact related to its meaning in this text. On the other hand, meanings are not readily available in a corpus, from which we can select, for example, expressions with adjectives referring to size. Moreover, there is no exhaustive list of communicative intentions that are realized by size adjectives. For instance, *short* is obviously a size adjective, even including cases when it is used for referring to time, as in *short stay*. What is more, the latter use is systematically related to size, so it should be considered in the treatment of size adjectives. However, this means that the complete description of size adjectives could cover adjectives that cannot refer to size proper, but only to temporal properties, e.g. *brief*. Thus, *brief* should be also included in the description. Incidentally, several types of uses of other words referring to temporal properties can also be included, for instance, *young* and *old*, because they are related to uses of size adjectives: *his little brother, two small children*.

When such words as *fine* are considered, the choice is less obvious. Even though, it is not considered as a size adjective, it is frequently used for this purpose, e.g. CCED definitions "Something that is fine is very delicate, narrow or small" and "A fine detail or distinction is very delicate, small, or exact". Such uses are relatively frequent: *distinctions, needle, threads, tuning* belong to the most significant collocates of *fine* according to the Cobuild collocation sampler. Moreover, when uses of translation equivalents of *fine*, are studied in German and Russian, they also often refer to size properties, e.g. *fine distinction* is translated into German and Russian as *feiner Unterschied, tonkoe razlichie*. This provides the reason to include *fine* in the list, but the description is not intended to cover other uses of *fine*, like "You use fine to describe something that you admire and think is very good" (CCED).

2.2. Problems with existing descriptions of size adjectives

WordNet (Miller, 1990) makes full use of the formal model, which treats meanings as concepts (Figure 1). Miller refers to concepts as *synsets*, i.e. sets of synonymous senses, and intends to map *all* English words onto *all* concepts existing in English. For instance, WordNet lists 11 synsets for *little*, including:

1. limited or below average in number or quantity or magnitude or extent;
5. of little importance or influence or power; of minor status;
6. (informal terms) small and of little importance;
8. contemptibly narrow in outlook, e.g. "a little mind consumed with trivia"; "petty little comments";
11. used of persons or behavior; characterized by or indicative of lack of generosity.

It is clear that synsets 5, 6, 8, 11 have something in common, namely, the negative attitude towards the object designated using *little* in this sense, but there are no formal means to relate them in the definition. The formal model also assumes that any given use points to one concept

from the list of senses of a word, unless the use is ambiguous. However, when uses of a word in a text are studied, it is not always easy to select the exact concept in the lists of senses. For instance, *little* is used twice in the following example:

(1) - *Oh, you wicked little thing! - cried Alice, catching up the kitten, and giving it a little kiss to make it understand that it was in disgrace.*

The first use (*little thing*) can refer simultaneously to synsets 1, 5, 6 and 8, even though the example is not ambiguous. *Small* also has 11 synsets, $small_1$ equals to $little_1$, while $small_2$= $minor_{10}$, which is described as "limited in size or scope", which is almost indistinguishable from $small_1$. At the same time, there is no appropriate sense in the WordNet list to classify *a little kiss*, the second use of *little* in (1), which in principle can refer to "below average in magnitude", though there is no unit to measure the magnitude of kisses.

The list of synsets and its relationship to words is language-specific. For instance, English provides an array of words referring to the small-scale generic size specification: *little, small, slight*, and *minor*. They are typically translated into German as *klein* (small). However, we cannot assume that *klein* shares synsets with the four English adjectives, because German has another array of words referring to small amount specification: *gering, klein, knapp, mäßig, wenig*. Similarly, in Russian another array of words is used for the same purpose: *malenkij, malyj, melkij, neboljshoj*. The words in the German and Russian lists have different uses, which do not fit nicely into the synsets available for the four English words. For instance, *malyj* collocates with *kolichestvo* (small amount), *predprinimatel'stvo* (small business), *ploschad'* (small area), *sily* (little efforts), *skorost'* (low speed), *vysota* (low altitude), *moschnost'* (low power), *srok* (short term), etc, but WordNet has no synsets that include both *small* and *low*. This requires several separate Russian synsets, which differ from synsets available for English.

Unlike WordNet, which aimed at listing all the senses possible for a word, other studies of size adjectives focused mostly on physical properties of objects designated by size adjectives. For instance, Tucker (1998: 138) proposes a systemic network for describing size adjectives starting with four features: three dimensions (height, length and width) and generic size references. However, the description is not complete. First, it makes no reference to non-spatial senses, including those for specifying *little* in the example (1) and non-dimensional uses, like *great, major*, etc. The network also misses several important dimensional adjectives that are listed in Table 1: *broad, deep,*[3] *thick, thin*. Finally, it does not address restrictions on spatial uses of size adjectives. Every physical object has three dimensions, so one can describe it in terms of its length, width and height. However, not all references are equally possible. A house, as a physical object can be referred to as *long* or *high/low*, but it is almost never referred to as *short*. At the same time, a lying pole can be described as *long*, while the same pole, when it is in upright position, is either *long* or *high* (here, both *long* and *high* refer to the same dimension). On the other hand, even though, the shape of a cigarette closely resembles the one of a pole, it is impossible to refer to a cigarette as *high*, even it is held vertically[4]. Similarly, a rope attached to a high point shares all the spatial properties of a pole, but it cannot be regarded as *high* in English. Width is typically defined as the second dimension (the assumption is that the largest dimension is referred to as length), but there are many cases, when *wide* refers to the largest dimension, as in *wide table*.

At the same time, in spite of some problems with covering the set of real uses, the network of size adjectives proposed in (Tucker, 1998) is useful as the starting point for the investigation of their uses. It was extended by gathering all potentially relevant frequent size adjectives (listed in Table 1) and by studying respective monolingual corpora, as described in the next section.

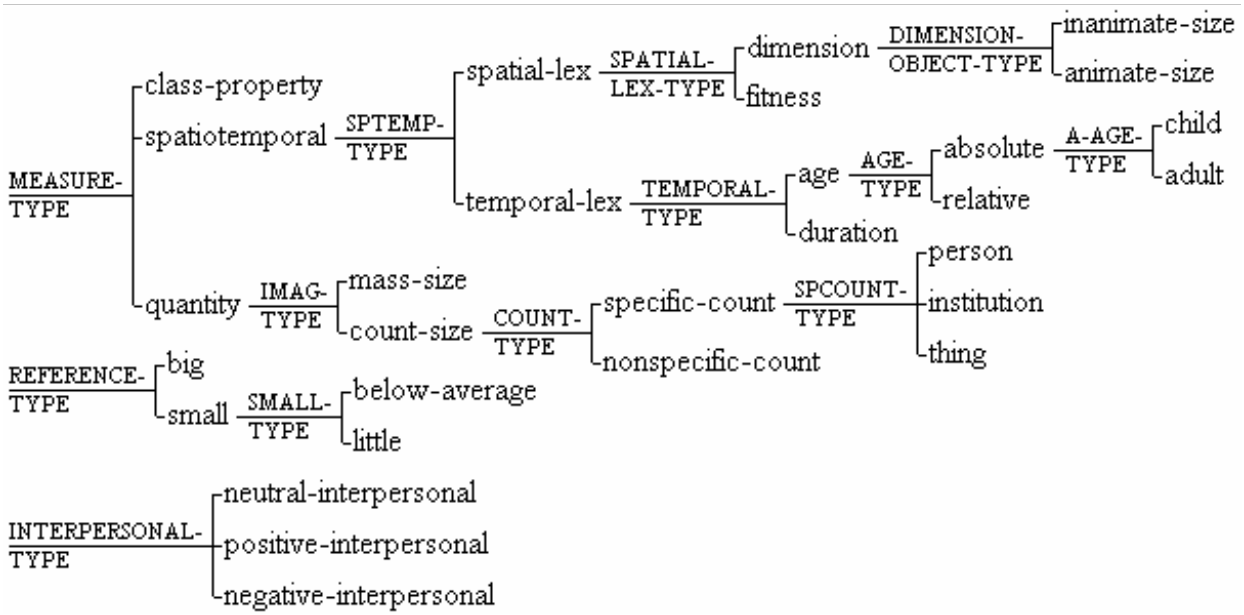


Figure 2. Basic options for uses of size adjectives

3. The systemic network for the most frequent uses of size adjectives

The section presents a systemic network that should describe the most frequent uses of size adjectives in the three languages. According to the same principles as used for development of systemic grammars (Martin, 1987), a feature in the lexical network is motivated, if it has some reflex in form, i.e. if several lexical items are available for expressing related, but different communicative intentions, they correspond to several distinct features. For instance, the height of an object can be expressed in English using *high* and *tall*. Since *high* cannot be used for referring to the height of animate objects, the distinction between animate and inanimate objects should be presented in the network. A feature is also motivated in the lexical network, if there are constraints on realization following its choice, for instance, different co-occurrence patterns. In principle, each type of use should correspond to a unique set of features in the lexical network. Feature names in the discussion below are enclosed in square brackets.

Basic options of the lexical network that describes the most frequent uses of size adjectives are presented in Figure 2. The network starts with three systems: MEASURE-TYPE, which describes what is measured, when a size adjective is used, REFERENCE-TYPE, which describes the reference system for measurement (whether something is big or small), and INTERPERSONAL-TYPE, which describes the attitude of the speaker towards the object of description. The latter system is relevant for any quality specification and can influence the choice of size adjectives, as it is discussed below.

Another system of options shown in Figure 3 describes the set of dimensions for referring to objects and concepts of various kinds: generic size references ([non-directional], this includes not only *big*, *small*, etc, but also *great*, *huge*, *slight*), references to linear size (*long*), to flat surfaces ([two-dim-size], *wide*), volume references (*thick*) and references to the vertical dimension, which has two possible measurement directions: from the ground upwards ([height-size], e.g. *high*, *tall*) and below the ground or towards an unattainable surface ([depth-size], *deep*). A reference to an unattainable surface helps in describing non-vertical uses of *deep*, like

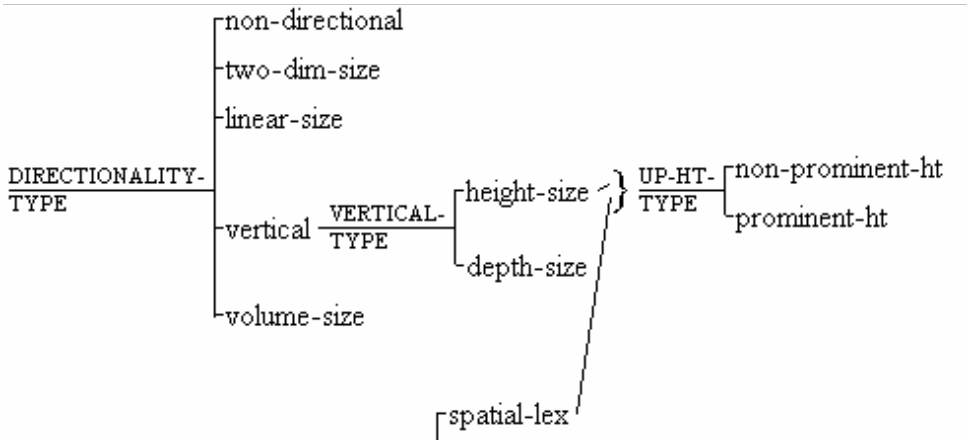


Figure 3. The options for dimensionality

deep wardrobes, shelves, borders, etc. All directions can be set for all types of uses of size adjectives with the exception of the distinction between *high* and *tall* in English. Unlike *high*, *tall* is applicable only to spatial measurements and is synonymous with *high* only for inanimate objects, otherwise, it is required for animate objects and has *short* as the antonym.

Studies of size adjectives typically describe their uses referring to physical size, e.g. (Bierwisch, 1987), (Tucker, 1998: 137ff). However, it is just one of multiple options. In the network in Figure 2, it corresponds to the feature [dimension] within subsystems of MEASURE-TYPE. It can be regarded as the primary meaning source, because other types of uses of size adjectives are extensions referring to qualities of other type, as if they were physical size qualities. The metaphor of representing values as size seems to be natural, because the physical size of an object is one of the most evident properties for visual perception, so size properties provide the basis for communicating properties in other domains, in which a parameter can vary in certain respects. A reference to the size in this context implies a map from a qualitative state onto a value, which represents the measure for the state. Thus, words that designate the mapping are naturally applicable both to sizes and quantities.

However, uses of size adjectives for referring to size qualities constitute a small proportion of total uses. For instance, the most significant nominal collocates of *big* are:[5]

> blow, boost, boys, break, breakfast, brother, budget, business, chance, city, club, company, crowd, cut, day, deal, difference, disappointment, event, eyes, fan, fish, football, game, gun, heart, hit, house, impact, issue, mac, man, mistake, money, name, occasion, part, picture, players, problem, question, race, screen, star, step, surprise, thing, ticket, time, toe, word

Only *eyes, gun, house, man, screen, toe* are big in physical size, even *big thing* and *big step* are used in reference to the importance of a problem or a decision.

In addition to [dimension], another possible spatial use of size adjectives is to refer to the degree of [fitness], when an object (often the human body) fits into the space in another object (often, a piece of clothing). This often leads to specific lexicogrammatical restrictions, e.g. for clothing *weit* vs. *knapp* in German, *velik* vs. *mal* in Russian (only the short forms of adjectives are used in this sense); these pairs correspond to *loose* vs. *tight* in English. In Russian, this option is also used for coding the living space *tesnyj* (cramped), *prostornyj* (spacious)[6].

Within the spatiotemporal measurements, [spatial-lex] and [temporal-lex] features are distinguished. In language (at least, in the languages under consideration) temporal qualities are often expressed by the same lexical means as spatial ones, e.g. *This spans a large period of time*,

	count with *small*	count with *little*	total count
boy	228	686	13290
boys	113	143	8054
girl	50	996	15762
girls	24	239	9621

Table 2. Frequency values for *small* and *little* collocating with children.

they also share such spatial properties as source and destination, e.g. *from April to June*. Temporal qualities that can be expressed by size adjectives refer either to one's age ([age]) or to a time interval ([duration]). In the former case the distinction is between the age specification in relative terms, e.g. *his little brother*, i.e. younger than the referent, and in absolute terms, e.g. *I have a little boy of 8*. Finally, there can be a difference in referring to the age of a child or an adult: *groß* in German and *boljshoj* in Russian are applicable to a child, but not to an adult in the sense *Das Kind ist schon groß geworden, Rebenok uzhe bolshoj* (the child has grown up). Similarly, *little* in English is applicable to a child, while *young* to an adult. Duration in the three languages is typically described by size adjectives in one-dimensional terms, e.g. *short* (*a short stay*), *kurz, korotkij*, or without a reference to dimensions at all, by means of adjectives applicable only to the duration, e.g. *brief* (*a brief visit*), *kratkij*.

Quite frequently size adjectives are used to refer to [quantity] with two possible features of [count-size] and [mass-size] measurements. The latter are most typically realized by [vertical] size adjectives, like *high* and *low*. The [count-size] measurements collocate with *number* as well as with more or less specific designations of groups of various kinds by means of generic size adjectives: *a small group of students, a big crowd*. Yet another possibility for expressing [quantity] is realized by grammatical means: *a lot of, many, some*. They can be linked to existing options in the grammatical network for quantity specification (Matthiessen, 1996), (Tucker, 1998: 119ff).

The third basic option in the system MEASURE-TYPE is [class-property], which does not specify a measurable parameter, like [quantity], but the class of objects that are semiotically considered as large or small. This feature covers uses of *little kiss* and *little thing* in the example (1), as well as *big names, low achievers, deep feelings,* etc. In several cases, the feature controls the choice between near-synonymous lexical items. For instance, *a large fish* refers to the physical size of a particular fish, while *a big fish* belongs to the class of big fishes (also, figuratively). The expression *a large city* is not idiomatic; it is normally used in expressions referring to the number of people living in it or the area it occupies. Uses of *a big city* are different. CCED defines as:

> The big city is used to refer to a large city which seems attractive to someone because they think there are many exciting things to do there, and many opportunities to earn a lot of money.

So, in the case of [non-directional] size specifications (Figure 3), the feature [quantity] is preferably realized by *large,* while the feature [class-property] by *big*.

The choice of a lexical item can also be influenced by the speaker's evaluation of an object. WordNet defines a synset *little=small=young* for denoting young children. However, *little* is

typically used, when the author announces the positive attitude to a child being described. On the contrary, *small* is typically used in less favorable contexts. For instance, *little* collocates with *girl* much more strongly than with *boy*, while *boy* collocates stronger with *small*, cf. Table 2, in which the first two columns present the number of the joint occurrences of two words, and the last column presents the total number of occurrences of the respective nouns in the BNC. The difference in the collocation frequency does not necessarily imply the difference in semantics of words *boy* and *girl*. However, the difference is based on the fact that girls are more often referred to in the paternalistic tone, i.e. the word *girl* has the greater probability to be used in conjunction with the interpersonal meaning of sympathy. Thus, it is more often realized with *little* than with *small*. Also, the plural noun forms for *girls* and *boys* correspond to about 61% of uses of the singular form, but the plural compound *little girls* corresponds to just 24% (21% for *boys*) of respective singular uses. There is no such a big drop for *small* (about 50% for both *boys* and *girls*). This can be also explained by the fact that the greater interpersonal value is expressed in the exchange about a particular child, not about children collectively. The difference between patterns of uses of *little* and *small* in English is also discussed by Stubbs (2001), who compares frequent positive collocations of *little* to predominantly formal contexts in which *small* is used (*pretty little girl* vs. *comparatively small quantity*).

Most often the choice of an adjective chosen in [class-property] is idiomatically linked to the noun it modifies: *fine detail, short temper, narrow views,* etc. A useful tool for describing such idiomatic relationships has been introduced by Mel'chuk (Mel'chuk, Polguère, 1987). He speaks of lexical functions, which are functions (in the mathematical sense) that associate keywords with their values. The keyword of a lexical function is a lexical expression, and its value is another lexical expression. [7] The set of about 60 lexical functions used by Mel'chuk includes `Magn`, the lexical function for expressing the intense degree, e.g. `Magn`(*range*)=*wide*, `AntiMagn`(*range*)=*narrow*. The notion of lexical functions can be profitably used in systemic studies as a method for describing the contribution of lexical expressions. For instance, in

(2) *Factory shops contain a <u>wide range</u> of cheap furnishings,*

`Magn` is a functional label that refers to the size of a collection of furnishings. Another lexical function is `Bon`, a positive evaluation of an object, e.g. `Bon`(*aid*)=*valuable*.

The model based on the systemic network can be used for descriptions that are similar to lexical functions. In systemics, the influence of one word on the choice of another word can be described by preselecting features in the network (Halliday, Matthiessen, 1999: 43-44). In the case of size adjectives, dimensional options in the network are preselected according to lexical semantics of a noun. However, the systemic network can do more than lexical functions do. A lexical function offers the word-to-word mapping, because it simply links two lexical expressions: its output is "all or nothing" without any relationship to possible communicative intentions of the speaker or the context. Also the theory of lexical functions provides no way to express the network of relations between them, for instance, it cannot describe uses that are simultaneously `Bon` and `AntiMagn`, as it is the case in the distinction between *small* and *little* denoting young children. At the same time, the lexical choice in the systemic network is controlled by combination of several features, including, for example, rhetorical properties of the interaction between the speaker and the hearer, as defined by the system INTERPERSONAL-TYPE in Figure 2. The system of such choices is not restricted to size

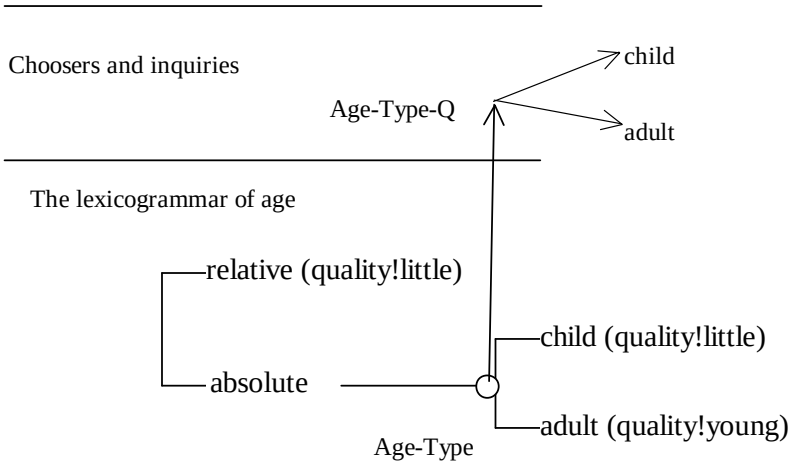


Figure 4. The three views on the lexical network

adjectives, more delicate options in other cases can be related to a more general system of APRAISAL (Martin, 2000).

Another system of options, which is not shown in Figure 2, refers to the level of emphasis. It is applicable to any quality group and controls either grammatical or lexical means for emphasizing the value of the quality. In the three languages, only generic size specifications have lexicalized emphasis: *huge, tiny, great, immense*, etc, while the size in specific dimensions can be emphasized using grammatical means at the level of the adjectival group: *extremely deep, awfully high, very long*, etc. Lexical items used for emphasizing the size cover a range of other options in the network, for instance, *a huge piece of canvas, an immense cloud of smoke*, *a tiny living room* ([spatial-lex]), *huge profits, a tiny fraction of all home owners* ([mass-size]), *a huge problem, great cultural achievements, a tiny glimmer of hope* ([class-property], this is the most frequent type of uses).

4. The trinocular view on the lexical network

Halliday and Matthiessen (1999: 504) discuss the following trinocular view on the lexicogrammar:

1) "from around" to consider what choices are available in the lexicogrammar, for instance, for [declarative] and [interrogative] for the system of indicative clauses;
2) "from above" to look at the distinction in meaning which is reflected in the choice between [declarative] and [interrogative] in the system;
3) "from below" to consider differences between [declarative] and [interrogative] in terms of their syntagmatic realization.

The same trinocular view can be applied to the lexical network. The section above considers options for using size adjectives by comparing, for instance, the distribution of options in [temporal] expressions between [age] and [duration], and for [age] expression between [relative] (*my little brother*) and [absolute] (*a little boy of 8*). This corresponds to the view "from around" the lexicogrammar.

We can also consider realization of features like [absolute] or [relative] age specification by specific lexical items and/or grammatical constraints on the structure of the nominal group. For instance, the [relative] age specification is realized by *little*, while the [absolute] age specification can be realized by *young* for an adult and either *little* or *small* for a child, depending on preference

options, such as the interpersonal attitude and other restrictions, i.e., *small* refers more to the size of a child, while *little* to the age. This means that we have three lexical items *little, small* and *young,* which uses can be studied in texts as related to the features in the network. This corresponds to the view "from below" the lexicogrammar.

Finally, choices between features are based on semantic distinctions between meanings intended by the speaker according to their relationship to lexicogrammatical options available in language. In the case of the lexicogrammar of age, the choice between the two features [child] and [adult], ultimately the choice between *little* and *young,* depends on the age of a person in question (Figure 4). This corresponds to the view "from above" the lexicogrammar.

The next subsections consider the views from above and below in greater detail.

4.1. The view "from above" as the chooser-inquiry interface

Many systems in the lexicogrammatical network allow a simple mapping from features to reasons for choosing the features. When the lexicogrammar is implemented in the framework of KPML/PENMAN, two computational constructs are used: *choosers* that make decisions on which feature in the network is to be selected, and *inquiries*, computational procedures that act as semantic "experts" for making small-scale decisions that can be done within the semantic environment of the interaction, cf. (Bateman, 1997), (Halliday, Matthiessen, 1999: 374ff). For instance, inquiry Age-Type-Q in Figure 3 makes a decision whether it is contextually appropriate to refer to a given person as a child or as an adult. The inquiry implies that the answer depends not only on one's age, but also on contextual conditions of the interaction. Similarly, the choice between the two features [count-size] and [mass-size] (Figure 2) is based on the answer from an inquiry: whether the object is semiotically presented as a set of distinct elements. Such cases stress the point that the classification of lexicogrammatical choices is a natural extension of semantics.

However, some features in the network have a more complex relationship to reasons for their choice. As an example, consider options for choosing dimensions of measurement from Figure 3: what are reasons distinguishing uses of *large, long, wide* or *high*? When the semiotic dimensionality for the size of a *non-spatial* object is realized, it depends on lexical properties that preselect features in the network. For instance, various types of strong feelings are realized using *deep*, importance is most typically realized non-directionally (*great achievements, big*/*little things, major factors*), etc. Thus, unless otherwise stated, a strong feeling preselects [depth-size], importance—[non-directional] (in terms of features of the network in Figure 3). However, the choice of semiotic dimensions for *spatial* objects is based on various properties of the object to be described: *large balls, long hairs, wide beds*. The choice includes several semantic inquiries that check such properties as:

1. the predominant dimension of an object: whether it exists and, if yes, what is its orientation in the physical space (this is a prerequisite for considering its as large in [linear] or [vertical] dimensions). If no predominant dimension exists, another inquiry checks whether the object has a flat surface or not (this is a prerequisite for considering it as two-dimensional). Otherwise, the size of the object should be described as [non-directional].
2. the functional orientation of the predominant dimension: is the direction relevant for the normal function carried out by an object? (otherwise, the object is considered as long, cf. the example with **a high cigarette*);

3. the anthropocentric orientation of an object: what is its position with respect to a human? Humans typically face objects they operate with, so semiotic dimensions of objects are aligned with respect to the human face and arms. This is the reason for considering *tables, screens, walls* as *wide* along their longest dimension;
4. the shape of an object; for instance, pen-shaped objects are ambivalent with respect to linear vs. generic size references, e.g. *long* vs. *large pens, nails*, when their cross dimension is discussed, it is *thick*, not *wide*, etc;
5. the direction of measurement for vertically oriented objects. The most natural direction of measurement starts from the point of attachment of an object. If an object is attached from their upper end, the measurement starts from above. This is the reason for referring to vertically oriented ropes, icicles, etc, as *long*, not *high*, cf. (Rakhilina, 2000).

4.2. The view "from below" as a lexicographical database

The view from below is natural for corpus-based studies, because in a corpus we can consult uses of lexical forms of words, while lexicogrammatical features and meanings intended by uses are not readily available for inspection.[8] The view "from below" implies classification of various uses of size adjectives in terms of features in the lexical network.

The development of the network starts with basic considerations on the structure of uses of size adjectives. The starting point in the study was the network from (Tucker, 1998). The network was used for annotation of dictionary senses (CCED, Wahrig, Ozhegov for English, German and Russian respectively) and examples of uses of size adjectives by means of the Systemic Coder (O'Donnell, 2002). The Systemic Coder is a text markup tool that greatly facilitates annotation of texts using features from the systemic network. Traditionally it has been used for annotating texts with grammatical features, but its supports arbitrary networks of choices.

Each example was annotated with a set of features in the lexical network. If no feature in the network was adequate to reflect a use, the network was extended or reorganized to cover the example. Thus, the network has been built up inductively and the resulted network is quite different from the original one, but it is aimed to cover the most frequent uses and senses detected in dictionaries. The task of handling a relatively large number of words (66 for the three languages in total) with a large number of polysemous senses (356) requires a database, which can help in consulting the uses of words and comparing conditions in which they occur. The database is represented in the XML format and inherits the TEI guidelines for encoding printed dictionaries (Sperberg-McQueen, Burnard, 2001), because TEI is a well-established and widely used format. However, the purpose of the TEI dictionary section is to provide standard means for encoding any information available in *existing* dictionaries. Lexicographical databases require additional means, because they are aimed at development of *new* descriptions of behavior of lexical items. The TEI guidelines also provide no means for relating uses of words to communicative intentions. For more information on means available in the database, see (Sharoff, 2002b).

The database allows to check in which cases a lexical item is annotated with a feature and what lexical items are annotated with a feature. For instance, the user can ask a query to the database for size adjectives that can refer to duration. The output lists *long, short, brief* and, surprisingly, *little*. The latter occurs in such contexts as *a little laugh, a little talk,* etc. The database also helps in comparing uses of lexical items with identical sets of features. When such a query is applied to one language, the database outputs cases of context-dependent synonymy (the synonymy can be

restricted by collocations with respective nouns). When it is applied to several languages, the database outputs cases of context-dependent translation equivalence in the three languages, even if the original annotation was based on *mono*lingual corpora, for instance, the database outputs cases of uses of German *kurz* and *klein* and Russian *korotkij, kratkij* and *malyj* as possible translation equivalents for uses corresponding to *brief, little, short* referring to duration in various contexts.

5. Representing translation equivalence in systemic networks

The contrastive study represented as a systemic network can be helpful for the tasks of machine or computer-aided translation and multilingual generation. The process of machine translation starts "from below": each lexical item instantiates a set of features in the network that can be potentially related to meaning intentions responsible for selection of the features. The amount of possible features and meaning intentions is vast, because of high polysemy of size adjectives. For instance, the [little] and [non-directional] features are always selected for *little* in the REFERENCE-TYPE and DIRECTIONALITY-TYPE systems, but when the program selects the most appropriate interpretation for *little* within the MEASURE-TYPE system in Figure 2, it should choose from the following set: [class-property], [animate-size], [inanimate-size], [absolute-child-age], [relative-age], [mass-size], [count-size], [duration]. However, many features from the MEASURE-TYPE system can be filtered out using restrictions imposed by the modified noun. For example, a reference to an object of the class of children (*girl, son, child*) sets [absolute-child-age] as the most probable feature; alternatively, a reference to *a little table* leaves the only possibility of [inanimate-size] specification, a reference to *money* leaves the only possibility of [mass-size] specification, etc. When a pronoun is modified (*When I was little*), the task is to relate known meaning intentions about the object identified by the pronoun, as well as the reference time used in the clause, to the most probably feature in the set. In the case of a machine translation system, the result of understanding is represented as a list of inquiries with answers ranked in terms of their probability. For translated texts, we can also assume that the set of meaning intentions is the same for a source text and its translation, though we cannot assume this in general (even for texts written in the same genre), because sets of meaning intentions may differ for sociocultural patterns of uses for different languages, in particular, for interpersonal options.

Under this assumption we can use a uniform set of meaning intentions for guiding generation of texts into other languages. Generation should not necessarily lead to the choice of the same options in the systemic network, as it was for the source text. There are three levels to accommodate for multilingual differences, corresponding to three views on the lexicogrammatical network:

1) differences in lexicogrammatical realization ("from below"**)**

Of course, different lexical items are chosen for different languages, though this also includes differences in the structure of the quality group or realization at different ranks, for instance, the most natural way to translate *little kiss* in (1) into German is by means of *Küsschen*, a word for *kiss* with a diminutive suffix.

2) differences in available lexicogrammatical options ("from around"**)**

This involves two options: 1) some features in the network are more or less specific for a language and 2) the same set of features is available for several languages, but different features are selected for a specific language. With respect to the first option, German and Russian lack resources for specific references to the height of animate beings, like the system distinguishing

high and *tall* in English. This means that the option animate/inanimate in Figure 2 is not effective for German and Russian.

As for the second option, concerning features shared between the languages in the study, one and the same feature is selected in similar situations in the three languages, i.e. size adjectives, like *large, high, long, wide,* regularly correspond to *groß, hoch, lang, breit* and *boljshoj, vysokij, dlinnyj, shirokij,* respectively. This also concerns many non-spatial senses, like *high quality, hohes Qualität, vysokoe kachestvo*, or *long list, lange Liste, dlinnyj spisok*. For each dimension, say, *deep,* the correspondence between features covers wide ranges of uses, e.g. *deep appreciation, contempt, cut, mourning, sleep, sigh* are rendered in German and Russian in a similar way: *tiefe Anerkennung, Verachtung, Schnitt, Trauer, Schlaf, Seufzer* vs. *glubokie priznateljnostj, prezrenie, nadrez, traur, son, vzdoh*. However, there are few cases, in which another dimension is selected. For instance, *deep delight* is rendered in German and Russian in terms of non-directional properties: *große Freude, ogromnoe naslazhdenie*. In such cases, the modified noun (*delight, Freude, naslazhdenie*) controls the expression of intensity by preselecting options in the network independently from the set of meaning intentions.

Some properties expressed by a size adjective in one language can be rendered as a measurable quality without a reference to the size, cf. the example with *high wind* vs. *starker Wind* and *siljnyj veter* (lit. strong wind). Other examples of this kind are: *low pulse, schwacher Puls, slabyj puls* (lit. weak pulse), *low visibility, geringe Sicht* (small sight), *plohaja vidimostj* (bad sight). Such cases are also modeled by preselection of specific features in various regions of the network (in this case, outside the domain of size specification)

3) differences in underlying meaning intentions ("from above"**)**

The option is least frequent for the three languages. It means that there are significant differences in meaning intentions concerning frequent non-idiomatic uses of size adjectives. Given that the study deals with three Standard Average European languages that share a great deal of metaphorical means for talking about things, such differences are rare, but they exist. In particular, there are two possible ways for referring to the smaller side of the scale in German and Russian, thus, the REFERENCE-TYPE system in Figure 2 should be more delicate. Some adjectives mean really small, e.g. *klein, malenjkij, uzkij*; they are also regularly used as translation equivalents for *small*. Other adjectives are used for the description of objects that are smaller than their large counterparts, and are below the average or expected value for them, though they do not reach the smaller side of the scale, e.g. *gering, mäßig, neboljshoj, neshirokij*. When the option is applicable, the choice is made with respect to several criteria:

1. is it important for rhetorical reasons to distance the description from the smaller side of the scale? If yes, choose [below-average].
2. the size of the object in general: is the object inherently larger than a human? This concerns such objects as mountains, boulevards, airports. If yes, choose [below-average].
3. the size of the object in the current situation: is it possible to consider the object or amount in the particular context as much smaller than usual? If yes, choose [little].

There is also an interpersonal difference, which distinguishes uses of [below-average] size adjectives from respective [little] adjectives. When *uzkij* (narrow) or *nizkij* (low) are used, they often include a negative characterization of a value, e.g. *uzkoe mesto* (bottleneck, lit. narrow place), *nizkoe kachestvo* (low quality). On the contrary, *neshirokij* (not wide) and *nevysokij* (not high) are used, when the speaker's goal is to de-emphasize the negative properties of an object being described, often in comparison to its price: *neshirokij kanal svjazi* (narrow communication

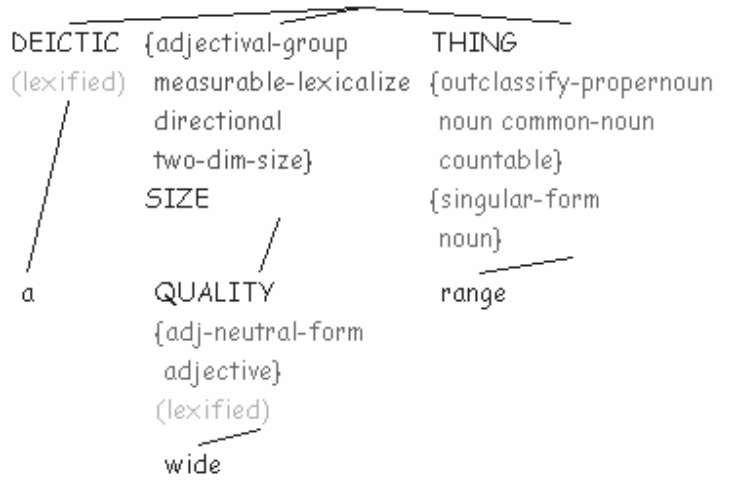


Figure 5 The generation of a nominal group with lexicalized size description

channel), *skanirovanie s nevysokim razresheniem* (low resolution scanning), *nevysokie tseny* (low prices).

The description proposed above has been tested by development of a small fragment of a multilingual generation system aimed at choosing size adjectives in English, German and Russian. The multilingual network has been modeled using KPML (Bateman, 1997) and existing multilingual resources for English, German and Russian (Bateman, et al, 2000). Since no semantic model existed for getting meaning intentions from the environment, or a parser for getting them from texts, input specifications were supplied manually in the format of the Sentence Plan Language, SPL (Kasper, 1989). SPLs consist of concepts and relations as well as inquiry-response pairs that are minimally necessary for specification of the semantic of the utterance to be produced. The following is an SPL example that can be used for generation of size adjectives:

```
(3) (EXAMPLE :NAME   NP-WIDE-2
       :GENERATEDFORM          "A wide range"
       :LOGICALFORM (R / OBJECT :LEX RANGE
          :PROPERTY-ASCRIPTION (Q / QUALITY :REFERENCE-TYPE-Q BIG)))
```

The semantic specification for generation is stored in slot `:logicalform`. The lexical realization of the basic concept (its name is R) is specified exlicitly using `:lex RANGE`. Its properties are less explicit: the inquiry-response pair `:reference-type-q big` claims that the value of a concept should be treated in this way. This is the minimal input specification that outputs *a wide range* (the tree of constituents output by KPML is shown in Figure 5).

In the case of uses of size adjectives referring to physical objects, the input SPL provides exact information on possible meaning intentions associated with the object:

```
(4)(EXAMPLE :NAME   NP-WIDE-1
       :GENERATEDFORM          "A wide table"
       :LOGICALFORM (R / OBJECT :LEX TABLE
          :PROPERTY-ASCRIPTION (Q / QUALITY
                 :REFERENCE-TYPE-Q BIG
                 :ANTHROPOCENTRIC-ORIENTATION-Q ACROSS)))
```

i.e. that a table can be measured as big and it is oriented with respect to a human operating with it (in a “complete” generation system such answers should be provided by the environment).

6. Conclusions

The study reported in the paper shows an approaching for describing lexical semantics from the viewpoint of their uses in texts. Even though the lexicographic research aimed at the enumeration of senses provides an invaluable resource for analysis of uses of lexical items in texts, the assumption that each use points to a concept in the list of senses (the assumption is used, for instance, in WordNet) leads to certain descriptive problems. When the list of senses gets more elaborate, more uses become ambiguous, since often they start to refer to more senses from the dictionary. The communication-centered perspective onto lexical semantics shifts the attention from the meaning of a lexical item to possible uses of groups of lexical items. The question of the lexical semantic analysis in this case is not "What is the meaning of a lexical item?", but "How things are meant by lexical items?"

The paper also presents a case study of uses of size adjectives using the systemic network as a test for the proposed model. Four distinctive features of the model are:

- the description is *functional*: lexical items are described in terms of their use for achieving communicative intentions of the speaker;
- the description is *computational*: the model is based on the systemic network, so that the description can be used in computational applications, in particular, for machine translation or multilingual generation;
- the description is *wide*: it covers the most frequent lexical items in the domain (the size adjectives under consideration belong to the core lexicon that covers 75% of all uses in respective languages);
- the description is *multilingual*: it represents similarities and differences in uses of words across (typologically diverse) languages.

7. References

Bateman, J.A., (1997) Enabling technology for multilingual natural language generation: the KPML development environment. *Journal of Natural Language Engineering* 3, 15-55.

Bateman, J.A., Sharoff, S., (1998) Multilingual grammars and multilingual lexicons for multilingual text generation. In: *Multilinguality in the lexicon-II. (ECAI'98 Workshop 13). European Conference on Artificial Intelligence*, Brighton, U.K., August, 1998. 1-8.

Bateman, J.A., Teich, E., Kruijff, G., Kruijff-Korbayová, I., Sharoff, S., Skoumalová, H., (2000) Resources for multilingual text generation in three Slavic languages**.** *Proc. of LREC2000: Language Resources and Evaluation Conference*, Athens, Greece, May 30-June 2, 2000. 1763-1767.

Bierwisch, M., (ed.) (1989) *Dimensional adjectives: grammatical structure and conceptual interpretation*. Berlin: Springer Verlag.

Dirven, R., Taylor, J. (1988) Conceptualization of vertical space in English. In *Topics in cognitive linguistics*. ed. by Brygida Rudzka-Ostyn, Amsterdam: John Benjamins, pp. 379-402.

Halliday, M.A.K. (1994) *Introduction to Functional Grammar*. 2nd edition. London: Edward Arnold.

Halliday, M.A.K., Matthiessen, C.M.I.M. (1999) *Construing experience through meaning: a language-based approach to cognition.* London: Cassell.

Hasan, R. (1987) The grammarian's dream: lexis as most delicate grammar. In *New Developments in Systemic Linguistics.* Volume 1. M.A.K. Halliday and R.P. Fawcett (eds.), London: Pinter Publishers. 184-211.

Hunston, S., Francis, G. (1999) *Pattern grammar: a corpus driven approach to the lexical grammar of English.* Amsterdam: John Benjamins.

Martin, J.R. (1987) The meaning of features in systemic linguistics. In M.A.K. Halliday, R.P. Fawcett (eds.) *New Developments in Systemic Linguistics.* Vol. 1. London: Pinter Publishers. 14-40.

Martin, J.R. (2000) Beyond exchange: APPRAISAL systems in English. In S. Hunston, G. Thompson, (eds.) *Evaluation in Text: Authorial Stance and The Construction of Discourse*. Oxford: Oxford University Press. 142-175

Matthiessen, C.M.I.M. (1996) *Lexicogrammatical cartography: English systems.* Tokyo: International Language Science Publishers.

Mel'chuk, I., and A. Polguère (1987) A Formal Lexicon in the Meaning-Text Theory (or How to Do Lexica with Words). *Computational Linguistics*, 13 : 3/4, 261–275.

Miller, G., (ed.) (1990) WordNet: an online lexical database. *International Journal of Lexicography* **3** (the special issue on WordNet).

O'Donnell, M. (2002) Automating the coding of semantic patterns: applying machine learning to corpus linguistics. In: *Proc. 29th International Systemic Functional Congress*, Liverpool, July, 2002. Also, cf. http://www.wagsoft.com/Coder/

Rakhilina, E, (2000). *Kognitivnyj Analiz Predmetnyh Imen: semantika i sochetaemostj*. Moscow: Russkie Slovari.

Sharoff, S. (2002a) 'When I use a word, it means just what I choose it to mean'. In C. Inchaurralde, and C. Floren, (eds.) *Interaction and Cognition in Linguistics.* Hamburg: Peter Lang. 277-293

Sharoff, S. (2002b) Meaning as use: exploitation of aligned corpora for the contrastive study of lexical semantics. In *Proc. of Language Resources and Evaluation Conference (LREC02).* May, 2002, Las Palmas, Spain. 447-452.

Sharoff, S. (2003) Methods and tools for development of the Russian Reference Corpus. In *Proc. of the Corpus Linguistics Conference.* Lancaster, March, 2003.

Sperberg-McQueen, C. M., Burnard, L. (eds.) (2001) *Guidelines for Electronic Text Encoding and Interchange.* http://www.hcu.ox.ac.uk/TEI/P4X/index.html

Stede, M. (1996) Lexical Semantics and Knowledge Representation in Multilingual Sentence Generation. Ph.D. thesis, University of Toronto.

Stubbs, M. (2001) *Words and Phrases: Corpus Studies of Lexical Semantics.* Oxford: Blackwell.

Tucker, G.H. (1998) *The lexicogrammar of Adjectives: a Systemic Functional Approach to Lexis.* London: Cassell.

Wanner, L. (ed.) (1996) *Lexical Functions in Lexicography and Natural Language Processing*. John Benjamins: Amsterdam.

Wanner L. (1997) Exploring lexical resources for text generation in a systemic functional language model. PhD Thesis, University of Saarbrücken.

Sources of frequency information

English

Cobuild Collocation Sampler. http://titania.cobuild.collins.co.uk/form.html#democoll

CCED, 1995. *Collins Cobuild English Dictionary*. 2nd edition. Glasgow: HarperCollins.

Kilgariff, A. 1996. BNC database and word frequency lists.

http://www.itri.bton.ac.uk/~Adam.Kilgarriff/bnc-readme.html

German

COSMAS, http://corpora.ids-mannheim.de/~cosmas/

XLEX, http://xlex.uni-muenster.de/

Russian

Sharoff, S. 2002. The Russian Frequency List. http://bokrcorpora.narod.ru/frqlist/frqlist-en.html

---

Notes

[1] Of course, a comprehensive study of the English grammar cannot completely exclude topics pertaining to lexical semantics, so in (Halliday, 1994) they are addressed to some extent in chapters on types of processes and cohesion.
[2] For another implementation of the lexical resources as a separate stratum, cf. (Stede, 1997).
[3] *Shallow* is not included in Table 1, because it is much less frequent (15 ipm in the BNC). Most frequently, small degrees of depth are addressed in English by means of negation, e.g. *There is a very lovely lough close by called Sweethope, …, not very deep and full of splendid trout.*
[4] The example is from (Dirven, Taylor, 1988).
[5] The data are taken from the T-score list provided by the electronic service of the Cobuild collocation sampler: http://titania.cobuild.collins.co.uk/form.html#democoll
[6] *Cramped* and *spacious* in English are beyond the frequency threshold; *narrow* or *small* are often used in situations, in which *tesnyj* is used in Russian.
[7] For an introduction to the concept and potential uses of lexical functions, cf. (Wanner, 1996).
[8] They can be available in hand-coded small corpora, but unlike lexical items, which form is theory-neutral, the annotation scheme depends on a theoretical model used for annotation. For instance, the semantic concordance of WordNet (a selection of texts manually annotated with WordNet synsets) depends on the assumptions on communication made in WordNet and even on the set of synsets used in a particular version of WordNet. The described database is also theory-dependent.