

# A Model of Idiomaticity<sup>1</sup>

BEATRICE WARREN

## 1. *Defining idiomaticity*

As an introduction, I will offer the following two definitions of idiomaticity:

- (i) natively like selection of expression (inspired by Pawley and Syder (1983))
- (ii) that which one has to know over and above rules and words (inspired by Fillmore et al (1988))

The latter definition breaks with the traditional view that knowing a language involves two types of knowledge: rules and lexical items - period. Although it is common knowledge that there is more to knowledge of a language than dictionary items and syntax, Fillmore's suggestion nevertheless represents a breakthrough in linguistic theory. Surprisingly, the fact is that it is only in the last few decades that we have seen this insight empirically demonstrated and theoretically accounted for.

In this connection it should perhaps be pointed out that we must distinguish between the study of idiomaticity and the study of idioms. Idioms in the sense "opaque invariant word combinations" have been studied by theoretical linguists quite extensively, but these bona fide idioms do not contribute to the idiomaticity of a text in any important way. Presence of such idioms in a text does not necessarily make it idiomatic; nor does their absence make it unidiomatic.

Now, if knowing dictionary items and syntax does not ensure natively like selection of expression (i.e. idiomaticity), this raises the

---

<sup>1</sup> This paper was originally published in the *Proceedings of the Ninth Conference for English Studies*.

question: why not? The answer that a number of linguists have given is: human memory capacity. Bolinger (1976:2) was probably one of the first to point out the influence of memory in shaping natural languages, which was something he considered the then dominant transformational-generative theory had overlooked. Since then a number of linguists have made similar claims, probably independent of each other. Pawley and Syder (1983) point out that certain situations and phenomena recur within a community. It is natural that standard ways of describing such recurrent "pieces of reality" develop. A native speaker of a language will—as a matter of course—have learnt these standard ways of expression which can consist of more than one word or certain clausal constructions. Sinclair (1991) contrasts the open choice principle with the idiom principle. The open choice principle says that syntax is there to specify the slots into which memorised items—normally single words—can be inserted. The idiom principle says that a language user has available to him a large number of memorised semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments. Mel'cuk (1996) suggests that the memorised expressions outnumber single words. Jackendoff (1997:156) likewise points out that there are a vast number of memorised expressions. Thus, he concludes, memorised expressions can hardly be a marginal part of our language. Hopper (1998:166), like Bolinger, objects to the generative approach that stresses the uniqueness of each utterance treating it as if it were completely novel, and suggests that everyday language to a very considerable extent is built up of combinations of prefabricated parts. Langacker (1998:25) makes a distinction between stored low-level patterns, many of which incorporate particular lexical items, and high-level schemas, which are general and productive patterns, but suggests that the low-level structures "do much, if not most of the work in speaking and understanding".

So, summing up: the answer to the question: "Why should we need to know more than words and rules of how to combine them?" is: "Because we naturally memorise what is repeated." Moreover, it is often pointed out that it is also a question of economy of effort. Retrieving more or less readymade combinations of words requires less mental effort than composing an utterance word for word (see, e.g., Wray 2002:92). As will become apparent, I do not think that frequency and economy is the whole truth.

In concluding this introduction, let us return to the characterisation of idiomaticity inspired by Pawley and Syder. That is, idiomaticity consists in knowing what situations and phenomena require standard expressions—

although alternatives are normally conceivable—and in knowing what these would be. This is a general characterisation of idiomaticity. In the following a more precise characterisation will be attempted in which the point of departure is the non-native learner's difficulties in acquiring idiomatic language.

## *2. The model*

A more precise characterisation of idiomaticity could be the following. Idiomaticity involves:

- (i) preferences for discourse structure

The very manner in which information is presented in a text may be language specific. I support this claim in particular on the results of the following three studies: Mauranen (1996), Strömqvist (2003) and Wiktorsson (2003), but no doubt there are others I could adduce.

Mauranen compared Finnish and Anglo-American writers' discourse patterns in academic writing and found—as had been previously established—that: “Finnish writers tend to use less metadiscourse than Anglo-American writers, and to employ final-focus, or inductive, argumentative strategies as opposed to initial-focus, or deductive strategies, which are preferred by Anglo-Americans” (1996:143). Strömqvist with co-workers investigated how motion events were described in narrative discourse. The study involved 17 different languages. It was found that “speakers of Romance and Semitic languages detail relatively little information about direction when they relate the motion event, whereas speakers of Germanic languages detail relatively much information. And speakers of Romance and Semitic show a preference for detailing information about the Source, speakers of Germanic about the Path, and speakers of Slavonic about Goal”. Wiktorsson found that essays written by Swedish university students of English were characterised by writer visibility to a greater extent than comparative essays written by native speakers of English.

Probably because of its elusive character, the importance of this aspect of idiomaticity is emphasised comparatively rarely in teaching students to write a foreign language. Features of this kind are after all

tendencies which we are dependent on expert discourse analysts to be confident that they actually exist.

Idiomaticity further involves:

- (ii) knowledge of language-specific propositional expressions including so-called formal idioms and lexicalised sentence stems

I include in this category proverbs, allusions and clichés etc., which are often included in studies of idioms (see, e.g., Alexander (1978), Makkai (1972:128-129), but also lexicalised sentence stems and formal idioms. Lexicalised sentence stems are defined by Pawley and Syder (1983:192-193) as units of clause length which are more or less constrained syntactically and lexically and which are “not true idioms but rather regular form-meaning pairings”. Formal idioms were first described by Fillmore et al (1988). They are constructions with idiosyncratic meanings that do not derive from lexical items but which are inherent in the syntactic frame of the idiom. An often quoted example is *Him be a doctor*, the frame of which is non-nominative NP + non-finite VP + complement and which expresses incredulity. This particular construction does not specify any particular lexical item. Most of the formal idioms, however, are at least partially lexically specific as demonstrated by the following examples discussed in the literature:

verb *one's way* PP: *John joked his way into the meeting* (Goldberg 1995)

verb [Time-NP] *away*: *John drank the afternoon away* (Jackendoff 1997)

*What is X doing Y*: *What is this scratch doing on the table?* (Kay and Fillmore 1999)

*do a* [proper -NP]: *you could do an Arnold Schwarzenegger, just break the lock!* (Pentillä (ms))

*it+be high time* complement.: *it is high time she did something about it* (Lavelle and Minugh 1998)

The feature that these examples have in common is that the meanings they express are at least partially inherent in the construction. Note also that these meanings tend to be evaluative in character, expressing in particular reprobation (*it is high time that...; what is X doing Y*). They have attracted

linguists' interest not only because of their constructional meanings but also because they often manifest not only syntactic and semantic but also phonological and pragmatic constraints.

From the non-native learner's point of view, idiomatic expressions in this category are possibly comparatively unproblematic. Since they are so idiosyncratic, they are either learned or refrained from. The real stumbling blocks for the non-native speaker are expressions which are condoned by the grammar and standard meanings of words but which nevertheless are not used by native speakers. If there *is* a problem with expressions of this kind, predictably it will occur when a learner attempts to translate verbatim a formal idiom into the target language. A Swedish learner might, for instance, render *Vad var det nu du hette?* with *What was it now that you were called?* instead of *What's your name again?*

Formal idioms tend to be clausal constructions. This is true also of the following group of idiomatic expressions I have singled out as forming a particular group:

(iii) expressions in social interaction.

Examples include *excuse me, can I help you, many happy returns of the day, (I am) sorry, (I beg your) pardon* and many more. These are phrases that are performative in that they are not used *about* particular situations but *in* particular situations. They differ from the expressions in group (ii) not only functionally, but also in that as a rule they are lexically specified (i.e. they are less schematic).

At least the most frequent ones are listable and probably explicitly taught and therefore comparatively well known to the foreign learner. Note that some of these are one-item phrases (although originally probably clausal): *cheers* (when toasting), *speaking* (telephoner). I make this point because it is sometimes claimed that idioms are necessarily combinations of words. Such a view—although not strictly correct from a synchronic point of view—is understandable since knowledge of the combinatory potentials of words to form phrases represents an essential feature of idiomaticity. Hence the fourth feature is:

(iv) combinatory potentials of words

It is well known that knowing a word involves knowing what other words it can combine with to form syntactic units. Verbs, for instance, seek above all nouns as partners, as do adjectives, whereas nouns, apart from verbs and adjectives, often combine with other nouns. I will here concentrate on verb-noun combinations, with particular focus on verb-object noun combinations for reasons that will eventually become evident.

As is also well known, the early transformational-generative linguists fully realised that not any lexical item can fit in the slots that syntax makes available. Verbs had to be supplied with not only *subcategorizing features* but also with *selectional restrictions*. Selectional restrictions specify that the object noun in the case of *read*, for instance, would have to be a piece of writing. Nowadays there is also general agreement that verbs have *argument structures*. A verb such as *run* would have an Agent as a subject argument, a verb such as *sink* would have a Theme as a subject, etc.

Specifying thematic roles and selectional restrictions of verbs involves specifying what I refer to as *generalised meanings*, a notion which will be developed presently. It is not possible to know the meaning of, say, *drink* without knowing that there has to be some agent performing the action of drinking and there has to be something that is drunk and that has to be liquid. So supplying words with features like this prevents combinations such as *colourless green ideas sleep furiously*, and serves to predict what combinations are *possible*, at least in the best case. But it does not account for features of idiomaticity, which involves knowing which particular combinations are *conventional* in a language community although other combinations are conceivable. As has already been pointed out, failing to realise that accounting for what is possible is not "the whole story" has been a sin of omission among theoretical linguists, which only now is beginning to be rectified.

The notion of generalised meanings is inspired by usage-based models of language acquisition, in particular Tomasello's (see, e.g., Tomasello, (2000)). Tomasello maintains that in their early language development children reproduce not adult words but adult utterances. They begin by repeating specific combinations of language. It is only when they have heard the same word in different contexts that they are able to construct some general meaning by abstracting semantic commonalities of these different uses. It is now that they can begin to

produce combinations they have never heard before. In other words, the first step is repeating combinations. Producing unheard combinations is a later development and is evidence that the child has been able to analyse utterances into semantic units and abstract semantic commonalities. This abstracted, i.e. decontextualised and general meaning, is what I refer to as generalised meaning.

The construction of generalised meanings can be illustrated as in Figure 1. The arrows in this figure are intended to symbolise the bottom-up kind of approach involved in constructing a generalised meaning in the case of native learners.

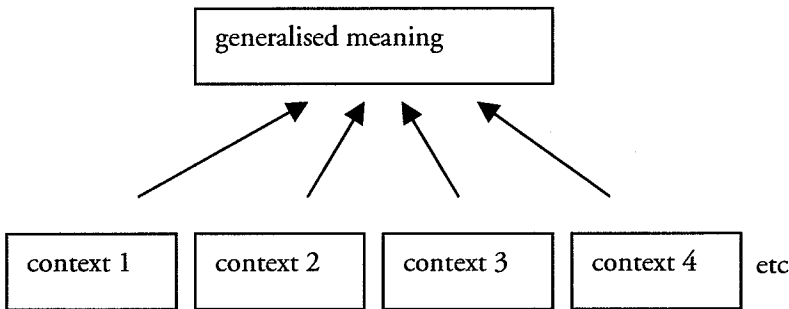


Figure 1. The native learner's construction of generalised meanings

A generalised meaning of a verb will allow any kind of word combination as long as the selectional restrictions and thematic roles specified by this meaning are met. In the case of transitive *drop*, for instance, this would condone *drop a pen, a glass, a key or a piece of amber*, i.e. some novel combination which one nevertheless will recognise as correct, but it would not condone, say, *\*drop love* or *\*drop sunshine*. However, having constructed a generalised meaning does not mean that the language user erases from memory all uses which gave rise to this meaning. Some uses form combinations which will be memorised not only because they are frequent but—I suggest—because they are associated with a certain salient type of situation or phenomenon, i.e. they are often form-meaning pairings and should in my view have the status of lexical items. At any rate, they are generally recognised as more or less fixed phrases which represent language-specific uses. In the case of transitive *drop* they would

include combinations such as *drop bombs*, *drop someone/something at a place*, *drop one's voice*, *drop charges*, *name drop* and *drop a hint*, etc.

However, whereas the native learner will construct some generalised meaning of a word by means of abstracting semantic commonalities of different uses of this word (type frequencies), the non-native learner is likely to construct a generalised meaning by equating it with the generalised meaning of a first language word, i.e. by transfer. That is, the non-native learner's strategy naturally tends to be a top-down approach. Provided that the generalised meanings of first and target language word are indeed equivalent, this will enable the non-native learner to form all the combinations that the generalised meaning condones, but the language-specific uses may be more problematic. (This is illustrated in Figure 2.) For instance, a Swedish learner of transitive *drop* will have to learn, apart from its generalised meaning, also English specialised uses such as the phrases exemplified above (*drop a bomb*, *drop a charge*, *drop a hint*, etc.) and also the manner in which *Swedish* specialised uses are rendered in English (see Table 1).

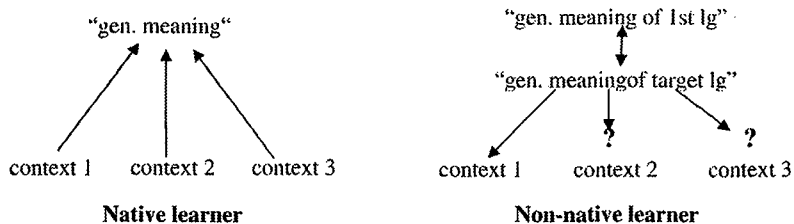


Figure 2. Schematic and simplified representation of the difference between native and non-native learners' acquisition of generalised meanings.

The point I wish to demonstrate is that learning the vocabulary of a foreign language involves considerably more than generalised meanings of single words. Yet generalised meanings are what we teach the learner of a foreign language and are what we test in vocabulary tests. And generalised meanings are what lexicologists focus on, although they have been aware that describing the combinatory potentials of, for instance, verbs in terms of thematic roles and selectional restrictions underrepresents the native speaker's collocational knowledge.



<b>Generalised meaning of English <i>drop</i><sup>2</sup></b> <i>drop a pen, a glass, a key etc.</i>	<b>Generalised meaning of Swedish <i>tappa</i></b> <i>tappa en penna, ett glas, en nyckel etc.</i>
<b>Language specific uses of <i>drop</i></b> <i>drop a bomb</i> <i>drop charges</i> <i>drop a hint</i> <i>drop one's voice</i>	<b>Swedish equivalents (verbatim translations in parentheses)</b> <i>fälla en bomb (fell a bomb)</i> <i>lägga ner åtal (put down charges)</i> <i>ge en vink (give a hint)</i> <i>sänka rösten (sink one's voice)</i>
<b>English equivalents</b> <i>lose one's patience</i> <i>be in a bad mood</i>  <i>lose one's grip/lose control</i> <i>not feel like doing something</i>	<b>Language specific uses of <i>tappa</i></b> <i>tappa tålamodet (drop one's patience)</i> <i>tappa humöret (drop one's good mood)</i>  <i>tappa greppet (drop one's grip)</i> <i>tappa lusten (drop one's inclination)</i>

Table 1. Some examples of *drop/tappa*+object combinations

The descriptions that lexicologists have offered have traditionally involved a threefold division, i.e. open combinations, idioms and collocations as demonstrated in Figure 3.

<sup>2</sup> Arguably transitive *drop* has two generalised meanings:(i) "accidentally let something fall" and (ii) "cause something to fall". The generalised meaning of *tappa* corresponds only to sense (i).

## A Model of Idiomaticity

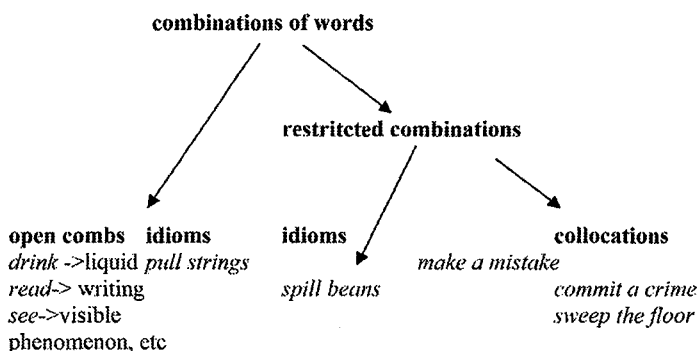


Figure 3. Traditional classification of word combinations

Open combinations are considered productive and compositional and to form the norm. Idioms constitute obvious exceptions since they are neither productive nor compositional. Collocations are often described simply as habitual combinations of words and tend to receive little attention<sup>3</sup>. My version of the native speaker's knowledge of the combinatory potential of words is different. As is illustrated in Figure 4, I suggest the following classification of restrictions: on the one hand, there are words that require a certain semantic profile of their collocate (i.e. grammatical objects in the case of verb-object combinations) and on the other hand, words that require a certain lexical item as their collocate. The first kind of restriction can be exemplified by *look forward to*+ positive situation or *commit*+immoral act. These restrictions represent tendencies, i.e. they may be waived. The latter kind of restriction represent fixed phrases which are stored and which are normally form-meaning pairs. The

---

<sup>3</sup> This is not to deny that there have been attempts to raise the linguistic status of collocations. To my knowledge the first to do so was Lyons, who points out that "it must be remembered that many such phrases (i.e. high frequency phrases, *my addition*) are synchronically speaking, no longer to be considered as units of collocations at all, but as simple grammatical units." (1966:296-297).

Cruse defines collocations as "sequences of lexical items which habitually co-occur" in 1991(p 40). In 2000 (pp 296-297) he does acknowledge that there are arbitrarily restricted collocations which merit inclusion in the dictionary, but leaves it at that.

Allerton (1989: 36), realizing that there are syntactically and lexically unmotivated "locutional co-occurrence restrictions", which a language-user needs to master, suggests that these justify the introduction of "idiomatics" as a special branch of lexicology.

fixed phrases are in turn divided into transparent combinations, which in traditional terminology would be referred to as collocations, and opaque combinations, i.e. in traditional terminology idioms.

Let us first consider the first type of restriction. These types of constraints have been revealed by studies of concordances from large corpora and are sometimes referred to as semantic prosodies. (They have been described by, above all, Stubbs (1995)). Consider as an example *Peter is looking forward to the meeting*. The noun *meeting* is evaluatively neutral, but as a complement of *look forward to* a positive feature is coerced. As just pointed out, these

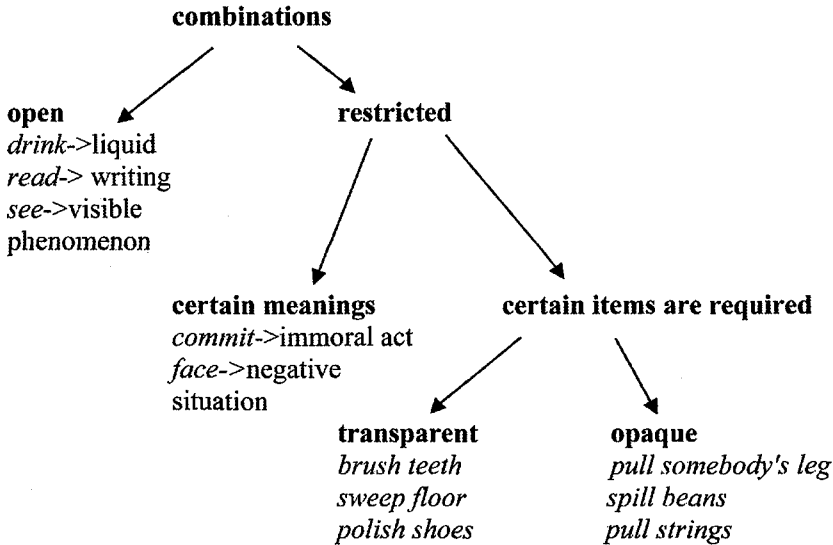


Figure 4. Alternative classification of word combinations

constraints can be cancelled. It is, for instance, possible to modify *look forward to* with the adverbial *with mixed feelings* yielding *Peter is looking forward to the meeting with mixed feelings*, which brings about a change of the interpretation of *meeting*. Some verbs seem to require a more specific semantic character of their objects. *Commit* in the sense of 'do', 'perform' requires that the act carried out is immoral: *commit a sin, a crime, adultery*, etc. The prepositional verb *deal with* in the sense 'be about' requires that the subject represents a 'communicative product' (book, article, talk) and

that the object should be a theme but not just any theme. *The letter dealt with his arrival* would not be normal unless the arrival in question involved some complications. We expect the theme which is the object of *deal with* to be something the relating of which is not quite straightforward.

These required meanings are types of selectional restrictions but differ from what is normally understood by this term in that they are not mandatory and more specific. It is possible—at least for a foreign learner—to feel satisfied that (s)he knows what, say, *commit* and *deal with* mean without fully grasping these kinds of combinatory constraints. A Swedish learner of English, for instance, may very well equate the meaning of *commit* with *begå*. The two words are good translation equivalents. In parallel with *commit*, *begå* combines naturally with the Swedish words for crime, adultery, murder, sin etc. Yet there appear to be differences in their combinatory potentials: In Swedish *misstag* (“mistake”) is a common grammatical object of *begå*, but in English the combination *commit mistakes* seems less natural. Possibly the difference between *commit* and *begå* is that *commit* requires a certain meaning (“immoral act”) of its object, whereas *begå* is less restrictive requiring simply a negative feature of meaning of the object. To develop sensitivity to tendencies of this kind requires a great deal of exposure to a language.

Let me finally point out that the existence of these lexical item+certain meaning combinations may be a reflection of the pattern-creating mental activities which attempt to abstract commonalities among stored expressions and which in the end may affect generalised meanings.

As is illustrated in Figure 4, fixed phrases are divided into transparent (traditionally termed collocations) and opaque combinations (traditionally bona fide idioms), but I would like to emphasise the similarities of these two types of combinations rather than their differences. In my view what collocations and idioms have in common is more important than their differences. Just as *pull strings* is a form-meaning pair representing a particular type of action, which is made evident by the fact that *\*move strings* or *\*pull threads* would not work, so is *brush teeth*. It represents a particular type of action involving a certain type of brush on which tooth paste is spread and which is applied to all the teeth in somebody’s mouth. So, in spite of the fact that both *brush* and *teeth* can be said to have their conventional meanings, the meaning of the combination is not compositional (cf. Fillmore’s frame semantics (1985)). The form-meaning status of the phrase is further made evident in that *polish teeth* or *brush*

*dentals* would either not mean the same or be unidiomatic. If we can agree that idiomaticity represents “nativelike choices of expression”, then *get up in the morning, brush teeth, polish shoes, clear the table, dial a number, get the wrong number*, etc., etc. are as idiomatic as the generally recognised idioms<sup>4</sup>. From a communicative point of view, they are likely to be more important to master than the *bona fide* idioms since they most probably outnumber these both as to their total number and individual frequencies.

The approach forming the basis of the division displayed in Figure 4 departs from the traditional account mainly as far as the status of collocations are concerned. As already pointed out, collocations are traditionally characterised as combinations of words that appear together with greater than random probability. I repeat that frequently they are more than that. They often serve to pick out salient types of situations and phenomena. This in turn amounts to the claim that there are considerably more lexical units in a language than lexicologists and lexicographers account for. One important reason for the undetected lexical status of many collocations is probably their transparency and the fact that they tend to be syntactically unconstrained, in particular verb-object collocations. Transparency is often mistakenly equated with compositionality. True, if some combination is compositional, it is necessarily transparent, but it does not follow that a transparent combination is necessarily compositional.

---

<sup>4</sup> This approach to idiomaticity departs from the common view that the more inflexible and the more opaque a phrase is, the more idiomatic it is. Cowie (1984:x-xiii) and Howarth (1996: 1-47), for instance, suggest a fourfold classification of phrases ranging from least to most idiomatic exemplified in Howarth (p33) by the following combinations:

free collocation	<i>blow a trumpet</i>
restricted collocation	<i>blow a fuse</i>
figurative idiom	<i>blow your own trumpet</i>
pure idiom	<i>blow the gaff</i>

This type of classification is based on the (in my view) mistaken desire “to eliminate from the description (of phrases, *my addition*) those combinations whose occurrence can be accounted for by normal grammatical and syntactic processes” (quoted from Howarth, p47). Syntactic regularity and literal uses of words do not ensure non-idiomaticity. According to the definition of idiomaticity adopted here, the examples above are all idiomatic, also *blow a trumpet*, which implies “play the trumpet” (cf. *blow into a trumpet*).

In view of the multitude of conventionalised phrases a learner of a foreign language has to acquire, it is not surprising that nativelike mastery is difficult to attain. Yet, there are learners who come pretty close to such mastery. In Wiktorsson's study (2003) in which the frequencies of prefabs (i.e. conventionalised multiword combinations) in essays by Swedish university students of English and by native speakers were compared, it was found that there were no differences as to quantity. However, a comparison between essays by less advanced Swedish learners of English (i.e. upper secondary students) and university students showed that the more advanced students were, the more prefabs their essays contained. This suggests, as expected, that the better students are at English, the more prefabs they will know. What may at first blush appear surprising is the fact that upper secondary as well as university students know so many fixed phrases in spite of the fact that they receive little explicit instruction concerning conventionalised combinations of the type *brush teeth*, *clear the table*, *sun rises*. These seem to be picked up subconsciously and fairly effortlessly, probably because the meanings are normally there already<sup>5</sup> and the forms are transparent, which means that there are no new meanings and no new words to learn. What is new are mnemonically motivated combinations of words. It seems then that explicit instructions are not necessary for the acquisition of transparent multiword units. Exposure to the target language, however, is a *sine qua non*.

I hasten to add, however, that not all conventionalised phrases are equally easily learned. It can be hypothesised that phrases containing non-salient and apparently unmotivated items such as prepositions and particles require some effort to be memorised correctly. The same kind of difficulty applies to the delexical verb (*do*, *get*, *give*, *have*, *make*, *put* and *take*) in delexical verb+noun constructions, as pointed out by Allerton (1984:33) and Altenberg and Granger (2001). Also stylistically sophisticated phrases representing abstract events such as *lay down rules*, *exert pressure*, *assume importance* can be assumed to be less easily learned.

This then concludes my classification of idiomaticity features. The reader will hardly have failed to notice a hierarchical organisation going from discourse to phrase level:

---

<sup>5</sup> That learners are aware at some level of the need to find the correct combinations of words for a particular meaning is supported by the fact that users of the *English-Danish Cobuild* dictionary report that they use this dictionary not only for English into Danish translations but for finding the right English collocation (see Zettersten 2002).

- discourse level (i.e. organisation of contents)
- clause level: (i.e. (i) propositional (ii) performative)
- phrase level (i.e. word combinations)

However, it should be admitted that the model leaks. For instance, some of the formal idioms are arguably phrase level constructions, i.e. those in which the subject is not specified and, conversely, intransitive verb+subject combinations are arguably clause-level constructions. Also, there is no hard and fast division between lexical item+certain meaning combinations and lexical item+ lexical item(s) combinations ( see again Figure 4) as demonstrated by the expressions referred to as *prefabs with restricted variability* discussed by Erman and Warren (2000:41) exemplified here by *tappalförloral\*bli av med tålamodet* and *to a great/large/big extent*. The reason behind the hierarchical organization of the model is a matter of presentational clarity rather than a claim as to how the language user mentally organises features of idiomaticity.

### 3. *Some theoretical repercussions*

It should come as no surprise to the reader that an important source of inspiration for the account of idiomaticity in this study has been Construction Grammar. For instance, idiomatic expressions on phrase and clause level fit Goldberg's definition of constructions, which is:

C is a construction iff<sub>def</sub> C is a form-meaning pair  $\langle F_i, S_i \rangle$  such that some aspect of  $F_i$  or some aspect of  $S_i$  is not strictly predictable from C's component parts or from previously established constructions. (Goldberg 1995:4)

Given that the kind of phrasal multiword combinations exemplified above are indeed form-meaning pairs, this will have considerable consequences for lexicology and lexicography. Lexicographers would have to include many more items in dictionaries<sup>6</sup>. Lexicologists can no longer be satisfied with sense relations such as synonymy, antonymy, hyponymy and meronymy. A realistic account of associative links between words in the

---

<sup>6</sup> To a certain extent, this requirement is in practice already met in dictionaries based on large corpora concordances. Again, see Zettersten (2002).

mental lexicon would have to include associations of the kind *tooth: tooth brush, tooth paste and brush teeth; bed: go to bed, be in bed, be ill in bed, bedtime, go to bed with someone*. Above all, they would have to account for multiword lexical items, which is not the same as accounting for single words, since there are important differences between these two types of lexical items. One obvious difference is that single words have no syntactic structure in contrast to multiword combinations which can normally be manipulated (although not always in a uniform manner, which is a further complication). Another difference is that single words are often unmotivated, whereas multiword combinations, excepting bona fide idioms, are motivated. Connected to this is a third difference: whereas it is sometimes possible to replace a standard expression with an alternative descriptive expression, single words are not replaceable in this manner. Such non-standard alternatives appear to be possible to a greater extent in the case of verbal than in nominal multiword expressions. In fact, one reason for the focus on verbal multiword combinations in this survey is that their lexical status is less clearcut than the lexical status of nominal multiword combinations. *Tooth paste* and *shoe polish*, for instance, are normally accepted as lexical units without question, whereas the lexical status of *brush teeth* and *polish shoes* would probably not be as readily recognised. We may tentatively connect this with the fact that nominal multiword expressions tend to denote entities which more clearly represent units than verbal multiword expressions which typically denote transient events extended in time in such a way that it is not possible to perceive beginnings and ends simultaneously.

Such non-standard alternatives may be more or less acceptable to the native ear. Consider, for instance:

(1) *Please, remove the dirty dishes from the table.*

for: *please, clear the table*

(2) *I will adhere to my promise.*

for: *I will keep my promise*

(3) *We related the truth.*

for: *we told the truth.*



(4) *He covered his body with a shirt and a pair of trousers.*

*for: he put on a shirt and a pair of trousers.*

The fact that the descriptive nature of multiword lexical units does not preclude alternative ad hoc descriptive expressions justifies the view that multiword lexical units is a matter of idiomaticity as well as vocabulary. (Cf. Allerton's suggestion that "idiomatics" should be introduced as a special branch of lexicology.)

#### *4. Summing up*

It has been suggested above that idiomaticity should be characterised as nativelike selection of expressions. This in turn implies that accounting for "all and only the possible structures in a language" is not an adequate aim in linguistic theory. Being overproductive, it misses the target.

It has also been suggested that features of idiomaticity can be found on different levels, ranging from discourse to phrase levels. Discoursal idiomatic features are thought to be the most elusive. Below this level, features of idiomaticity are divided into clausal and phrasal constructions. Clausal structures, in turn, are subcategorised into two functional classes: propositional and performative. Apart from being functionally different, there are some linguistic differences between these. Performatives tend to be less schematic, although sometimes they are abbreviated obscuring their clausal origin. The native as well as the non-native learner are often explicitly taught performatives since it is important to know what to say in common interactive situations such as leave-taking and greeting, apologising, thanking or congratulating someone.

Of particular importance are the combinatory constraints of single words. There are different types of such constraints. There are those involved in forming decontextualised and general meanings, i.e. so-called selectional restrictions and and—in the case of verbs—thematic roles. According to usage-based models of language acquisition, such generalised meanings are formed by abstracting semantic commonalities from different uses. A generalised meaning will enable the language-user to use the word creatively (=in unheard contexts) and yet be confident that it is used correctly.

It was, however, posited that some combinations will resist decontextualisation and be stored verbatim forming more or less strictly

## A Model of Idiomaticity

form-meaning pairs. These will normally be language-specific expressions which the non-native learner will have to learn in addition to generalised meanings. It was tentatively suggested that, since these phrases are normally mnemonically motivated combinations of words representing meanings occurring also in first language, explicit instructions may not be necessary for their acquisition provided there is exposure to the target language.

The lexical status of such multiword combinations is often not recognised. However, many linguists have in the last few decades recognised the large quantity of such expressions and concordances of large corpora confirm their numerousness.

Apart from selectional restrictions of the traditional kind and thematic roles, combinatory restrictions of words can also be in terms of so-called semantic prosodies. That is to say, a particular word typically combines with words of a particular type of—normally evaluative—meaning which is not warranted by generalised meanings. It is posited that for such constraints to be acquired exposure to the target language is particularly important.

*Lund University*

## References

- Alexander, R.J. (1978) "Fixed expressions in English: a linguistic, psycholinguistic, sociolinguistic and didactic study", *Anglistik und Englischunterricht* 6:171-188.
- Allerton, D.J. (1989) "Three (or four) levels of word co-occurrence restrictions", *Lingua* 63:17-40.
- Altenberg, B. and Granger, S. (2001) "The grammatical and lexical patterning of *make* in native and non-native student writing," *Applied Linguistics* 22, 2: 173-194.
- Bolinger, D. (1976) "Meaning and memory", *Forum Linguisticum* I: 1-14.
- Cowie, A.P. Mackin, R. and McCaig, I.R. (1975-1984). *Oxford Dictionary of Current Idiomatic English* Vols.I-II. *General Introduction*. Oxford: Oxford University Press.
- Cruse, A. (1991) *Lexical Semantics*. Cambridge: Cambridge University Press.
- Cruse, A. (2000) *Meaning in Language*. Oxford: OxfordUniversity Press.
- Erman, B. and Warren, B. (2000) "The idiom principle and the open choice principle", *Text* 20, 1:29-62.
- Fillmore, C. (1985) "Frames and the semantics of understanding", *Quarderni di Semantica* Vol. VI no.2: 222-254.
- Fillmore, C., Kay, P. and O'Connor M.C. (1988) "Regularity and idiomaticity in grammatical constructions: the case of *let alone*", *Language* 64:501-538.
- Goldberg, A. (1995) *A Construction Grammar Approach to Argument Structure*. Chicago: The University of Chicago Press.
- Hopper, P. (1998) "Emergent Grammar" in Tomasello, M. (ed.) *The New Psychology of Language*. Mahwah, New Jersey and London: Lawrence Erlbaum.
- Howarth, P. (1996) *Phraseology in English Academic Writing*. Tübingen: Max Niemeyer Verlag.
- Jackendoff, R. (1997) "Twistin' the night away", *Language* 73:3, 534-559.
- Kay, P. and Fillmore, C. (1999) "The *What's X doing Y?* construction" *Language* Vol 75.1: 1-33.
- Langacker, R. (1988) "A usage-based model", in Rudzka-Ostyn, B. (ed.) *Topics in cognitive linguistics*. Amsterdam: John Benjamins.
- Lavelle, T. and Minugh, M. (1998) "And high time, too: A corpus-based study of one English construction" in Lindquist et al (eds.) *The Major Varieties of English*, Acta Wexionensia No 1.

## A Model of Idiomaticity

- Lyons, J. (1966) "Firth's theory of 'meaning'" in *In Memory of J Firth*, pp 288-302. London: Longman.
- Makkai, A. (1972) *Idiom Structure in English*. The Hague and Paris: Mouton.
- Mauranen, A. (1996) "Discourse awareness and non-native speakers of English", *Zeitschrift für Fremdsprachenforschung* 7, 2:137-153.
- Mel'cuk, Igor. (1996) "Lexical functions: A tool for the description of lexical relations in a lexicon" in Leo Wanner, L. (ed.) *Lexical functions in lexicography and natural language processing*. Amsterdam/Philadelphia: John Benjamins.
- Pawley, A. and Syder, F. (1983) "Two puzzles for linguistic theory: nativelike selection and nativelike fluency" in Richards, J. C. and Schmidt, R. W. (eds.) *Language and Communication* 7. 1: 191-226. London: Longman.
- Penttilä, E. (ms) "Do a(n) X as an idiomatic construction in English", paper given at the *Symposium for the Relationship between Syntax and Semantics*, Helsinki, 2-4 September 1999.
- Sinclair, J. (1991) *Corpus, Concordance and Collocation*. Oxford: Oxford University Press.
- Strömquist, S. (forthcoming) "A picture of linguistic diversity" in Eriksson, Y. *Language and Visualisation*. Lund University: Department of Cognitive Science.
- Stubbs, M. (1995) "Collocations and semantic profiles". *Functions of Language* 2, 1:23-55.
- Tomasello, M. (2000) "First steps toward a usage-based theory of language acquisition", *Cognitive Linguistics* 11(1/2): 61-82.
- Wiktorsson, M. (2003) *Learning idiomaticity..* Lund Studies in English 105. Stockholm: Almqvist & Wiksell.
- Wray, A. (2000) *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
- Zettersten, A. (2003) "On collocations in bilingual lexicography". In Aijmer, Karin and Britta Olinde, *Proceedings from the 8th Nordic Conference on English Studies*. Gothenburg Studies in English 84: 151-161.