Thierry Fontenelle
*University of Liège*

# Using Lexical Functions to Discover Metaphors

### Abstract

This paper deals with the construction of a collocational database and with the potential applications of such a database in general and applied linguistics. The starting point of the database is the machine–readable version of the Collins–Robert dictionary (Atkins & Duval 1978/1987): this dictionary has been chosen because of the wealth of information it contains on various types of co–occurrence properties (collocations, selection restrictions, etc).

The first part of the paper describes the various types of information contained in our database; the second part focuses on one potential application and shows how new metaphors can be discovered by analyzing the content of the database.

## 1. Collocations and co–occurrence knowledge

Collocations have been the topic of a lot of research for the past 20 years, mainly because they pose a number of problems for the language learner who has to learn how to produce (encode) well–formed sentences in a foreign language (see Cowie 1986). Knowledge about permitted or forbidden word combinations will enable the learner to produce the following well–formed sentences or phrases and to avoid generating the odd–sounding expressions in brackets:

|  |  |
|---|---|
| a school of fish | (vs *a swarm/pack of fish) |
| a pack of dogs | (vs *a school/gaggle of dogs) |
| a speck of dust | (vs *a blade/drop of dust) |
| a drop of water | (vs *an item/a speck of water) |
| a confirmed bachelor | (vs *a big/high/hardened bachelor) |
| a heavy smoker | (vs *a big smoker) |
| to make a mistake | (vs *to do a mistake) |
| to pay attention | (vs *to make/do attention) |

As can be seen above, the term collocation is used here to refer to various types of idiosyncratic (i.e. unpredictable) combinations of nouns and adjectives, verbs and direct objects, nouns and nouns, etc. They are to be distinguished from totally frozen expressions such as idioms because they can usually be submitted to various types of syntactic manipulations (passivisation, insertion of material, pluralization...). However, the examples above show that they are not free insofar as one element (called the BASE, to use Hausmann's terminology) is responsible for the selection

of the other element (the COLLOCATOR). The noun *mistake* selects and, to some extent, as Allerton (1982) argues, 'tailors' the meaning of the accompanying verb *make*, which is used figuratively. In an encoding perspective, the information about which verb can co–occur with a given noun should sometimes be found in the entry for the noun (the base). Specialized dictionaries have been designed to that end, e.g. the BBI Combinatory Dictionary of English (Benson *et al.* 1986) or Ilgenfritz *et al.* (1989). These dictionaries, however, suffer from severe limitations insofar as they simply provide lists of collocates, without any attempt at performing any kind of semantic interpretation (although one has to admit that Benson *et al.* do sketch a methodology to provide a semantic interpretation in their introduction to the dictionary — see their discussion of EN (eradication/nullification) and CA (creation/activation) verbs — but, unfortunately, these labels are not used in the dictionary proper).

A much more refined approach to the semantic modelling of collocations can be found in Mel'chuk's Explanatory and Combinatory Dictionary (Mel'chuk 1984, 1988). The ECD, which is the lexical component of the Meaning–Text Theory, is a production–oriented dictionary insofar as it describes, among many other things, the lexical collocations in which an entry word participates; to that end, it resorts to a set of 50–odd lexical functions which are used together with a keyword to "signify a set of either phraseological combinations related to the keyword or those words which can replace the keyword under certain conditions" (Steele 1990:41). The general form of such functions is $f(X)=Y$, where X is the keyword (the entry in the ECD, the base of the collocation) and Y is the collocate that has to be selected to express the meaning denoted by $f(X)$. For example, Magn, which means 'very', 'intensely' or 'to a high degree' will be used in the following collocations:

Magn (bachelor) = confirmed        Magn (thirst) = unquenchable
Magn (célibataire) = endurci       Magn (thank) = warmly

The function $Oper_{i/j}$ applied to a noun yields the semantically empty verb which takes the keyword as direct object and the ith/jth actant as subject, as in:

$Oper_1$ (attention) = pay          $Oper_1$ (pressure) = exert
$Oper_2$ (attention) = draw         $Oper_1$ (suicide) = commit

Noun–noun collocations also form an important part of the ECD: Mel'chuk distinguishes several important lexical functions that capture idiosyncratic relationships linking a noun and the expressions of a single unit or instance (Sing) or of a group or aggregate (Mult):

Sing (dust) = speck, fleck, particle    Mult (fish) = school, shoal
Sing (grass) = blade                     Mult (bee) = swarm

Unfortunately, no such dictionary exists for English. Three volumes have been published for French, which represents a coverage of around 250 lexemes. Moreover, I do not know of any computerized collocational dictionary. The project currently under way at the University of Liège just aims to fill this lexicographical gap by constructing a collocational database that offers a wide range of access keys. The following section briefly describes this project.

## 2. Constructing a collocational database

It is generally admitted that constructing a lexical database from scratch is a waste of time. In keeping with current research on the reusability of lexical resources (see i.a. Heid/Martin/Posch 1991), the Liège database draws on the machine–readable version of the Collins–Robert English–French dictionary (Atkins & Duval 1978/1987), which was made available to us under research contract with the publishers. The availability of this computerized version makes it possible to manipulate and re–format the lexical information it contains. Collocations are explicitly specified in the italics part of the dictionary, following a well–defined system of representation for typical objects, typical subjects, typical nouns modified by adjectives, etc. I have argued elsewhere (Fontenelle 1992, 1994, in press) that lists of lexical collocations for any word in italics can be readily extracted from this bilingual dictionary (there are approximately 100,000 such items in the dictionary). Using the printed version only would require the user to leaf through the entire volume to have access to all the occurrences of these words in italics. The following examples from the printed dictionary illustrate how collocations and typical arguments are coded, using a system of parentheses, brackets and italics:

**school**[2] *n [fish]* banc *m*
**speck 1** *n [dust, soot]* grain *m*
**swarm**[1] **1** *n [bees]* essaim *m*
**deadly** *adj hatred* mortel, implacable
**mortal** *adj hatred* mortel
**bark** *vi [dog]* aboyer

Square brackets surround the typical subject of a verb or the typical noun complement of the headword. Typical nouns which can be modified by a given adjective are unbracketed.

The main problem is that the base of the collocation is not accessible to a human user because of the constraints imposed by the alphabetical order. The computerized version of the dictionary enables the user to retrieve the

set of items that are associated with any word in italics. The Liège database contains just this type of information, together with the lexical function that represents the semantic interpretation of the pairs of collocations in Mel'chuk's formalism. Figure 1 illustrates a few sample records from the database (simplified for the sake of clarity since the French translations given by the dictionary are provided in a separate field; moreover the base is disambiguated and translated too and its translation is provided separately):

| LF | Base | Collocator |
|------|--------|------------|
| Mult | fish | school |
| Sing | dust | speck |
| Sing | soot | speck |
| Mult | bee | swarm |
| Magn | hatred | deadly |
| Magn | hatred | mortal |
| Son | dog | bark |

Figure 1: sample entries from the collocational database

The lexical function (LF) has to be assigned manually, which poses a certain number of problems (in this respect, Heylen *et al.*'s assessment of Mel'chuk's lexical functions for NLP is particularly illuminating). It should be realized that such a database can be used in various ways, language teaching probably being the most straightforward application (indeed, the system makes it possible to construct queries such as "list the nouns that refer to a *group* of fish", "which verb expresses the typical sound made by a dog?", "which adjectives can co-occur with *hatred* to express a high degree of the concept?"). Unlike a traditional dictionary in which the alphabetical order is the only access key, this database enables users (translators, language students, linguists) to access information via any field, i.e. the base, the collocator, the translation of the lexical function, whether in isolation or in combination. As such, it resembles the kind of collocational database

described by Heid (1992) but unlike this tool, which is primarily designed for technical translators, the Liège database mostly contains general language, since the source from which it is extracted is a general–purpose dictionary. The following section illustrates a potential application, viz. the discovery of metaphors.

## 3. Collocations and metaphors

Lakoff & Johnson argue that "the essence of metaphor is understanding and experiencing one kind of thing in terms of another" (1980:5). Their contention is that the metaphorical structuring of concepts is reflected in the phrasal lexicon of the language (ibid:52). To illustrate their hypothesis, they show that metaphors such as ARGUMENT IS WAR are reflected in everyday language, as testified by the following expressions: *He attacked every weak point in my argument*; *He shot down all of my arguments*, etc. (ibid:4). To support their claim, Lakoff & Johnson provide the reader with an impressive list of metaphors ranging from LOVE IS A JOURNEY (*This relationship is a dead–end street*; *We've gotten off the track*) to TIME IS MONEY (*How do you spend your time these days?*; *That flat tire cost me an hour*) or AN ARGUMENT IS A BUILDING (*We need to construct a strong argument for that*; *We need more facts or the argument will fall apart*). Browsing through our collocational database and focusing on the realization of the Mult lexical function reveals that other metaphors are also commonly used. Consider the following examples retrieved from the database (queried against the occurrences of Mult to express a group/set of something):

    Mult (arrow) = cloud, rain, shower, storm
    Mult (bullet) = rain
    Mult (missile) = storm
    Mult (stone) = shower

Arrow, bullet, missile, and stone all refer to some kind of projectile used as a weapon to hit a target. The context in which these items are used usually denotes a war–like situation. Yet these terms are used in co–occurrence with words pertaining to an altogether different field of experience, viz. the field of meteorological phenomena (cloud, rain, shower...). This observation enables us to posit the existence of the metaphor A PROJECTILE IS A METEOROLOGICAL PHENOMENON. Paraphrasing Lakoff (1993:28), we may say that a conceptual mapping applies to a source domain ontology (the field of projectiles) and maps it onto a target domain ontology (the domain of meteorological phenomena). Yet, the concept of projectile should be construed much more broadly, as testified by the following examples retrieved from the database:

Mult (curse) = stream
Mult (insult) = shower, storm
Mult (abuse) = spate, storm

The bases of the above collocations all share a strong negative connotation: they involve two participants, an agent (the 'attacker') and a patient (the 'victim', who is physically or morally affected by the 'projectiles' loosed off by the attacker). The whole scene — or frame, to use Fillmore's terminology of frame semantics (Fillmore 1982) — is viewed as a situation where natural elements rage and bluster furiously. It should be noted that meteorological phenomena seem to be seldom associated with positive elements: it would be odd to say *a torrent/cloud/rain of cheers*. However, *a storm/thunder of applause* and *a storm of cheers* are attested in the dictionary, which forces us not to be too restrictive; cheers and applause are also viewed as some sort of 'projectile' launched by an agent against a patient and, as such, they acquire properties typical of meteorological phenomena.

It should be noted that this set of metaphors also exists in French, although one should realize that the two systems do not function in exactly parallel ways. As noted by Mantha & Mel'chuk (1988:47), the French noun *grêle* (hail) is associated with negative, harmful elements only (compare *une grêle de balles/d'injures* vs *une grêle de compliments/de fleurs*), very much like *shower* or *cloud* in the examples above. *Pluie,* however, is neutral in this respect since *une pluie de compliments* (positive) is attested next to *une pluie d'injures* (negative). Moreover, the projectile represented by the argument of the function Mult need not have a target in all cases. In Mult (spark) = cascade/shower, the sparks are not directed at anybody in particular while Mult (bullet) = rain or Mult (insult) = shower/storm definitely involve an actant which can be considered as a target (the patient or victim). In the French ECD (Vol. II), Mel'chuk explicitly specifies the optional or compulsory character of the target in the definition of the meteorological phenomenon itself *(pluie, grêle...)*.

What is important here is that we have a double system of metaphors. Curses, insults, abuse, cheers and applause are words and shouts. The metaphor WORDS ARE PROJECTILES makes it possible to use expressions from the field of projectiles to talk about speech and words. The general category of projectiles (which also contains true, non–metaphorical projectiles such as bullets, arrows and similar weapons) is then structured in terms of the metaphor A PROJECTILE IS A METEOROLOGICAL PHENOMENON.

Strong evidence for the existence of the metaphor WORDS ARE PROJECTILES is provided in the following entries for *barrage* in the Longman Dictionary of Contemporary English (LDOCE, Procter 1978) and the American Heritage Dictionary (AHD):

**barrage**

**a. LDOCE**

1. the firing of a number of heavy guns at once so that the exploding shells fall well forward in a line behind which soldiers are protected as they advance upon the enemy
2. (of speech or writing) a large number of things put forward at almost the same time or very quickly one after the other: *a barrage of questions*

**b. AHD**

1. a heavy curtain of missiles or artillery fire
2. a rapid outpouring: *a barrage of questions*

The second definitions both refer to a metaphorical use of the word *barrage* and the lexicographers were aware that this conventionalized metaphor ought to be captured and represented in the dictionary.

It is then crucial to realize that metaphors can be used to account for some co–occurrence phenomena which would otherwise be considered as purely idiosyncratic. This is very much in keeping with Nunberg *et al.* (forthcoming) who suggest that the meaning of many seemingly opaque (i.e. non–compositional) idioms can be explained in terms of situational metaphors. The collocations described above result from a mechanism of 'composition' (see Lakoff 1993:29) whereby two metaphorical mappings overlap and combine within a single phrase.

**4. Conclusions**

In this paper, I have outlined the general structure of a collocational database which is currently being constructed at the University of Liège. This database contains bilingual (English–French) collocations extracted from the machine–readable version of the Collins–Robert dictionary. The concept of lexical function, adapted from Mel'chuk's Explanatory Combinatory Dictionary, is used to model the semantics of these collocations. The database provides users with various access points (the base, the collocator, the lexical function...) and can therefore be used as a large–scale repository of collocational information to investigate the structure of the lexicon. One application described in this paper is the study of metaphors through the close scrutiny of the realization of the Mult lexical function.

## 5. Acknowledgements

**References**

Allerton, D.J. 1982. *Valency and the English Verb*. London: Academic Press.
Atkins, B.T., Duval, A. 1978/1987. *Collins–Robert French–English English–French Dictionary*. London & Glasgow: Collins; Paris: Dictionnaires Le Robert.
Benson, M., Benson, E., Ilson, R. 1986. *The BBI Combinatory Dictionary of English*. Amsterdam/Philadelphia: John Benjamins Publishing.
Cowie, A.P. 1986. "Collocational Dictionaries – A comparative view" in Murphy (ed.), *Fourth Joint Anglo–Soviet Seminar*, London: British Council, pp.61–69.
Fillmore, C. 1982. "Frame Semantics" in The Linguistic Society of Korea (ed.), *Linguistics in the Morning Calm*, Seoul: Hanshin Publishing Co., pp.111–137.
Fontenelle, T. 1992. "Collocation acquisition from a Corpus or from a Dictionary: a Comparison" in Tommola *et al.* (eds), *EURALEX'92: Proceedings of the Fifth Euralex International Congress*. University of Tampere, Finland, pp.221–228.
Fontenelle, T. 1994. "Towards the construction of a collocational database for translation students". *META*, 39:1, pp.47–56.
Fontenelle, T. (in press) "Using a bilingual computerized dictionary to retrieve support verbs and combinatorial information". *Acta Linguistica Hungarica*, 41:1–2, 1990–1991.
Hausmann, F.J. 1985. "Kollokationen im Deutschen Wörterbuch. Ein Beitrag zur Theorie des lexikographischen Beispiels", in Bergenholtz & Mugdan (eds) *Lexikographie und Grammatik*. Lexicographica Series Major 3, pp.118–129.
Heid, U. 1992. "Décrire les collocations: deux approches lexicographiques et leur application dans un outil informatisé". *Terminologie et Traduction*, n 2–3, C.E.C., Luxembourg, pp.523–548.
Heid, U., Martin, W., Posch, I. 1991. *Feasibility of standards for collocational description of lexical items*. EUROTRA–7 study, C.E.C., Luxembourg.
Heylen, D. 1993. *Collocations – Final Report of the ET–10/75 Study on the Lexicalisation of Semantic Operations*. Utrecht University, Stichting Taaltechnologie.
Ilgenfritz, P., Stephan–Gabinel, N., Schneider, G. 1989. *Langenscheidts Kontextwörterbuch Französich–Deutsch*. Berlin–München: Langenscheidt.
Lakoff, G., Johnson, M. 1980. *Metaphors We Live By*. Chicago: The University of Chicago Press.
Lakoff, G. 1993. "The syntax of metaphorical semantic roles", in Pustejovsky, J. (ed.), *Semantics and the Lexicon*. Dordrecht: Kluwer Academic Publishers, pp.27–36.
Mantha, S., Mel'chuk, I. 1988. "Le Champ Lexical "Phénomènes Atmosphériques" – Formulations des Définitions", in Mel'$chuk, I. *et al. Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico–sémantiques* – Vol.II. Montréal: Les Presses de l'Université de Montréal, pp.41–47.
Martin, W. 1992. "Remarks on Collocations in Sublanguage". *Terminologie et Traduction*, n 2–3, C.E.C., Luxembourg, pp.157–164.
Mel'chuk, I. *et al.* 1984/1988/1992. *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico–sémantiques* – Vol.I (1984), Vol.II (1988), Vol.III (1992). Montréal: Les Presses de l'Université de Montréal.
Nunberg, G., Sag, I., Wasow, Th. (forthcoming) "Idioms", to appear in *Language*.
Steele, J. (ed.) 1990. *Meaning–Text Theory: Linguistics, Lexicography and Implications*. University of Ottawa Press.