Meta
Journal des traducteurs
Translators' Journal

Towards the Construction of a Collocational Database for Translation Students

Thierry Fontenelle

Aller au sommaire du numéro

Découvrir la revue

Citer cet article

Fontenelle, T. (1994). Towards the Construction of a Collocational Database for Translation Students. *Meta*, *39*(1), 47–56. https://doi.org/10.7202/002756ar

Résumé de l'article

Cet article présente une expérience effectuée au cours d'un projet d'élaboration d'une base de données bilingue (anglais-français) de collocations à l'Université de Liège. Cette base de données offrira aux utilisateurs plusieurs points d'accès ainsi que des informations lexico-sémantiques. On tentera de démontrer que ces particularités sont présentement absentes dans le domaine des dictionnaires de collocations et que cette base de données réussira à combler ce manque.

Cet article est diffusé et préservé par Érudit.

Érudit est un consortium interuniversitaire sans but lucratif composé de l'Université de Montréal, l'Université Laval et l'Université du Québec à Montréal. Il a pour mission la promotion et la valorisation de la recherche.

https://www.erudit.org/fr/

TOWARDS THE CONSTRUCTION OF A COLLOCATIONAL DATABASE FOR TRANSLATION STUDENTS

THIERRY FONTENELLE
University of Liège, Liège, Belgium

Résumé

    Cet article présente une expérience effectuée au cours d'un projet d'élaboration d'une base de données bilingue (anglais-français) de collocations à l'Université de Liège. Cette base de données offrira aux utilisateurs plusieurs points d'accès ainsi que des informations lexico-sémantiques. On tentera de démontrer que ces particularités sont présentement absentes dans le domaine des dictionnaires de collocations et que cette base de données réussira à combler ce manque.

INTRODUCTION

For over a decade, there has been a strong revival of interest in lexical studies, and numerous institutions, universities and private companies have embarked on large-scale, long-term linguistic studies, focussing primarily on the lexical aspects of the linguistic description of natural languages. After Amsler's seminal work on the machine-readable version of the Merriam-Webster Pocket Dictionary (Amsler 1980), immediately followed by Michiels's dissertation on the computerized version of the *Longman Dictionary of Contemporary English* (Michiels 1982), researchers started to realize that there was more in dictionaries than meets the eye. Various projects were launched to investigate the possibility of re-formatting, re-using and extracting the information contained in the computerized versions of lexical resources (dictionaries, term banks, thesauri, etc.). Most of these research projects have aimed (and still aim) at exploiting these resources in a natural language processing (NLP) perspective. The boom of the information market and the increasing need for translation and other linguistic services have indeed made it necessary to develop various types of NLP products, ranging from machine translation systems to natural-language front-ends to information-retrieval or speech-processing systems. The major stumbling block in the development of such systems, however, is the construction of the large lexicons these systems generally have to draw upon. It is a well-known fact that building these lexicons from scratch is both time-consuming and costly and the idea soon emerged that linguistic information about words could be derived from already existing computerized lexical resources.

    Applications in computational lexicography, however, tend to be limited to natural language processing and often discard other (less ambitious?) goals such as language teaching and language learning, perhaps because the computer scientists and computational linguists who research machine-readable dictionaries often tend to be more computer-oriented than teaching-minded. Yet, it should be borne in mind that teaching the behaviour of words to language students and formalizing the syntactic, semantic or pragmatic properties of lexical items for computers often boil down to describing the same linguistic facts, albeit in an altogether different terminology. The special attention paid by computational lexicographers to the computerized versions of *learners'* dictionaries (*cf.*

Boguraev and Briscoe 1989, who devote a whole book to the *Longman Dictionary of Contemporary English*) testifies to this absence of clear-cut distinction between computational and pedagogical applications.

In this paper, I would like to report on an experiment carried out at the University of Liège in the field of applied linguistics. As will be shown below, this experiment is part of a project under way at Liège aimed at constructing a database of bilingual (English-French) collocations. It is to be used, among many other applications, by French-speaking university students of English in their quest for the 'right term'. Although students are rightly taught to turn to a dictionary such as the *BBI Combinatory Dictionary of English* (Benson *et al.* 1986), it will be shown that this database can fill a collocational gap in lexicography because it will provide users with various access points and because it will also include lexical-semantic information. The starting point for this database is the machine-readable version of the Collins-Robert (English-French, French-English) dictionary (Atkins and Duval 1987), in keeping with the prevailing idea that constructing lexical knowledge repositories from scratch should be avoided.

MULTI-WORD LEXICAL UNITS

In a recent paper, Cowie argues that "journalistic prose draws very heavily on verb-noun collocations that are already well-established and widely known" (Cowie 1992: 1). This does not seem to be typical of journalistic prose only, however, and research in applied linguistics has shown that native speakers often memorize ready-made word combinations. They usually have some predisposition to store these combinations as wholes, which accounts for the pervasiveness of lexical collocations in everyday language. The term 'lexical collocation' is used here to refer to the privileged, idiosyncratic relationship that holds between some verbs and their subjects/objects or between some nouns and the adjectives that modify them. It should be noted that this term may also denote the special relationship between two nouns or between a verb and an adverb. Some authors (*e.g.* Cowie 1986, 1991; Howarth 1993) also distinguish between free collocations and restricted collocations. The former refer to combinations that "allow substitution of either, or any, of their elements without semantic change in the other element or elements" (Cowie 1986: 62). The combination *fire staff*, for instance, is considered as free because other verbs can be substituted for *fire* (*dismiss, lay off, sack*) and *staff* can be replaced by *worker, employee, clerk*, etc. (it is more a co-occurrence restriction on the type of nouns that can be inserted in the direct object slot than a collocation). In contrast, restricted collocations are semi-fixed combinations and are characterized by the limited collocational range of their elements. In the case of verb + direct object restricted collocations, Howarth (1993: 4) distinguishes between delexical collocations (usually with a semantically empty verb such as *have, make, do, get...*), collocations with a figurative verb (*adopt a policy, hazard a guess*) and technical or specialized collocations (*snuff out a candle*). Typically, the verb often loses its primary meaning to acquire a figurative meaning under the influence of the noun it accompanies (or 'supports', to use Gross's terminology; *cf.* Gross 1981). The French noun *choix*, for example, is responsible for the selection of the verbs *faire* or *opérer* and it can be safely asserted that *opérer* in *opérer un choix* has not the meaning of "pratiquer une intervention chirurgicale sur" illustrated by *opérer un patient*. Similarly, the delexical verb *pay* in *pay attention* has acquired a figurative or nearly empty meaning in the vicinity of *attention*. It should be stressed here that knowledge of such restrictions is of paramount importance since cross-linguistic divergences are the rule (*payer attention* is unacceptable in French, the delexical verb *faire* having to be used instead). Learners should therefore be made aware of these restrictions and taught how to identify, memorize and use these semi-frozen expressions that account for a native speaker's phraseological

competence. This is what I wish to show in the following sections, focussing more espe-
cially on the teaching of certain collocations, highly frequent in business English, to
French-speaking learners of English.

COLLOCATIONS IN BUSINESS ENGLISH

The general idea that lies at the root of the experiment described below developed
from the observation that many advanced university students of English had difficulty in
selecting and using the frequent collocations used pervasively in macro-economics texts.
In translation classes, French-speaking students had been asked to translate a business text
containing many references to rises and drops in prices. It appeared that many students,
for the sake of simplicity or out of sheer ignorance, tended to use only verbs such as *in-
crease* or *decrease* in combination with a noun like *price*. Many other verbs, however,
collocate with that noun and the following authentic examples from the Oxford Uni-
versity Press Pilot Corpus testify to the richness and diversity of the English lexicon in
that particular area:

- Oil prices have *soared* above forty dollars a barrel.
- Tokyo stock and bond prices and the value of the yen all *plunged* yesterday.
- If free competition were introduced, prices would *plummet*.
- IBM's share price *rocketed* in London and New York.
- House prices will *slump* by 10% next year.
- They were reluctant to raise their investment plans, even though prices were *booming*.

The preceding examples illustrate the use of *price* as the subject of an intransitive
verb. The following sentences show that it can also be used as the direct object of several
transitive verbs:

- They are *slashing* prices to secure sales.
- The price of unleaded petrol should be *lowered*.
- The Economy Minister *boosted* petrol prices by up to 700 per cent.
- The increase in young house buyers will *push* house prices *up* by an additional one per
  cent next year.
- The ministry of food reported that it must *raise* food prices.

It should be noted that French also resorts to specific, often metaphorical verbs to
refer to the behaviour of prices (*les prix s'effondrent, dégringolent, augmentent, diminuent,
montent en flèche, flambent...*). Such language is often typical of journalese but exerts a
strong influence on everyday language as well. It is therefore important that students
should be aware of such collocations and able to use them adequately when translating a
text into a foreign language, since they are going to be judged by their ability to manipu-
late these ready-made chunks of language. As Howarth (1993: 1) notes, this forces us to
move away from the compositional, generative view of language processes to adopt a
more lexicon-centered (phraseological) approach to translation and foreign-language
teaching. The following section describes how these typical chunks can be extracted from
the Collins-Robert dictionary.

COLLOCATIONS IN THE COLLINS-ROBERT DICTIONARY

Compilations of lexical collocations for English are quite rare, with the notable
exception of the *BBI Combinatory Dictionary of English* (Benson *et al.* 1986). Compu-
terized collections of collocations for English are even rarer and I do not know of any
machine-readable dictionary of bilingual collocations. The Collins-Robert dictionary is
not primarily intended to be used as a collocational dictionary but its availability in a

machine-readable version at the University of Liège under contract with the publishers was too tempting to leave untapped the wealth of information it contains on the semantic, syntactic and combinatory properties of words. This dictionary, which is probably the best bilingual English-French, French-English dictionary currently on the market, partly owes its reputation to the extensive use it makes of metalinguistic information. A study of the microstructure of the dictionary reveals that it relies heavily on the use of parentheses, brackets and italics in the coding of collocations and typical arguments (typical objects, subjects of verbs, etc). The following examples show that the collocational range of English transitive verbs is represented as a list of lexical items appearing in italics immediately before the French translation(s). For the sake of clarity, I use only examples illustrating collocations with *price* since we will be mainly concerned with this lexical item.

> **boost** *vt* (**b**) (*Comm, Ind etc: increase*) *price* hausser, faire monter; *output, productivity* accroître, développer; *sales, product* promouvoir, faire monter en flèche; *confidence etc* renforcer.
> **lower** *vt* (**b**) (*fig*) *pressure, heating, price, voice* baisser.
> **slash** *vt* (**b**) *prices* casser*, écraser*; *costs, expenses* réduire radicalement; *speech, text* couper *or* raccourcir radicalement.

The relationship between a verb and its typical subjects is conveyed via square brackets surrounding the subject, as in the following examples:

> **skyrocket** *vi [prices]* monter en flèche
> **plummet** 2 *vi [aircraft, bird]* plonger, descendre *or* tomber à pic; *[temperature]* baisser *or* descendre brusquement; *[price, sales]* dégringoler; *[spirits, morale]* tomber à zéro
> **boom** 1 *vi* (**b**) (*Comm, Fin, St Ex*) *[prices]* monter en flèche
> **slump** 2 *vi* (**a**) *[popularity, morale, production, trade]* baisser brutalement; *[prices, rates]* s'effondrer

Typical adjectives are also represented in italics. The following examples show that adjectives other than *high* can be used to refer to 'above average' prices:

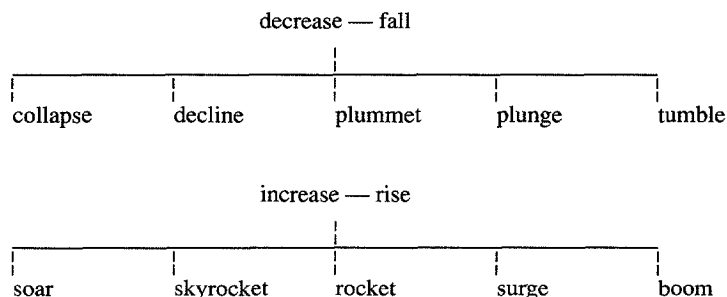> **extortionate** *adj price* exorbitant, inabordable; *demand, tax* excessif, exorbitant
> **steep** *adj* (**b**) (**fig*) *price* élevé, excessif; *bill* salé*; *story* raide*
> **inflated** *adj tyre etc* gonflé; (*Med*) *lung* dilaté; (*fig*) *style* enflé, boursouflé; *value* exagéré; *prices* exagéré, gonflé

A list of lexical collocations for the word *price* (or for any other word in italics) can be extracted automatically from the machine-readable version of the bilingual dictionary, a task which would be virtually impossible for a human user who would be obliged to leaf through the entire printed volume (833 pages for the English-French part) to have access to all the occurrences of *price/prices* in italics. I have argued elsewhere (Fontenelle 1992a) that such an extraction method can usefully complement corpus-based statistical analyses since the dictionary is used here as a corpus of pre-digested information. Combining a query on *price* in italics with a query on part-of-speech information (vi = intransitive verb, vt = transitive verb...) reveals that one can *advance, agree, boost, bring down, drop, increase, lower, mark down, push up, put up, raise* or *slash* prices (this is just a small sample of what one can do to prices; the complete lists as retrieved from the dictionary can be found in Fontenelle 1992a). *Price* can also co-occur in subject position with numerous intransitive verbs, such as *advance, boom, collapse, decline, decrease, drop, fall, go through the roof, increase, plummet, plunge, rise, shoot up, skyrocket, soar* or *toboggan*, to quote but a few.

Several remarks ought to be made at this juncture. The first thing to note is that the lists given above can be divided into two further classes as a function of the upward or

downward movement denoted by the verbs, some referring to a rise, others to a drop in prices. Another point worth noting is that the lists also contain two types of items, namely verbs that could be said to belong to some kind of 'core' vocabulary on the one hand (*cf.* Carter 1987) and more 'exotic' variants on the other hand. *Increase, decrease, rise* and *fall*, for instance, undoubtedly belong to the former category while *skyrocket, plummet, tumble, shoot up* or *go through the roof* are clearly less common and much more peripheral. It may be safely argued that the latter category contains near-synonyms that stand in a relation of hyponymy to the more general (core) verbs in a taxonomical hierarchy such as the following:

decrease — fall

collapse	decline	plummet	plunge	tumble

increase — rise

soar	skyrocket	rocket	surge	boom

The existence of such a taxonomy with multiple hyperonyms is confirmed by the treatment of these verbs in the Collins Cobuild English Language Dictionary (Sinclair 1987). This learner's dictionary is, to my knowledge, the only one that overtly captures the three basic lexical relationships (synonymy, antonymy, hyponymy). This information is to be found in the extra-column which also contains grammatical codes. The entries for *rocket* and *soar* illustrate how the taxonomy can be reconstructed:

rocket 2	If things such as prices and profits **rocket**, they get bigger very quickly and suddenly; an informal use. EG *His profits rocketed ... Land sales rocketed*	V ♠ rise = soar
soar 1	If something connected with money or business **soars**, it suddenly increases or becomes more valuable. EG *They watched the value of their shares soar from from less than £50 to £87... The development costs of the RB211 engine had soared... Rice production soared from 694,000 tons to 913,000 tons.*	V:IF+PREP THEN *from / to* ♠ increase = rocket

The up-arrow (♠) indicates hyperonyms and the two entries clearly show that *rocket* and *soar* are co-hyponyms that share the same superordinate terms (*rise* and *increase*).

To return for a moment to the pedagogic aspect of the problem, let me mention here that before being submitted to the collocations retrieved from the dictionary, learners tended to avoid the more exotic items and used exclusively more neutral, core items. When asked to translate a few French sentences into English without any dictionary or reference work (see below), even advanced learners stuck to the general hyperonyms, as if they were intuitively aware of Carter's observation that "the more core a lexical item is, the more partnership it will contract with other lexical items" (Carter 1987: 37). This behaviour would probably be referred to as a type of 'avoidance strategy' by some authors (Kleinmann 1977), but as Blum and Levenston (1978: 401) note, true avoidance presupposes choice. A more adequate term, then, might be 'apparent avoidance' since lack of vocabulary generally accounts for students' resorting to these core items.

Translators who translate "les prix s'effondrent/dégringolent" and "les prix flambent/grimpent en flèche/s'envolent" by "prices decrease/fall" or "prices increase/rise" respectively use only a superordinate term, and the absence of a qualifier, *i.e.* an additional explanation of a word in translation, results in what Blum and Levenston (1978: 404) call 'the depletion of meaning'. Some students, who were aware of their lack of vocabulary but sensed that using a hyperonym would not be sufficient, added such qualifiers in an attempt to overcome this depletion of meaning (and came up with translations such as "prices increase *quickly/rapidly*" or "prices rise *steeply/sharply/dramatically*").

TEACHING COLLOCATIONS

Teaching grammatical rules to language students is certainly not a matter of forcing them to study these rules by heart. On the contrary, they have to be able to abstract away from familiar situations, to extrapolate in order to generate novel sentences to which they have not yet been exposed. Teaching collocations, however, is a very different type of activity, since the learner has to be able to store and manipulate prefabricated chunks of language. To quote Cowie, it seems to be "impossible to perform at a level acceptable to native users (...) without controlling an appropriate range of multi-word units. Moreover, the demand of creative expression in the foreign language rests, as it does for native speakers and writers, on prior knowledge of a repertoire of such expressions" (1992: 10). Rote learning is therefore bound to play an important part in this type of lexical acquisition, given the idiosyncratic, unpredictable character of collocations.

An experiment was recently carried out with 76 third- and fourth-year university EFL students. They were first asked to translate a few French sentences referring to rises or drops in prices without using any dictionary (as described in section 4 of this paper). As already said, they scored poorly in their command of the collocations of *price*. For example, more than 50% of the students resorted to the two hyperonyms *rise* and *increase* to translate the sentence "les prix grimpent en flèche". This, however, was due to a lack of vocabulary since they had never been systematically exposed to the appropriate type of sub-language. In an attempt to make them aware of the richness of the English lexicon, they were then provided with lists of potential collocations extracted from the Collins-Robert dictionary. Authentic texts were also read to illustrate the use of these collocations in context. A few weeks later, when asked to take a similar test, only one student persisted in using *increase* while the majority (over 80%) would use non-core verbs such as *rocket* or *skyrocket*, which were hardly attested at all in the first part of the test.

True avoidance strategies were still detected since some students continued using a superordinate term (*increase/decrease* etc.) to translate a marked term in French (les prix s'envolent/dégringolent...). It can safely be argued that the students no longer lacked knowledge of the appropriate vocabulary, though some of them were not yet confident enough to use a term they knew only passively. They then resorted to this 'escape route' to avoid making mistakes or because they were intuitively aware, as I said earlier, that a core vocabulary item is likely to participate in more co-occurrence relations, although part of the meaning (intensification, connotation) is generally lost.

Interestingly enough, a few students committed errors in the syntactic potential of the verbs elicited. A strictly intransitive verb, for example, was used transitively. The French sentence "Les détaillants peuvent casser les prix" was translated by one student as "*Retailers can plummet prices*", although *plummet* can only be used intransitively; some students also believed that *raise* or *rise* are mutually interchangeable, that "*without rising the price of unleaded petrol*" is acceptable in English. This makes it abundantly clear that extreme caution should be exercised in presenting lists of collocations to students. Pairs

of items (*price-rise*; *price-plummet*; *raise-price...*) are obviously the basis of the learning process, but when presenting the pair of collocates the semantics of the link between the members of the pair should also be specified. Some system has therefore to be devised to indicate that verbs such as *plummet* or *rise* can be used to answer the question: *What can prices do?* Similarly, the learner has to be taught that *raise* can collocate with *price* but only to answer the question: *What can one do to prices?* Various theories are at the linguist's disposal to perform that kind of semantic interpretation. In the following section, I briefly sketch the framework chosen in the Liège project for a collocational database.

LEXICAL FUNCTIONS AND MEL'ČUK'S COMBINATORY DICTIONARIES

At the end of the sixties, Igor Mel'čuk and his colleagues (Apresyan *et al.* 1969) described the general foundations for a new type of dictionary that was meant to reflect the combinatorial properties of words. At the heart of Mel'čuk's Meaning-Text Theory, we find the so-called lexical functions which describe the various semantic and collocational relations that can hold between lexemes. To quote Steel and Meyer, "a lexical function (f) is 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). A very important function in this theory is the *Oper* lexical function, which links the keyword to the semantically empty verb that supports the keyword and takes it as direct object. The following examples illustrate this function whose usefulness need not be demonstrated in a language learning perspective:

$Oper_1$ (influence) = exert
$Oper_1$ (attention) = pay
$Oper_1$ (suicide) = commit

Another important function is the *Magn* operator which links the keyword to the items that express the highest degree of the concept denoted by the keyword, as in:

Magn (thirst) = unquenchable
Magn (liar) = arrant
Magn (bachelor) = confirmed
Magn (smoker) = heavy

Mel'čuk (1984) distinguishes more than 50 lexical functions, ranging from the traditional lexical relations (*Syn* = synonym of; *Anti* = antonym of; *Gener* = hyperonym of) to more 'exotic' but nonetheless interesting lexico-semantic relations. *Son*, for example, refers to the typical sound made by the keyword, as in *Son (frog)* = *croak* or *Son (bee)* = *drone*. The *Mult* and *Sing* functions refer to the aggregates or the units/portions of the keyword respectively, as in the following examples:

Mult (fish) = school, shoal
Mult (dog) = pack
Sing (dust) = fleck, speck
Sing (salt) = grain

In many cases, this information is present in traditional mono- or bilingual dictionaries but is to be found under the collocator and not under the base of the collocation (the keyword), to quote Hausmann (1985). In the Collins-Robert, for example, the entry for *fish* does not provide the user with information on how to refer to a 'group of fish'. Only the entries for *shoal* and *school* explicitly mention that some kind of link (to be

determined by the user) holds between these items and the word *fish*, though this information can only be used for decoding purposes, not for encoding text:

> **school**[2] *n [fish]* banc *m*
> **shoal** *n [fish]* banc *m*

It should be realized that the base of the collocation refers both to a real word and to some kind of metalanguage. As such, the word in italics can function as the head of a thesauric class (*a school of fish* is a well-formed phrase, next to *a school of sardines* or *of herring*). Lexicographers usually only use a superordinate term because exhaustive listing of all the hyponyms would be too space-consuming in most cases. *Son (dog) = bark* has been quoted to death but linguists are too prone to forget that *dog* can be replaced by *greyhound, poodle, Alsatian*, etc.

A translator needs to have tools at his or her disposal that will facilitate the text-encoding process. The collocational database currently being constructed intends to meet just these needs. The availability of the Collins-Robert in a machine-readable version makes it possible to extract potential pairs of collocates and Mel'čuk's lexical functions can be used to perform the semantic interpretation of these combinations. Of course, although the extraction process can be carried out automatically, the semantic interpretation proper has to be done by a lexicographer (and the first step is to disambiguate the words in italics, which are sometimes used in their various senses, as I point out in Fontenelle 1992b). The English-French part of the Collins-Robert dictionary contains more than 100,000 such collocations in italics, which means that, once the construction is completed, translators (and language students) will have access to an on-line database which can answer queries like the following:

> * Give me the transitive verbs that can take *price* as direct object and that refer to a drop in prices;
> * Give me the intransitive verbs that can take *price* as subject and that refer to a rise in prices;
> * Give me the adjectives that refer to 'high' prices.

On the basis of the information extracted from the dictionary, the database contains the following facts (simplified for the sake of clarity since the database contains bilingual information):

> CausPredMinus (price) = beat down, bring down, depress, drop, knock down, lower, mark down, reduce...
> IncepPredPlus (price) = advance, boom, go through the roof, hike, jump, increase, rise, rocket, skyrocket, soar...
> Magn (price) = exorbitant, extortionate, high, unbeaten...

As can be noticed, several lexical functions can be combined and the resulting clusters of functions may serve to express sometimes highly complex relationships. *Pred*, for instance, is a verbaliser of nouns and adjectives and means "be the keyword". It is very often used in combination with *Plus* (more) or *Minus* (less) and with the aspectual operators *Caus* (denoting the cause) or *Incep* (denoting the beginning of a process). It will be noted that this depth of analysis enables lexicographers to distinguish the verbs *raise* and *rise*, since the former verb appears as the value of the function *CausPredPlus* ($\approx$ to cause to be more) while the latter is the realization of the function *IncepPredPlus* ($\approx$ to begin to be more). Needless to say that the same methodology can also be applied to other fields, for example in the study of quasi-legal texts. Examining the collocations of the word *law*

typically reveals that several transitive verbs are at the translator's disposal to refer to the 'destruction', the 'nullification' of a law (one can *abrogate, abolish, annul, repeal, rescind, revoke* or *do away with* a law). In this respect, Mel'čuk has identified the *Liqu* lexical function (for liquidate), which denotes the destruction of something (*e.g. Liqu (file) = erase, delete...*). The database will then specify that *Liqu (law) = abrogate, abolish, repeal...* As pointed out by Cohen (1993: 511) who adopts a similar methodology for her *Lexique de cooccurrents*, the main drawback of such an approach is that the use of the database presupposes prior knowledge of the lexical functions; I am convinced, however, that such a project can serve the translators' and students' needs in their quest for the right term.

CONCLUSIONS

The main aim of this paper was to show that computerized dictionaries, which are the object of a great deal of research in the field of natural language processing, can also be used in applied linguistics to teach translation students how to select various types of collocations. The database which is currently being constructed at the University of Liège will be used to enable French-speaking students of English to expand their vocabulary and achieve greater fluency by exposing them to well-established, ready-made phraseological units that they would not be able to discover in the printed version of the Collins-Robert dictionary, given the constraints of alphabetical order. The collocational database will also serve as a repository of decision-making material to be used as a reference tool during the encoding process. It should fill a gap in lexicography since it will provide users with several access keys (the base, the collocator, the lexical function, the translation...) to make the collocational information more easily accessible than in the BBI dictionary.

The underlying theory is Mel'čuk's Meaning-Text Theory and, more specifically, the set of lexical functions which lie at the root of the explanatory combinatory dictionary. An experiment has been described to illustrate how collocations could be recognized, extracted and formalized for students translating business texts into English. These students are generally tempted to use core items or words that bear a strong resemblance to their French counterparts, and it is hoped that the kind of data contained in our collocational database will contribute to enlarging their vocabulary while developing in them a feel for idiomaticity.

REFERENCES

AMSLER, Robert (1980): *The Structure of the Merriam-Webster Pocket Dictionary*, PhD thesis, the University of Texas at Austin.

APRESYAN, Y., I. MELČUK and A. ŽOLKOVSKY (1969): "Semantics and Lexicography: A New Type of Unilingual Dictionary", Kiefer (Ed.) *Studies in Syntax and Semantics*, Reidel, Dordrecht-Holland, pp. 1-33.

ATKINS, Beryl T. and Alain DUVAL (1987): *Robert & Collins Dictionnaire Français-Anglais, Anglais-Français*, Paris, Le Robert, Glasgow, Collins.

BENSON, Morton (1985): "Collocations and Idioms", Ilson (Ed.) *Dictionaries, Lexicography and Language Learning*, ELT Documents:120, Oxford, Pergamon Press, pp. 61-68.

BENSON, Morton, BENSON, Evelyn and Robert ILSON (1986): *The BBI Combinatory Dictionary of English*, Amsterdam-Philadelphia, John Benjamins Publishing Company.

BLUM, S. and E. A. LEVENSTON (1978): "Universals of Lexical Simplification", *Language Learning*, 28-2, pp. 399-415.

BOGURAEV, Branimir and Ted BRISCOE (1989): *Computational Lexicography for Natural Language Processing*, London and New York, Longman.

CARTER, Ron (1987): *Vocabulary: Applied Linguistics Perspectives*, London, Unwin Hyman.

CARTER, Ron and Michael McCARTHY (1988): *Vocabulary and Language Teaching*, Harlow, Longman.

COHEN, Betty (1993): "Méthodes de repérage et de classement des cooccurrents lexicaux", *Terminologie et Traduction*, n° 2/3-1992, Luxembourg, Commission des Communautés européennes, pp. 505-511.

COWIE, Anthony P. (1986): "Collocational Dictionaries — A comparative view", Murphy (Ed.) *Fourth Joint Anglo-Soviet Seminar*, London, British Council, pp. 61-69.

COWIE, Anthony P. (1991): "Multiword Units in Newspaper Language", Granger (Ed.) *Perspectives on the English Lexicon: A Tribute to Jacques Van Roey*, Cahiers de l'Institut de Linguistique de Louvain, CILL 17.1-3, pp. 101-116.

COWIE, Anthony P. (1992): "Multiword Lexical Units and Communicative Language Teaching", Arnaud and Béjoint (Eds) *Vocabulary and Applied Linguistics*, London, Macmillan, pp. 1-12.

FONTENELLE, Thierry (1992a): "Collocation Acquisition from a Corpus or from a Dictionary: a Comparison", *EURALEX'92 Proceedings, Fifth EURALEX International Congress*, University of Tampere, pp. 220-228.

FONTENELLE, Thierry (1992b): "Co-occurrence Knowledge, Support Verbs and Machine-Readable Dictionaries", F. Kiefer, G. Kiss and J. Pajzs (Eds) *Papers in Computational Lexicography — COMPLEX'92*, Linguistics Institute, Budapest, Hungarian Academy of Sciences, pp. 137-145.

GROSS, Maurice (1981): "Les bases empiriques de la notion de prédicat sémantique", *Langages 63*, pp. 7-52.

HAUSMANN, F. (1985): "Kollokationen im Deutschen Wörterbuch. Ein Beitrag zur Theorie des Lexikographischen Beispiels", Bergenholtz and Mugdan (Eds) *Lexikographie und Grammatik*, Tübingen, Niemeyer, pp. 118-129.

HOWARTH, Peter (1993): "A Phraseological Approach to Academic Writing", MS, University of Leeds.

KLEINMANN, H. (1977): "Avoidance Behaviour in Adult Second Language Acquisition", *Language Learning*, 27-1, pp. 93-107.

MEL'ČUK, Igor (1984): *Dictionnaire explicatif et combinatoire du français contemporain: recherches lexico-sémantiques I, Montréal*, Les Presses de l'Université de Montréal.

MICHIELS, Archibald (1982): *Exploiting a Large Dictionary Database*, PhD thesis, University of Liège, mimeographed.

SINCLAIR, John (1987): *Collins Cobuild English Language Dictionary*, Glasgow, HarperCollins Publishers.

STEELE, James (Ed.) (1990): *Meaning-Text Theory: Linguistics, Lexicography, and Implications*, University of Ottawa Press.