A Bilingual Electronic Dictionary for Frame Semantics

Thierry Fontenelle

19 Rue du Merschgrund, L-8373 Hobscheid, Luxembourg fontenel@pt.lu

Abstract

Frame semantics is a linguistic theory which is currently gaining ground. The creation of lexical entries for a large number of words presupposes the development of complex lexical acquisition techniques in order to identify the vocabulary for describing the elements of a 'frame'. In this paper, we show how a lexical-semantic database compiled on the basis of a bilingual (English-French) dictionary can be used to identify some general frame elements which are relevant in a frame-semantic approach such as the one adopted in the FrameNet project (Fillmore & Atkins 1998, Gahl 1998). The database has been systematically enriched with explicit lexical-semantic relations holding between some elements of the microstructure of the dictionary entries. The manifold relationships have been labelled in terms of lexical functions, based on Mel'cuk's notion of co-occurrence and lexical-semantic relations in Meaning-Text Theory (Mel'cuk et al. 1984). We show how these lexical functions can be used and refined to extract potential realizations of frame elements such as typical instruments or typical locatives, which are believed to be recurrent elements in a large number of frames. We also show how the database organization of the computational lexicon makes it possible to readily access implicit and translationally-relevant combinatorial information.

Introduction

It is no longer useful to dwell on the costly and lengthy nature of the construction of a computational lexicon for language processing and word natural sense disambiguation. For nearly twenty years now, researchers have tried to tap the contents of machine-readable dictionaries with a view to extracting, formalizing and representing the linguistic information they contain and turning it into formats usable in machine translation, information retrieval, automatic dictionary look-up, question answering, etc. More recently, especially as a result of advances in dictionary making in the Anglo-Saxon world, corpora have become one of the main sources of information for populating the large computational lexica required by any NLP system. Although some researchers claim that pure dictionary research has run its course and that the time has come to envisage applications only, it is far from clear whether all the information contained in MRDs has really been tapped and whether the electronic versions of large commercial dictionaries have yielded all their secrets, making them intellectually less interesting and scientifically less worthy of attention. This is probably a moot point since the new generation of dictionaries are the result of scores of person-years of close scrutiny of corpus-based evidence which had to be dissected, digested, interpreted, condensed and regurgitated by teams of highly skilled lexicographers. Neglecting this data would boil down to reinventing the wheel with imperfect tools, which, in this author's view, pleads for a combination of linguistic resources, viz. existing dictionaries and textual corpora, rather than the exclusion of one resource in favour of the other.

Frame semantics

A seemingly new theory has attracted some attention recently in computational lexicography circles. Frame semantics is indeed at the heart of an ambitious project run by the University of Berkeley in the field of semantic tagging and corpus-based dictionary construction, viz. the FrameNet project (Fillmore & Atkins 1998, Baker et al. 1998, Lowe et al. 1997, Gahl 1998). The aim of this project is to describe word senses by using corpus evidence. At first glance, this very brief description is not original: ever since the publication of Cobuild, the first corpus-based English learners' dictionary (Sinclair 1987), nearly every dictionary project has attempted to do just that. The originality of the FrameNet project is that it aims at including in the resulting lexical database a description of the list of all possible constellations of so-called 'frame elements', which complements the 'traditional' morphosyntactic information one is used to finding in such lexicons. An additional feature of FrameNet is that each word sense is linked to a set of corpus-derived sentences been annotated with frame-semantic that have information. In a way, this can be seen as a form of semantic tagging (see also Fillmore & Atkins 1998).

What are frames?

The basic concept which underlies frame semantics is the frame. It can be defined as a general or specific structure involving entities (frame elements) which participate in these structures. A frame-based lexicon aims at describing the combinatory potential of a given lexical item, which boils down to explicitly indicating how each frame element can be realized at the surface level.

The frame therefore represents a sort of situation, a piece of reality in which keywords are contrasted with one another and can be classified as a function of the relationships which hold between the various actants or frame elements. One of the early examples described by Fillmore¹ is the so-called commercial transaction scene, which involves four frame elements: a seller (S), goods (G), a buyer (B) and the price/money (P). A speaker who

¹ Frame semantics can be seen as a sophisticated development of case grammar. It is not as recent as some might think, however, since Fillmore already laid the foundations of this theory nearly 20 years ago in what might be considered as a seminal paper in which the main concepts were introduced (Fillmore 1982). A decade later, thanks to recent advances in the field of corpus linguistics and the development of corpus query tools, the DELIS European LRE project was to produce the very first fragments of corpus-based lexical descriptions using frame semantics (in the field of perception and speech act vocabulary - see Heid 1994, 1996).

wishes to describe a commercial transaction may resort to a series of verbs such as *sell*, *buy*, *pay*, *charge* or *cost*. The choice of one of these verbs means that the speaker imposes a point of view from which s/he considers the situation as a whole. All these verbs can be contrasted as a function of the ways in which they enable the various frame elements to be realized syntactically. Consider the following sentences which can be considered as paraphrases insofar as they describe the same frame:

- 1. John sold the car to Peter for \$2,000.
- 2. Peter bought the car from John for \$2,000.
- 3. Peter paid John \$2,000 for the car.
- 4. John charged Peter \$2,000 for the car.
- 5. The car cost Peter \$2,000.

The sentences above clearly show that the various frame elements can occupy different positions. In terms of grammatical functions, they can be realized differently, which has strong implications for the lexical description of the verbs. For each lexical entry, the number and nature of the frame elements needs to be specified, together with information on how a given element is to be realized at surface level. Such a description will for instance indicate that the verb buy takes a Buyer (B) as first syntactic actant (subject), goods (G) as second syntactic actant (direct objet) and optionally a Seller (S), appearing in a prepositional phrase introduced by from and money (M), appearing in a prepositional phrase introduced by for. Similarly, the verb charge takes a Seller (S) as first syntactic actant (subject), money (M) as second syntactic actant (direct object), and optionally a Buyer (B) appearing as indirect object and goods (G) appearing as an optional prepositional phrase introduced by for.

It should be pointed out that, unlike case grammar, frame semantics does not postulate the existence of 'universal' frame elements. Rather, they should be seen as heavily dependent on the frame or scenario in which they are to be found. Very much like in plays or in movies, where an actor may play highly different parts, a given lexical item may be assigned different semantic functions, depending on which frame is activated. Consider the following sentences:

- 1. Her doctor bought a superb BMW for £25,000.
- 2. Her doctor drove his BMW at lightning speed around the city.
- 3. Her doctor was able to cure her cancer.

While (1) can undoubtedly be interpreted in terms of the commercial transaction scene described above (the noun *doctor* being an exponent of the Buyer frame element), (2) illustrates the DRIVING frame (see Baker et al. 1998). In this latter frame, the noun *doctor* plays the part of a DRIVER (a primary mover), which appears here as a subject, while the BMW is a VEHICLE and appears as a direct object. Other frame elements for this frame have been identified by the FrameNet researchers, i.e. a CARGO, a RIDER or a PATH, which surfaces in (2) as an oblique complement (*around the city*).

The last sentence above illustrates yet another frame, viz. the HEALTH frame, which is described at length in Lowe et al. (1997). In this frame, the noun *doctor* plays the part of a HEALER, i.e. an individual who tries to improve the health of a PATIENT. In (3), the healer frame element appears as the subject of the verb *cure*, but this verb can also appear with a different constellation of frame elements (a so-called Frame Element Group, or FEG), as is shown in the following examples excerpted from the Cobuild dictionary (Sinclair 1987):

4. It was used as a folk-medicine to cure snake-bite.

In (4), *cure* occurs with a MEDICINE frame element appearing in subject position and a WOUND surfacing as a direct object. Other possible frame elements in the HEALTH frame are PATIENT, DISEASE (see *cancer* in 3), BODY PART, SYMPTOM or TREATMENT.

Frame semantic tagging

Semantic tagging is currently a hot topic in computational lexical semantics. The idea here is to move beyond traditional part-of-speech or syntactic tagging in order to try and assign word senses to lexical items in a corpus. The assignment process can be manual, which is both tedious and time-consuming, and requires special lexicographical skills. It can also be automated and several projects now attempt to use large-scale lexical resources as gold standards, whether they are commercial dictionaries such as the CIDE dictionary (Procter 1995; see Harley & Glennon 1997) or research-oriented lexical databases such as WordNet (Fellbaum 1998).

The FrameNet researchers have developed a number of corpus tools which enable them to quickly browse through corpus data and assign the appropriate frame element tags to the sentences they are examining. They can then visualise the structure of the concordances in which different colours are used for the various frame elements: this enables the linguists to retrieve from the corpus, say, all sentences featuring a given frame element group (e.g. a verb surrounded by a given constellation of frame elements). The frame semantic annotation itself is purely manual, however, and heavily relies on the expertise of the coder, who has to become a skilled lexicologist wellversed in the linguistic theory which underlies the project. In the following sections, we would like to show how a separate resource which was not primarily built with this perspective in mind could be used to partially identify some frame elements and the combinatory potential of a number of lexical items.

A bilingual lexical-semantic database

After realizing that the collocational potential of bilingual commercial dictionaries had never been fully exploited, we embarked on the construction of a lexical-semantic database based on the machine-readable version of the Collins-Robert English-French dictionary (Atkins & Duval 1978, 1st edition). The original idea was to create a multi-access database in which the very rich and sophisticated collocational and thesauric material of the dictionary would be made readily accessible. In addition to the creation of access programs, which were to enable (linguists, lexicographers, NLP designers, users translators...) to surf on the dictionary in a highly opportunistic mode in order to discover implicit information, we also decided to add a semantic layer to the original data, which spurred us to enrich the dictionary with information on the lexical-semantic relationship linking headwords and a series of 'indicators' appearing at word sense level. For space reasons, we cannot go into the details here and will limit ourselves to a general presentation of this database. Fontenelle (1997a, 1997b) provide detailed explanations on the rationale of this project and on its possible applications.

The Collins-Robert bilingual dictionary

Good bilingual dictionaries such as the Collins-Robert dictionary (henceforth CR) provide users with information about contextual restrictions and the conditions which have to be met for a given translation to apply in a given context. They do not simply list possible translations in a row, but use a whole gamut of indicators, synonyms, collocations, semantic restrictions, subject field codes, etc. to guide the translation process. The following system was applied by the CR lexicographers:

- Typical subjects of a verb headword appearing in italics and between square brackets;
- Typical direct objects of a verb or typical noun modified by an adjective appearing in italics (unbracketed);
- Typical noun complements of a noun headword appearing in italics between square brackets;
- Synonyms, paraphrases, micro-definitions appearing in italics, between parentheses;
- Subject fields appearing in italics, between parentheses and with a initial capital letter.

The following examples illustrate this practice, which was applied consistently throughout the dictionary.

fluff vt **a** (also ~ **out**) feathers ébouriffer; pillows, hair faire bouffer. **b** (* do badly) audition, lines in play, exam rater, louper*

grunt vi [pig, person] grogner...
platoon n (Mil) section; [policemen, firemen etc] peloton;
(US Mil) ~ sergeant adjudant
sty n [pigs] porcherie

The information above shows that the dictionary contains a lot of crucial information which can be put to good use in a word sense disambiguation perspective, and more specifically in a translation selection perspective. It shows for example that the verb *fluff* should be translated as *rater* or *louper* in French if it applies to an exam, and that the translation *ébouriffer* is out in this particular context, since the latter normally applies to cases where *feathers* appears in direct object position.²

The availability of the dictionary in machine-readable form, and more specifically in database format³, makes it possible to access the data via other access keys than the traditional alphabetical ordering of the headwords, which is the only access path a user of the paper version can resort to. More specifically, the user can for instance focus on the occurrence of a given item appearing in italics somewhere in the micro-structure of a headword and ask the computer to list all headwords under which this italicized indicator appears. A quick glance at the four examples above shows that *pig* is used under *grunt* and *sty*, but the list of occurrences of *pig* in italics is quite informative. This item indeed appears *under boar, dig, food, geld, grunt, keep, mash, nuzzle, root, root up, rout, slop, snout, sow, sty,* and *swill.*

Lexical functions and Meaning-Text Theory

The data above is undoubtedly interesting insofar as it includes a variety of collocations and semantically-related words which bear some resemblance to what can be extracted when one computes statistics such as Mutual Information scores to discover significant co-occurrence relations (Church & Hanks 1990). The relationships between the various elements differ widely, however, and there is no explicit way of specifying that boar and sow refer to male and female pigs respectively and are therefore closer to each other than, say, grunt or sty. In order to make such distinctions explicit and add a semantic layer to the original dictionary, we decided to label the 70,000-odd pairs of semantically-related items with lexical relations. The mechanisms we opted for was based upon the lexical function paradigm developed by Mel'cuk in the framework of his Meaning-Text Theory (Mel'cuk et al. 1984). The list of lexical functions used in our database and the rationale which underlies the choice of additional relations can be found in Fontenelle (1997a). To illustrate the theory of lexical functions with data borrowed from the CR dictionary, it is sufficient at this stage to understand that a lexical function is a meaning relation between a keyword and other words or phraseological combinations of words. The general form of such a function is f(X)=Y, where X is the keyword and Y is the related item (usually, though not necessarily, a collocate) which has to be selected to express the meaning denoted by f(X). In the data above, the relationship between pig (the italicized item corresponds to the keyword X) and grunt can be represented in terms of the lexical function Son (typical verb for the sound of X), which is written as follows:

Son (pig) = grunt

Similarly, the relationship between *pig* and *sty* was coded in terms of the Sloc lexical function (typical location/place):

Sloc (pig) = sty

We have extended the original Meaning-Text Theory to cater for a number of additional links, such as part-whole relations⁴, or male/female relations. Focusing on the occurrences of *pig*, we are then able to retrieve the data below from the dictionary database. The order applied to display the information here is: dictionary headword, part of speech of the headword, italicized item, French

² One immediately sees the limitations of this approach: in order to save space, the lexicographers have indeed not been able to list all collocates and have selected the most salient or the most frequent ones. The problem is to match a sentence such as "The student fluffed his test" with the second sense of *fluff*, even though *test* is not listed as a possible collocate of the verb. This problem is addressed by the members of the DEFI team in Liège, who use the CR database in addition to a number of other bilingual and monolingual machine-readable dictionaries to automatically select the 'best' translation in context, which, in the present case, forces them, inter alia, to compute the semantic similarity between *test* (the disambiguating context) and *exam* (the information provided in one of the dictionaries). See Michiels (1998) and Dufour (1998) for more details on the DEFI project on word sense disambiguation and translation selection.

³ The structure of the database and the work which was necessary to transform the data from the typesetting tape into a database are described in Fontenelle (1997a).

⁴ Mel'cuk does not consider part-whole relations as lexical functions because they are not one-to-one relations. For information retrieval or language teaching purposes, however, such knowledge is undoubtedly essential and can provide crucial clues when disambiguating word senses. We therefore made use of the Lexical Function mechanism to formalize these relations whenever they were present in the dictionary.

translation of the headword, French translation of the italicized item, lexical function, if any.

boar (n) : $\sim pig \sim => verrat < m> (porc,male)$ dig (vi) : \sim pig \sim => fouiller (porc,) food (n) : $\sim pig \sim => patée < f> (porc,)$ geld (vt) : ~pig~ => châtrer (porc,) grunt (vi) : ~pig~ => grogner (porc,son) keep (vt) : ~pig~ => élever (porc,) $mash(n): \sim pig \sim => pâtée < f>(porc,)$ nuzzle (vi) : ~pig~ => fouiller du groin (porc,) root (vi) : ~pig~ => fouiller (avec le groin) (porc,) root up (vt sep) : ~pig~ => déterrer (porc,) rout (vi) : ~pig~ => fouiller (porc,) slop (n) : ~pig~ => pâtée <f> (porc,) snout (n) : ~pig~ => museau (porc,part) sow (n) : ~pig~ => truie <f> (porc,female) sty (n) : ~pig~ => porcherie <f> (porc,sloc) swill (n) : ~pig~ => pâtée <f> (porc,)

As can be seen above, the lexical function mechanism is not always rich enough to cope with some basic relations. A number of nouns are not assigned any lexical function because the list of 60-odd lexical functions normally includes standard relations, which occur with a large number of keywords and a large number of arguments. It is clear that, from a semantic perspective, some mechanism could be devised to capture the strong similarity between *food*, *mash*, *slop*, and *swill*, which all refer to the typical food of pigs. In terms of frame semantics, these four nouns could be seen as the exponents of a given frame element applying to pigs, which could be called FOOD, for instance.

The data above could also be represented diagrammatically, since the lexical function mechanism makes it possible to group together collocates which share a common meaning component with respect to the node (the keyword). In this way, the bilingual dictionary can be seen as a resource for constructing partial semantic networks, as is shown in Figure 1 (see also Fontenelle 1997b).

The retrieval program associated with the database makes

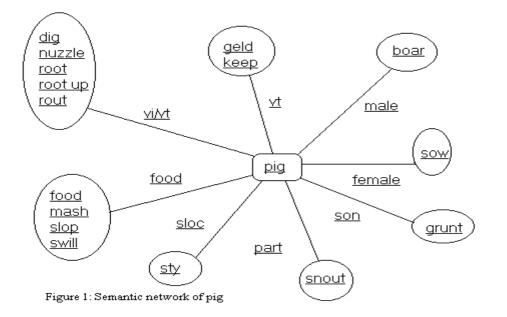
it possible to access the data via any element of the dictionary entry, including the lexical functions which were added subsequently. All these elements can be queried in isolation or in combination with each other. This makes it possible to ask, say, whether there are any verbs expressing the typical sound made by a pig, or to list transitive verbs (part of speech = vt) which can take the word *pig* as direct object, whatever the lexical function associated with it, if any.

Acquiring data for frame semantic descriptions

In this section, we would like to show how the CR database can be used to produce a partial description and fragments of dictionary entries in a frame semantic perspective. It should be pointed out that the Mel'cukian approach normally focuses on standard lexical functions, i.e. relations which are pervasive in general language. Therefore, lexical functions can be seen as a type of "universal" relation with usually unpredictable realizations. In comparison, frame elements are more likely to be highly specific and often apply only to a microscopic world which the frame semanticist tries to describe as minutely as possible. However, one may safely argue that a number of frame elements will probably recur repeatedly across a large number of frames. Frame elements referring to locatives or instruments, for instance, are cases in point. This is just an area where the CR database provides interesting data. Since the query programs also make it possible to concentrate on the realization of a given lexical function, without starting from a given keyword, it is possible to extract from the dictionary the list of all triples featuring the lexical functions Sloc or Sinstr, which denote typical locations or typical instruments associated with a keyword respectively. Such a query will generate hundreds of bilingual records, such as the following combinations:

Sinstr (conjurer) = wand (baguette magique) Sinstr (cowboy) = noose (lasso) Sinstr (hangman) = noose (corde)

Sloc (fox) = earth, hole, kennel (repaire, terrier)



Sloc (bishop) = see (siège épiscopal) Sloc (sentry) = shelter (guérite)

As will become obvious below, however, the dictionary database is also useful to identify the following linguistic elements when describing a given frame:

- The vocabulary used when activating a frame, i.e. the central verbs around which frame elements are going to revolve;
- Frame elements themselves;
- The semantico-syntactic relationship between predicates and frame elements.

As is argued below, all this information may cater for a preliminary and non-exhaustive description of a frame. The idea is then to have this data complemented with corpus data.

The "examination" frame

We would like to pay attention to the "examination" frame, which describes a situation in which someone goes in for an exam and has to satisfy a number of requirements in order to pass this exam. In order to identify the central predicates, i.e. the main vocabulary used to talk about this frame, the starting point can consist in retrieving the information contained in the database for the noun examination. Since it is impossible to predict that only examination has been used as a metalinguistic indicator in the microstructure of the dictionary entries, it is preferable to cast the net somewhat wider and query the database against occurrences of related terms such as exam or test. The list of items associated with these nouns includes the following verbs (see below): be in process, fail, fluff, go in for, hold, pass, prepare, set, sit, supervise, superintend, take, undergo...

A second task is to identify the frame elements which play a part in this frame. Apart from the nouns *examination*, *exam* and *test* themselves, which can be described as a type of central EVENT in this frame, the presence of at least two other frame elements can be identified on the basis of subscripts associated with the main actors (actants in the terminology used by Mel'cuk). The database contains the following records, which point to possible denominations for the first (S1) and second (S2) actants of the nouns *exam* and *examination*:

- entrant (n) : ~exam~ => candidat(e) (examen,s2)
- jury (n) : ~examination~ => jury <m> (examen,s1)

We suggest using the terms EXAMINER for the first actant and EXAMINEE for the second actant. Obviously, the information contained in the dictionary is very limited here and totally unsatisfactory since it does not cater for numerous other possibilities which only a corpus analysis would reveal⁵.

In the Meaning-Text Theory, subscripts also appear in the lexical functions associated with some of the verbs collocating with these nouns. Consider the following examples, excerpted from the database:

fail (vt) : ~examination~ => échouer à (examen,antireal2) fluff (vt) : ~exam~ => rater (examen,antireal2) go in for (vt fus) : ~examination~ => se présenter à (examen,oper2) pass (vt) : ~exam~ => être reçu à (examen,real2) prepare (vi) {TO PREPARE FOR} : ~examination~ => préparer (examen,preparoper2) sit (vt) : ~exam~ => passer (examen,oper2) take (vt) : ~exam~ => passer (examen,oper2) take (vt) : ~test~ => passer (test,oper2) undergo (vt) : ~test~ => subir (test,oper2)

All the verbs above can be used when describing the frame from the perspective of the second actant, in MTT parlance. This means that the second actant, viz. the person who is being examined or tested, is the subject of the verbs above. In this, one clearly sees that there are a number of nearly semantically empty verbs (which some linguists call 'support verbs'), which appear as the exponents of the Oper lexical function. Saying that somebody sits, takes, undergoes or goes in for a test or an exam boils down to saying that he or she is being examined or tested. The outcome of the test can be described in terms of the Real function, which indicates that the requirements have been met and that the outcome of the test is successful (X passed the exam), while AntiReal denotes a failure to comply with these requirements (X *fluffed/failed* the exam).

Note that the lexical functions can be used to account for a different meaning in a cross-linguistic perspective. Consider the following famous false friends in English and in French (*pass* an exam \neq *passer* un examen). These collocations can be represented as follows:

FR: Oper2 (examen) = passer

EN: Real2 (exam) = pass

The data retrieved from the CR database can be represented as in Table 1 below. This table shows the main predicates (verbs) used when activating the examination frame and the frame element groups (FEG) which can be identified on the basis of the information provided by the lexical functions contained in the database. Since three frame elements at least are possible, the figures indicate whether these frames occupy the position of a subject (1) or a direct object (2) of the verb in question. If the frame element appears in the form of a preposition phrase, the preposition heading this PP is indicated. Finally, the first column on the left is used to capture a very broad semantic category inferred from the lexical functions. These categories can be seen in the form of a process, with a beginning (the preparation), a middle (the examination itself and the set of semantically impoverished verbs which can be used to support the noun bases), and an end (the outcome, whether a success or a failure).

⁵ It would be interesting to resort to thesauri to expand the list of possible realizations for some of the frame elements identified here. It is clear that nouns such as *student*, *applicant*, *candidate*, *pupil*, etc. would fall within this category. Nouns such as *professor*, *teacher*, *examiner*, *president*, *jury*, *evaluator*, etc. would be the exponents of the EXAMINER frame element. Finally, it ought to be stressed that the EVENT frame element need not necessarily be realized by the nouns *exam* or *test*. A sentence such as *I failed my Maths A level* (CIDE, s.v. A level) reveals that terms like A level, B level, competition and other very specific items such as *International Baccalaureate* or *IB*

can be considered hyponyms of *examination*, which should be captured in a thesaurus (consider the authentic sentence: "Evans is to allow some pupils to take the International Baccalaureate instead of A-levels", Financial Times, 12 Feb. 2000, p.xii).

	Verb	Examiner	Examinee	Event
PREPARE	Set	1		2
(Prepar)	Prepare		1	For
MAKE/DO	Examine	1	2	
Oper / Func	Sit		1	2 / for
	Take		1	2
	Be in process			1
	Go in for		1	2
	Undergo		1	2
[+Control]	Supervise	1		2
	Superintend	1		2
	Hold	1		2
SUCCEED (Real, Fact)	Get through		1	
	Pass	1	2	
	Pass		1	(2)
FAIL (AntiReal, Liqu)	Carve up	1	2	
	Eliminate	1	2	
	Fail		1	(2)
	Fail	1	2	
	Fluff		1	2
	Plough	1	2	
	Refuse	1	2	
	Reject	1	2	
	Turn down	1	2	
	Weed out	1	2	

Table 1: Frame Element Groups in the "examination" frame

As can be seen above, table 1 also includes a number of frame element groups which do not involve an EVENT (i.e. a hyponym of *exam* or *test*). The verb *fail*, for instance, can appear with different constellations of frame elements, as the following sentences clearly show:

- 1. Many students_[EXAMINEE] failed the driving test_[EVENT].
- 2. The examiners_[EXAMINER] failed him_[EXAMINEE] because he had not answered all the questions.

In order to discover patterns involving EXAMINERS or EXAMINEES, we queried the CR database against the occurrences of a set of prototypical nouns standing for these frame elements, viz. *pupil, candidate, student* or *professor, teacher*. Some of the triples contained in the database are listed below. The semantico-syntactic behaviour of the verbs in question is formalized in Table 1 above, specifying for instance that the intransitive verb *get through* takes an EXAMINEE as a subject to express success or that an EXAMINEE can appear as the direct object (second actant) of a series of verbs expressing failure caused by an EXAMINER. In the latter case, an [EXAMINER] can *carve up / eliminate / fail / plough / refuse / reject / turn down / weed out* an [EXAMINEE].

carve up (vt sep) : ~candidate~ => massacrer [informal] (candidat,liqu)

eliminate (vt) : ~candidate~ => éliminer (candidat,liqu)

examine (vt) : ~candidate~ => examiner (<in> en) (candidat,real2)

fail (vt) : ~candidate~ => refuser (candidat,liqu)

fail (vi) : ~candidate~ => échouer (candidat,antifact0)

get through (vi) : ~candidate~ => être reçu (candidat,fact0)

pass (vt) : ~candidate~ => recevoir (candidat,real2)

plough (vt) : ~candidate~ => recaler [informal] (candidat,liqu)

refuse (vt) : ~candidate~ => refuser (candidat,liqu)

reject (vt) : ~candidate~ => refuser (candidat,liqu)

turn down (vt sep) : ~candidate~ => refuser (candidat,liqu) weed out (vt sep) : ~candidate~ => éliminer (<from> de) (candidat,liqu)

Refining the descriptions with corpus data

The data provided by the CR database should not be considered as the be-all and end-all of the exercise. Clearly, the dictionary database can only offer a starting point leading to a fragmentary description of the behaviour of a number of items participating in a given frame. Fragmentary as they may be, the frame element groups outlined in Table 1 above provide an interesting insight into the general structure of the examination frame. The combinatory potential of its components receives a preliminary description and the lexical functions prove to be interesting clues leading to the discovery of a number of frame elements and to the identification of basic semantic relations holding between them. The notion of subscripts used in Mel'cuk's Meaning-Text Theory to indicate the deep actants of a keyword (see the functions S1, S2, Oper1, Oper2, Real1, etc. above) is particularly interesting insofar as it helps identify the perspective from which the frame is seen when one selects a given predicate to activate it. Such functions are very general, however, and the proper labelling and identification of the frame elements can only be arrived at after a careful, indepth intellectual analysis. The predigested material contained in the database can be used to carry out this type of analysis, without forgetting that corpus data should then be used to complement the descriptions. Corpus evidence would for instance show that at least an additional frame element should be added to those we had already identified. Sentences such as the following (excerpted from the CIDE dictionary) are cases in point since they illustrate the use of other frame elements which could be called SUBJECT, as in (1) and (2) or RESULT, as in (3):

- 1. I passed in history but failed in chemistry.
- 2. She is taking Physics and Maths at A-level.
- 3. John got three passes and four fails in his exams.

In (1), the SUBJECT frame element is introduced by the preposition *in*, while it appears as a direct object of *take* in (2). It is usually realized as a noun corresponding to a traditional discipline studied at school (English, maths, geography...). In (3), the EXAMINEE sits an exam and gets a result which expresses his performance in terms of marks such as *passes*, *fails*, *As*, *Bs*, *Cs*, *distinction*, etc.

Conclusion

The idea of using a lexical-semantic database incorporating Mel'cukian lexical functions in a frame semantic perspective is only in its preliminary stage. Results are encouraging, however, given the emphasis laid by both theories upon a deep semantic description of the actants playing a part in a "linguistic" scenario and of their combinatory potential. Standard lexical functions are obviously too general in some cases to capture finegrained meaning distinctions. They can be used to identify core frame elements, together with their syntax, however, and the collocational database provided by the Collins-Robert bilingual MRD houses data upon which fragments of frame-semantic lexical entries can be based.

Acknowledgements

The original development of the Collins-Robert lexicalsemantic database took place while the author was working at the University of Liège. Thanks are due to the publishers for granting us access to the tapes of the dictionary and for allowing us to go on using it for research purposes.

References

- Atkins, B.T. & Duval, A. (1978): Robert & Collins Dictionnaire Français-Anglais, Anglais-Français, Paris: Le Robert/Glasgow: Collins. (3rd Edition edited by Sinclair, L. & Duval, A.).
- Baker, C., Fillmore, C. & Lowe, J.B. (1998): "The Berkeley FrameNet Project", in *Proceedings of ACL/COLING 1998*.
- Church, K. & Hanks, P. (1990): "Word Association Norms, Mutual Information and Lexicography", in *Computational Linguistics*, 16 (3), pp.22-29.
- Dufour, N. (1998): "Recognizing collocational constraints for translation selection: DEFI's combined approach", in *EURALEX'98 Proceedings* - 8th International Congress of the European Association for Lexicography, University of Liège, Liège: 109-118.
- Fellbaum, C. (ed.) (1998): *WordNet: An Electronic Lexical Database*, MIT Press, Cambridge, MA, London, UK.
- Fillmore, C. (1982): "Frame Semantics", in The Linguistic Society of Korea (ed.), *Linguistics in the Morning Calm*, Seoul, Hanshin, pp.111-137.
- Fillmore, C. & Atkins, B.T.S. (1994): "Starting where the dictionaries stop: the challenge for computational lexicography", in Atkins, B.T.S. & Zampolli, A. (eds) *Computational Approaches to the Lexicon*, Oxford University Press, pp.349-393.
- Fillmore & Atkins (1998): FrameNet and Lexicographic Relevance, *in Proceedings of the Granada Conference on Linguistic Resources*, May, 1998, pp. 417-423.
- Fontenelle (1997a): *Turning a Bilingual Dictionary into a Lexical-Semantic Database*, Tübingen: Max Niemeyer Verlag.
- Fontenelle, Th. (1997b): Using a bilingual dictionary to create semantic networks, *in International Journal of Lexicography*, Vol.10, N°4, OUP, pp.275-303.
- Gahl (1998): Automatic Extraction of Subcategorization Frames for Corpus-Based Dictionary Making, in *Euralex'98 Proceedings* (European Association for Lexicography), Liège, pp. 445-452.
- Harley, A. & Glennon, D. (1997): "Sense Tagging in Action", paper presented at the ACL 1997 Conference on Tagging Text with Lexical Semantics: Why, What and How?
- Heid, U. (1994): "Relating lexicon and corpus: computational support for corpus-based lexicon building in DELIS", *Euralex'94 Proceedings* (European Association for Lexicography), Amsterdam, pp.459-471.
- Heid, U. (1996): "Creating a multilingual data collection for bilingual lexicography from parallel monolingual lexicons", in *Euralex'96 Proceedings* (European Association for Lexicography), University of Göteborg, pp.573-590.
- Lowe, J.B., Baker, C. & Fillmore, C. (1997): "A Frame-Semantic Approach to Semantic Annotation", in *Tagging Text with Lexical Semantics: Why, What, and*

How? – Proceedings of the Workshop, Special Interest Group on the Lexicon, Association for Computational Linguistics, pp.18-24.

- Mel'cuk et al. (1984) : *Dictionnaire Explicatif et Combinatoire du Français Contemporain*, Presses de l'Université de Montréal.
- Michiels, A. (1998): "The DEFI matcher", in *EURALEX'98 Proceedings* 8th International Congress of the European Association for Lexicography, University of Liège, Liège, pp.203-211.
- Procter, P. (ed.) (1995): *Cambridge International Dictionary of English*, Cambridge University Press.
- Sinclair, J. (ed) (1987): *Collins COBUILD English Language Dictionary*, 1st Edition, HarperCollins, Glasgow.