Lexical Functions of *Explanatory Combinatorial Dictionary* for lexicalization in text generation

* Margarita Alonso Ramos
* Agnès Tutin
** Guy Lapalme

* Département de linguistique et de philologie
** Département d'informatique et de recherche opérationnelle
Université de Montréal
C.P. 6128, Québec, Canada
H3C 3J7

Abstract
In this paper, we show that the Lexical Functions (LFs) of the *Explanatory Combinatorial Dictionary* are well suited for addressing some lexical questions in text generation like collocations and textual phenomena. We first describe LFs and divide them into two subsets: paradigmatic LFs and syntagmatic LFs. In particular, we see how paradigmatic LFs can be used for coreferential relations and definite descriptions in text generation.

1. Introduction

Lexical choice cannot be processed in text generation without appealing to a lexicon which takes into account many lexico-semantic relations. The text generation system must be able to treat the immediate and the larger lexical context.
a) The **immediate lexical context** consists of the lexemes that surround the lexical item to be generated. This context must absolutely be taken into account in the case of collocational constraints, which restrict ways of expressing a precise meaning to certain lexical items, for example as in expressions like *pay attention*, *receive attention* or *narrow escape* (Wanner & Bateman 1990; Iordanskaja *et al.* 1991; Nirenburg & Nirenburg 1988; Heid & Raab 1989).
b) The **larger textual context** consists of the linguistic content of previous and subsequent clauses. This context is the source for cohesive links (Halliday & Hasan 1976) with the lexical items to be generated in the current clause, as in:

(1) Professor Elmuck *was lecturing* on lexical functions to third year students. The *lecturer* was interesting and *the audience* was very attentive.

In the second sentence, *lecturer* is coreferential with *Professor Elmuck* and *the audience* is coreferential with *third year students*. These semantic links are due to the lexico-semantic relations between, on the one hand, *lecturer* ("agent noun") and *lecture* , and on the other hand, between *audience* ("patient noun") and *lecture* .

In this paper, we will show that the Lexical Functions of the *Explanatory Combinatorial Dictionary* (hereafter ECD) (Mel'2cuk *et al.* 1988; Mel'2cuk & Polguère 1987) are well suited for these tasks in text generation.

> A LEXICAL FUNCTION F IS A CORRESPONDENCE WHICH ASSOCIATES A LEXICAL ITEM L, CALLED THE KEY WORD OF F, WITH A SET OF LEXICAL ITEMS F(L) - THE VALUE OF F. (Mel'2cuk 1988b).

We will first briefly describe the LFs, dividing them into two subsets:
- **paradigmatic LFs**, which can formalize semantic relationships such as *to lecture* - *lecturer* and *to lecture* - *audience*, and
- **syntagmatic LFs**, which formalize coocurrence relations such as *pay attention*. We will see that each subset addresses different kinds of lexical problems.

In particular, we will show that paradigmatic LFs can solve some interesting textual problems in text generation. We will mainly examine lexical coreferential relations and the introduction of specific definite NPs.

2. Different Kinds of Lexical Functions

2.1. Paradigmatic LFs vs syntagmatic LFs

There are two main kinds of LFs: paradigmatic LFs and syntagmatic LFs.

• **Paradigmatic LFs** are used to associate with a key-word a set of lexical items that share IN THE LEXICON a non trivial semantic component with the key-word. The value of the LF and the key word do not usually form a phrase. For example, S_0(*buy*) = *purchase* (derived noun) or Gener(*cherry*) = *fruit* (generic noun) are paradigmatic LFs.

• **Syntagmatic LFs** are used to formalize a semantic relation (possibly a null one) between two lexemes L1 and L2 which is realized IN THE TEXTUAL STRING in a non predictable way. In other words, lexical choice for expressing a given meaning (the name of the LF) in the context of a lexical item L1 (the key-word) is not free, but this choice is restricted (this is the value L2 of the LF for the key-word L1). For example, the LF for expressing the "intensity" meaning (FL Magn) in the context of the lexeme PLUIE [RAIN] is not free (**grande pluie*, ?*pluie intense*) but the choice is restricted (*pluie torrentielle*) [heavy rain]. In the formalism of LF, we have Magn(*pluie*)= *torrentielle*.

• **Paradigmatic LFs**

We sum up the paradigmatic LFs in Figure 1 below.

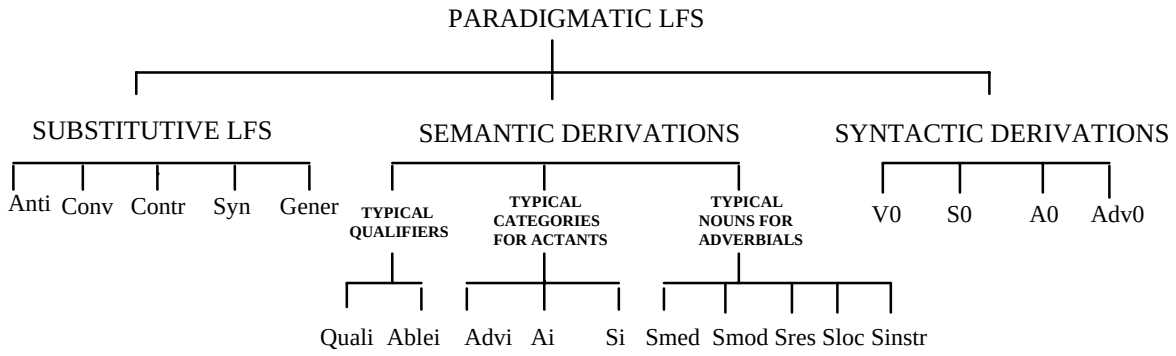

Figure. 1: Paradigmatic LFs

- Substitutive relations:

Substitutive FLs are FLs for which the FL value and the key-word are substituable in the syntagmatic string. Fl value and key-word belong to the same part of speech.

Syn: synonym	Syn(*calling*) = *vocation*
Anti: antonym	Anti(*small*) = *big*
Contr: contrastive	Contr(*laugh*) = *cry*
Conv$_{ijkl}$: conversive	Conv$_{3214}$(*sell*) = *buy*
Gener: generic word	Gener(*apple*) = *fruit*

Semantic derivations contain typical qualifiers, typical categories for actants, typical nouns for adverbials

- Typical categories for actants[1]:

S_i: typical noun for the i-th actant — S_1(*sing*) = *singer*

A_i: typical qualifier for the i-th actant — A_1(*love*) = *in love*

Adv_i: typical adverb for the i-th actant — Adv_1(*speed*) = *at a speed of*

- Typical nouns for adverbials:

S_{instr}: noun for typical instrument — S_{instr}(*paint*) = *brush*

S_{med}: noun for typical mean — S_{med}(*salt*) = [*to*]*salt*

S_{loc}: noun for typical place — S_{loc}(*box*) = *ring*

S_{res}: noun for typical result — S_{res}(*mix*) = *mixture*

S_{mod}: noun for typical mode — S_{mod}(*write*) = (*hand*)*writing*

- Typical qualifiers:

$Able_i$: adjective that qualifies what/who has the potential of being the i-th participant in the situation described by the key-word. — $Able_2$(*eat*) = *edible*

$Qual_i$: adjective that qualifies what/who has a high probability of being $Able_i$ — $Qual_1$(*cheat*) = *dishonnest*

- Syntactic derivations:

For syntactic derivations, the key-word is synonymous with the value, but they do not belong to the same syntactic category. The key-word and the value are not necessarily linked by a morphological derivation.

S_0: derived noun — S_0(*buy*) = *purchase*

A_0: derived adjective — A_0(*sea*) = *maritime*

Adv_0: derived adverb — Adv_0(*final*) = *finally*

• **Syntagmatic LFs** include, roughly, the following kinds (for a complete list of LFs, see Mel'2cuk & Polguère 1987 or Mel'2cuk & Zholkovsky 1988):

- Operator verbs: (**$Oper_i$**, **$Func_i$**, **$Labor_{ij}$**, **$Real_i$**, **$Fact_i$**, **$Labreal_{ij}$**, **Pred**)

Example:

$Oper_i$: Semantically empty verb which takes the i-th actant of the key-word as its subject and the key-word as its direct object. — $Oper_1$(*attention*) = *pay*; $Oper_2$(*attention*) = *receive*

- Qualifiers expressing a particular semantic relationship (**Bon**, **Epit**, **Magn**, **Pos_i**, **Ver**):

Example:

Magn: intensity qualifier — Magn(*escape*) = *narrow*

- Verbs expressing a particular semantic relationship with their actant (**Degrad**, **Nocer**, **Sympt**, **Excess**, **Obstr**, **Involv**, **Prox**, **Manif**):

Example:

Degrad: degradation — Degrad(*milk*) = *sour*

- Prepositions expressing a particular semantic relationship (**Propt**, **Instr**, **$Loc_{in/ad/ab}$**):

Example:

$Loc_{in/ad/ab}$: Place in/to/from — Loc_{in}(*street*) = *on* [*the* ~]

---

[1] In the ECD, a noun like *escape* will be described as a 3 actants noun: X's (actant I) escape from Y (actant II) through Z (actant III). The number of actants (semantic and syntactic) and the way they are realized superficially (for example, with which kind of prepositions) are noted in the government pattern of the lexeme. Government pattern exists not only with verbs and predicative nouns, but also with adjectives and adverbs.

Beside paradigmatic LFs and syntagmatic LFs, there are LFs which can be paradigmatic (the key word and the value share a semantic component) and can be syntagmatic insofar as they can constitute a phrase. These are LF like Mult: Mult(*ship*) = *fleet*, Cap : Cap(*school*) = *principal*, etc. We will call them **mixed** LFs.

In text generation, **Syntagmatic LFs** formalize collocational links that appear in the **immediate lexical context**. The importance of collocational constraints has been emphasized in the literature in text generation and machine translation, and the usefulness of LF formalism to cope with collocational constraints has been highlighted (Heid and Raab 1989; Wanner & Bateman 1990; Iordanskaja *et al*. 1991; Nirenburg & Nirenburg 1988).

In this paper, we will concentrate on **paradigmatic LFs** for treating **larger textual context.** We will see how they are useful to formalize some lexical textual problems: lexical coreference and introduction of specific definite NPs.

2.2 Single LFs, complex LFs and composed LFs

LFs can appear single, complex or composed.

Single LFs appear alone to describe lexical associations like Magn(*rain*) = *heavy*.

Composed LFs are functions for which the set of values is produced through a regular combination (functional composition) of values of the constituant LFs. For example, S_0(Gener(*étuver*)) [steam] = *cuisson* [cooking] can be decomposed: Gener(*étuver*) = *cuire* [to cook] and S_0(*cuire*) = *cuisson* [cooking].

Complex LFs are combinations which cannot be decomposed, like AntiMagn(*blessé*) = *légèrement* ["slightly injured"], that is, roughly LFs for which we cannot deduce the composition from the parts.

In the ECD, only the complex LFs and the single LFs are mentioned, because composed LFs can be reconstructed. Nevertheless, lexical relations that can be formalized by LF regular compositions must be studied because they appear in texts.

We will see in (3.3) how composed LFs can be exploited for creating coreferential relations.

3. Use of paradigmatic LFs for coreference in text generation

3.1 Lexical coreference in text generation

In text generation, textual context must absolutely be taken into account because texts like the following one are totally unacceptable:

(2) a. Prepare the carrots, the celery and the asparagus.
b. Cook the carrots, the celery and the asparagus in boiling water.
c. Take the carrots, the celery and the asparagus out after 10 minutes.

d. Cook *the vegetables* in boiling water and take *them* out after 10 minutes.

To avoid the unacceptable redundancy that we notice in 2b and 2c, we need to introduce anaphora, for example *them* or *the vegetables*, like in 2d.

For us, a textual element is a **lexical coreferential anaphor** of an antecedent, a textual element previously introduced if

- a) it has the same referent that its antecedent,

- b) it belongs to an open lexical class[2] , and
- c) the anaphor and the antecedent share a semantic link.

For example, in 2d, *the vegetables* is a lexical coreferential anaphor of *the carrots, the celery and the asparagus* but *them* is not one, because it does not belong to an open lexical class. In the following example,

(3) *Edith Cresson* arrived Monday at 9:00. At 11:00, *the Prime Minister of France* gave a press conference.

Prime Minister of France is a coreferential anaphor of *Edith Cresson*, but not a lexical one, because the coreferential link is not created by semantics but by world knowledge.

For creating lexical coreferential links in text generation, it is necessary to appeal to a large number of lexico-semantic relations. For example, let us imagine an underlying conceptual representation as a sequence of frame-like propositions:

```
{lecture:
        agent: Elmuck_1
        destination: Students_1
        topic: Fonctions_1
}
{interesting:
        agent: Elmuck_1
}
{attentive:
        agent: Students_1
}
```

Figure 2: Simplified conceptual input for 4a, 4b et 4c

Many lexicalizations are possible for the attribute values of this representation. For example, after the first sentence:

(4a) Professor Elmuck was lecturing on Lexical Functions to third year students.

Relying strictly on lexico-semantic and grammatical data, we could produce:

(4b) (He/ The teacher/ Professor Elmuck/ The Professor/ The lecturer) was interesting.
(4c) (?They/ The third year students/ The students/ The audience) was/were very attentive.

The lexical choice here depends strongly on the type of text to be generated. For example, coreferential links like *Professor Elmuck lectured ... lecturer* are frequent in journalistic texts but rare in technical reports. We will not here discuss strategies for lexical contextual choice based on text types but we will discuss how lexical links can be created. We will try to show that LFs allow us to formalize a large number of lexico-semantic relations for lexical coreference. We will see that single LFs and composed LFs can be used for this task.

3.1.1 Single LFs for coreference relations

First, lexical coreferential links may appear with what is called, according to Halliday & Hasan (1976) reiteration, i.e. repetitions, synonyms, near-synonyms, superordinates. These kinds of lexico-semantic links appear in LFs, for example in:

Syn(*churchman*) = *clergyman*
(5) Daniel confessed to *his churchman*. *The clergyman* blushed while listening to Daniel's sins.

---

[2] Lexical open classes contain the following parts of speech: nouns, verbs, adjectives and adverbs.

Gener(*lamb*) = *meat*

(6) Buy *New Zealand lamb*. *The meat* must be very fresh.

Beside these traditional coreferential relations, we can also use syntactic derivations LFs such as S_0, as in:

S_0(*buy*) = *purchase*

(7) Mary went to St Lawrence street *to buy clothes*. *Her purchases* made, she went back home.

Nervertheless, the use of a conversive relation does not seem to be allowed.

(8) $Conv_{3124}$(*sale*) = *purchase*
The sale of the house has been a long process. ?* *The purchase* has been uneasy.

Until now, we have identified some types of lexical coreference between lexical items which maintain lexico-semantic relations and enable a coreferential link between the antecedent (the key-word) and the lexical anaphor (the LF value). **Direct lexical coreference** occurs between lexical items that maintain a lexical relationship directly formalizable through a LF like Syn, Gener, S_0, $Conv_{ijkl}$ in some rare cases.

But, there can be coreference, not between the key-word and the LF value, but between an actant or an adverbial of the key-word and the value as we saw in (1) between *Professor Elmuck* and *lecturer* and between *third year students* and *audience*. LFs can thus be used to formalize **indirect lexical coreference** when coreference appears between lexical items and a dependent (actantial or adverbial) if one maintains a lexical relationship directly formalizable through a LF.

In the figure below, we give examples of direct and indirect coreference.

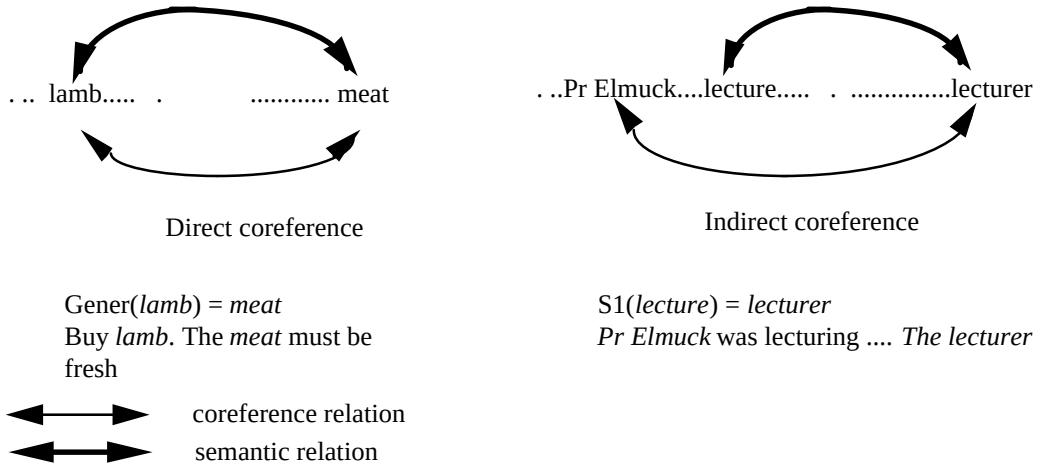

Figure 2. Direct and indirect coreference

In (1), *lecturer* is coreferential with *Professor Elmuck* and *third year students* is coreferential with *audience*. Coreference can be established with the help of the actantial lexico-semantic relations. *Professor Elmuck* is the subject of *to lecture* (first actant) and as such, it can be corefered to with the S_1 of *to lecture* , *lecturer*, and *third year students* can be refered to as the S_2 of *lecture*, *audience*. The actant number (1,2,3,...) corresponds to the semantic actant of the lexeme. For example, the lexeme [TO] LECTURE will be described as a three actants verb: Someone (actant 1) lectures to someone else (actant 2) about something (actant 3).

In the same way, typical nouns for adverbials can be used for an indirect coreferential relation. For example, in:

(9) Marguerite and Jean skated on *the Rideau canal* . *This skating rink* is 8 km long.

The prepositional phrase *on the Rideau canal* is dependent on the key-word *skate* in a certain semantic adverbial relation (location) and this phrase will be coreferred to by the typical adverbial, *skating rink*, of the same semantic relation (Sloc) of the key-word, *skate*.

The typical nouns for adverbials S_{med}, S_{loc}, S_{instr} and S_{mod} can be used for an indirect coreference relation.

Nevertheless, S_{res} (Typical result) is different. According to our analysis, the value of a S_{res} function is not a typical noun for adverbials because it cannot be coreferential with an adverbial of the key-word with which it is semantically linked. It is the typical noun denoting the result of a physical transformation. In the following example,

(10) Mix *the eggs*$_i$ *and the milk* with *the flour*$_j$. Pour *the* ***mixture***$_{i+j}$ in the pot.

the Sres of *mix*, *mixture* is coreferential with the set of *eggs*, *milk* and *flour* that has been affected by the action of mixing, i.e. with the actants II and III of *mix*. It is not a simple case of direct coreference because, in this case, these items have been affected by a complete transformation of their charasteristics.

3.1.2. Composed LFs for coreference relations

As we have suggested earlier, composed LFs can be profitably exploited because they also appear in coreference relations. In this section, we have only observed some of the most frequent compositions. This is just an exploratory work and we have not systematically studied what would prevent lexical coreference.

In the following table, we give some examples of LF compositions.

LFs or composition of LFs	**key word**	**values**
Gener Gener Gener Gener Gener Syn $Conv_{3214}$ S_0 S_0	achat [purchase] vente [sale] transaction [deal] auto [car] voiture [car] voiture [car] acheter [buy] acheter [buy] vendre [sell]	transaction [deal] transaction [deal] action [action] véhicule [vehicle] véhicule [vehicle] auto [car] vendre [sell] achat [purchase] vente [sale]
Gener o Gener Gener o Syn Gener o $Conv_{3214}$ Gener o S_0 S0 o $Conv_{3124}$ $Conv_{3214}$ o S_0	achat [purchase] auto [car] vente [sale] acheter [buy] acheter [buy] acheter [buy]	action [action] véhicule [vehicle] transaction [transaction] transaction [transaction] vente [sale] vente [sale]

Table 1: LFs and compositions of LFs for direct coreference links

The following facts must be noted about compositions:

- The Composition relation is not commutative

For example, S_0(Gener(*étuver*)) [steam] = *cuisson* [cooking] but Gener(S_0(*étuver*)) does not produce any value because S_0(*étuver*) does not have a value.

- Many compositions should be reduced, for example, Gener(Syn(X)) is equivalent to Gener(X) as we can see in Table 1. Reducible LFs should not be calculated (See Décary 1986 and Décary & Lapalme 1990 for an examination of reducible LFs).

Compositions involving direct coreference LFs produce direct coreference compositions:

S_0(Gener(*étuver*)) = *cuisson*

(11) Faire ***étuver*** *la viande*. A la fin de *la* ***cuisson***, ajouter les épices.
[Let the meat steam. At the end of the cooking, add the spices].

Compositions involving a direct coreference LF and an indirect LF (or an indirect coreference composition) produce indirect coreference compositions. For example,

Gener(S_1(*marathon*)) [marathon] = *coureur* [runner]
S_1(*marathon*) = *marathonien* [marathonian]et Gener(*marathonien*) = *coureur*

(12) *Jean* fait souvent le marathon. Ce *coureur* est infatigable.
[Jean often does the marathon. The runner is indefatigable]

In the following chart, we have shown compositions of two LFs that can be used in coreference relations.
DC indicates that the LF can create a direct coreference relation, IC indicates that the LF can create an indirect coreference relation. The entry in the table corresponds to LF_i o LF_j.
The product of the composition is indicated at the intersection. This can be:
- A DC: Ex: Gener (DC) o S0 (DC) => DC.
- An IC: Ex: Gener (DC) o Si (IC) => IC.
- A * before DC or IC indicates that the composition is reducible to a single function. Ex: Gener o Syn = Gener (*DC)
- No result: composition is absurd or does not produce anything.

	Type	Gener	Syn	S_0	S_i	S_{instr}	S_{med}	S_{res}	S_{loc}	S_{mod}
Gener	DC	DC	*DC	DC	IC	IC	IC	DC	IC	IC
Syn	DC	*DC	*DC	*DC	*IC	*IC	*IC	*IC	*IC	*IC
S_0	DC	DC	*DC	--------	--------	--------	--------	--------	--------	--------
S_i	IC	IC	*IC	*IC	--------	--------	--------	--------	--------	--------
S_{instr}	IC	IC	*IC	*IC	--------	--------	--------	--------	--------	--------
S_{med}	IC	IC	*IC	*IC	--------	--------	--------	--------	--------	--------
S_{res}	?DC	?DC	*DC	*DC	--------	--------	--------	--------	--------	--------
S_{loc}	IC	IC	*IC	*IC	--------	--------	--------	--------	--------	--------
S_{mod}	IC	IC	*IC	*IC	--------	--------	--------	--------	--------	--------

Fig. 4: Kinds of coreferential links produced by compositions of two LFs

Thus, LF formalism enables one to systematize many coreference links between open class words in a text. These links are either direct, or indirect if the link is formed by another lexical item. Finally, coreference links can be formalized by single LFs or by compositions of LFs.

In the perspective of text generation, this formalism appears very interesting for building coreferential relations. To point back to a referent already introduced, LFs and compositions of LFs offer many possible ways for lexicalizing a given referent. For example, let us suppose that after having introduced the following sentence,

(13) a.Laisser étuver la viande. [Let the meat steam.]

we have to refer again to the action *la viande étuve*. We could try to use a nominalization (S_0). But, as there is no nominalization for the verb *étuver*, we could use instead the nominalization of the generic term, S_0(Gener(*étuver*)) = *cuisson*. We could thus produce the following sentence:

(13) b. A la fin de la cuisson, ajouter les épices
[At the end of cooking, add the spices]

Of course, we do not claim here that LFs cover all the cases of coreference. It has been noted (Hirst 1981) that coreference often involves non linguistic relations (as we saw in example (3)).

3.2 Introduction of the specific definite article and lexical associations

Choosing between the indefinite and the definite article is sometimes hard in text generation: it depends on many factors, such as extra-linguistic knowledge, situational knowledge and lexical knowledge. Following the linguistic frame defined by Hawkins (1977) for the use of definite article, we will show in this section that the formalism of paradigmatic LFs can be usefully exploited to clarify some lexical knowledge.

3.2.1 Hawkins' approach of definite article

It has been noted that the specific definite article obeys many constraints, certain of which are textual. Within the frame of pragmatics, Hawkins (1977) identified four sets of shared knowledge between the speaker and the hearer that enable the introduction of a specific definite article:
- The **previous discourse set**: this includes the antecedents enabling one to make coreferential links.
- The **immediate situation set**: this enables the use of a definite deictic article. The definite NP is used to refer to something that is in the immediate environment.
- The **larger environment set**: this enables the introduction of referents whose knowledge is implicitly shared by the speaker and the hearer.
- The **association set**: this includes knowledge directly "activated" by lexical items of the previous discourse.

It seems possible and necessary to determine in a more linguistic way some of these shared sets. We will exclude from our study the immediate situation set, which is linked to the immediate environment, and the larger environment set, which contains world knowledge, because none is linguistic. Nevertheless, the previous discourse set and the association set can be described in a linguistic way.

3.2.2 The previous discourse set

The previous discourse set corresponds *grosso modo* to the antecedents that appear in the coreference relations. Each time a lexical coreferential anaphora is introduced, it can include a definite article. We will not go into detail here on these relations, because we have treated them in the previous section.

3.2.3 The association set

The definite article can be introduced in a text when it appears in a NP whose meaning is linked "paradigmatically" with one or several lexical items previously introduced.

The association set does not appear very clearly defined by Hawkins: the assocations sometimes seem to be encyclopaedic and they sometimes seem to be lexical. Besides, Hawkins does not detail the associations that really enable the introduction of the definite article. His perspective is more based on comprehension than on production.

We will clearly distinguish lexico-semantic relations from encyclopaedic ones. We will say that there is a lexico-semantic relation between two lexical items if they share a non trivial meaning

component. For us, the prototypical aspect will be part of the encyclopaedic knowledge (Lakoff 1988; Kleiber 1990).

As a set of associations, we will use paradigmatic LFs and we will only treat LFs whose values can be nouns.

Typical nouns for actants

When introduced in a text, typical nouns for actants are not always coreferential as in :

(14) Le Professeur Elmuck faisait une conférence. Le sujet était intéressant et l'auditoire était attentif.[3]
[Professor Elmuck was lecturing. The topic was interesting and the audience was attentive.]

In this example, the value of S_3(*conférence*) [lecture] = *sujet* [topic] and the value of S_2(*conférence*) = *auditoire* [audience] are introduced by a specific definite article. Is this always possible for typical actants? For us, the answer is probably "yes" for the following reason: the referents which correspond to typical nouns for actants are in some way implied. As soon as the key word is uttered, the actants are "activated" because they belong to the government pattern of the word. As they are typical nouns, they fit perfectly into the semantic slots for each actant.

Typical nouns for adverbials

Since they are introduced after the key word in a text, do the typical nouns for adverbials enable the introduction of the definite article? The relationships S_{loc}, S_{instr}, S_{med} and S_{mod} maintained with the keyword are looser than those of the typical nouns for actants. Nevertheless, they seem to enable the definite article to appear, as in the following examples:

S_{instr}(peindre) = pinceau
(15) Jean peignait ses murs. Soudain, (*la/ une) mouche s'est collée sur (le/ ? un) pinceau.
[Jean was painting his walls. Suddenly, (*the/ a) fly stuck on (the/ *a) paintbrush.]

S_{loc}(boxeur) = ring
(16) Le boxeur encaissait de nombreux coups. Soudain, il s'affala sur ((le / un*) ring / (*la / une) chaise).
[The boxer was taking many blows. Suddenly, he collapsed on ((the / *a) ring/ (*the / a) chair).]

S_{med}(peindre) = peinture
(17) Jean peignait ses murs. Soudain, ((la / de la) peinture / (*l' / de l') eau) coula par terre.
[Jean was painting his walls. Suddenly, (the/ •Ø) paint (*the/ Ø) water flew on the floor.]

In these examples, the indefinite article with the value of the function is used to introduce a referent which appears to be out of the situation created by the key word. Thus, in (15b), *un pinceau* (a paintbrush) cannot designate the paintbrush that Jean uses in the situation (15a) but another paintbrush. In (16b), the use of the indefinite article is not allowed because, in the situation (16a), there is only one ring, and no other ring can be thought of.

As we saw previously, S_{res} often appears as a complex coreferential anaphor when there is a change of state of the referents and in coreferential cases, the use of the definite article is very natural. There are nevertheless some cases as the following where a new referent is naturally introduced by a definite article, because, though the referent has not been previously introduced, it is implied by the meaning of the key word. For example, in:

S_{res}(photocopier) = photocopie
(18) a. Le professeur Elmuck a photocopié les examens des étudiants.
b. Il a fait tomber ((les / des) photocopies / (* les/ des) feuilles blanches) en faisant un geste brusque.
[Professor Elmuck photocopied the student exams. He dropped ((the/ Ø) photocopies) / (* the/ Ø) paper sheets) in making an abrupt gesture.]

---

[3]Because the meaning of the definite article can be slightly different in English and in French, we will give French examples here.

les photocopies is not coreferential. It introduces a new object which did not exist before in (18a) but which is the result of the whole process. The indefinite article in (18b) is used to designate the photocopies which are out of the situation (18a).

Substitutive relations

Contr and Anti are the only paradigmatic functions that do not enable introduction of coreferents. They do not seem to allow the introduction of the definite article either. Nevertheless, they introduce coherence relations because there is a thematic continuity in the texts, with the words belonging to the same semantic fields.

Thus, LFs formalize some of the semantic relations that allow the introduction of the definite article.
We should add the following remarks:
- Many other semantic relations can introduce a definite article. In particular, we should say that many of these semantic relations appear in the ECD within the definition of lexemes.
- Many of the associations enabling the introduction of definite article in texts are not lexico-semantic ones but encyclopaedic ones (Fradin 1983).
- We should envisage LF compositions for this application.

To sum up, LFs can in texts formalize some of the introductions of the definite article, either for coferential relations (previous discourse set, in Hawkins'words) or for the association set.

4. Conclusion

In this paper, we have shown the usefulness of paradigmatic LFs to treat lexico-semantic phenomena in text generation.

We first divided LFs into two subsets: paradigmatic LFs and syntagmatic LFs.

We showed that paradigmatic LFs are an appropriate formalism for treating lexicalization in textual context:
- They can be exploited to systematize lexical coreferential relations between a lexical anaphor and its coreferential antecedent. These relations are either direct, or indirect if they are realized through a third lexeme. Coreferential relations can be formalized through a single LF or composed LFs.
- Paradigmatic FLs enable to precise in a more linguistic way the "discourse previous set" and the "association set", defined by Hawkins (1977), for explaining the introduction of the definite NP:
• The previous discourse set contains the set of antecedents which are pointed back to by a coreferential lexical anaphor introduced with the help of the definite article.
• The association set can be partially formalized by some paradigmatic LFs. Typical nouns for actants and typical nouns for adverbials can easily be introduced with a definite article when they appear in the text immediately after their key-words, being implied by them.

Many things remain to be done:
. We should study the question of LF compositions in depth and establish a formal grammar of LF compositions. This problem has just been touched upon here.
. We should test LFs and LF compositions on a large corpus of texts to check whether the paradigmatic LFs are really productive for lexical cohesion.
. We should study thoroughly strategies for implementations that use LFs because we have not addressed this problem here. Nevertheless, let us indicate that some paradigmatic functions have been implemented in Kosseim 1991 for coreferential relations in text generation. Besides, text

generation systems using Meaning-Text Theory, and especially paraphrase (Mel'2cuk 1988b), necessarily make extensive use of some paradigmatic lexical functions (Iordanskaja *et al*. 1991, Boyer & Lapalme 1985).

Acknowledgements

We would like to thank Igor Mel'2cuk for fruitful discussions and helpful suggestions.

References

Alonso Ramos M. & A. Tutin (1991). *Les fonctions lexicales du Dictionnaire Explicatif et Combinatoire pour l'étude de la cohésion lexicale*, manuscrit, Université de Montréal.

Boyer M. & G. Lapalme (1985). Generating Paraphrases from Meaning-Text Semantic Networks, *Computational Linguistics*., I, 3-4. 103-117.

Décary M & G. Lapalme (1990). An Editor for the Explanatory Dictionary of Contemporary French (DECFC), *Computational Linguistics*, 16, 3, 145-154.

Décary M. (1986). *Un éditeur spécialisé pour le Dictionnaire Explicatif et Combinatoire du Français Contemporain*., M.Sc. Thesis, Département d'informatique et de recherche opérationnelle, Université de Montréal.

Fradin B. (1984). Anaphorisation et stéréotypes nominaux,*Lingua* 64, 325-369.

Halliday M.A.K & R. Hasan (1976). *Cohesion in English* , London, Longman.

Hawkins J.A. (1978). *Definiteness and Indefiniteness*, London, Croom Helm.

Heid U. & S. Raab (1989). Collocations in Multilingual Generation, *Proceedings of EACL*, 130-136.

Hirst G. (1981). *Anaphora in natural language understanding: a survey*, Berlin, Springer-Verlag.

Iordanskaja L., R. Kittredge & A. Polguère (1991). Lexical Selection and Paraphrase in a Meaning-Text Generation Model in C.L. Paris, W. R. Swartout & W.C. Mann eds.,*Natural Language Generation in Artificial Intelligence and Computational Linguistics* , 293-312.

Kleiber G. (1990). *La sémantique des prototypes: Catégories et sens lexical*, Paris, PUF.

Kosseim L. (1992). *Génération automatique de procédés cohésifs dans les recettes de cuisine*, M.sc. thesis, Département d'informatique et de recherche opérationnelle, Université de Montréal.

Lakoff G.(1988). Cognitive Semantics, in U. Eco, M. Santambrogio & P. Violi eds, *Meaning and Mental Representation* , Bloomington and Indianapolis, Indiana University Press, 118-154.

Mel'2cuk I. & A. Polguère (1987). A Formal Lexicon in the Meaning-Text Theory (or how to do Lexica with Words), *Computational Linguistics*, 13, 3-4, 1987, 261-275.

Mel'2cuk I. & Zholkovsky, D (1970). Sur la synthèse sémantique, *T.A. Informations*, 2, Paris, Klincksieck, 1-85.

Mel'2cuk I. *et al.* (1984). *Dictionnaire Explicatif et Combinatoire du Français Contemporain. Recherches Lexico-sémantiques I* .Montréal, Presses de l'Université de Montréal.

Mel'2cuk I.A. *et al.* (1988a). *Dictionnaire Explicatif et Combinatoire du Français Contemporain. Recherches Lexico-sémantiques II.* Montréal, Presses de l'Université de Montréal.

Mel'2cuk I. (1988b). Paraphrase et lexique dans la Théorie Sens-Texte,*Cahiers de Lexicologie*, LII, 5-50 et LIII, 5-53.

Meyer I. & J. Steele (1990) Lexical Functions in the Explanatory Combinatorial Dictionary: Kinds and Definitions, in J. Steele ed, *Meaning Text Theory* , Ottawa University Press.

Nirenburg S. & I. Nirenburg (1988) A Framework for Lexical Selection in Natural Language Generation, in *Proceedings of COLING 88* , Budapest, 471-475.

Putnam H. (1975) The Meaning of 'meaning', in *Mind, Language and Reality*, Philosophical Papers, 2, Cambridge, Cambridge University Press, 215-271

Wanner L. & J.A. Bateman (1990) Lexical Cooccurrence Relations in Text Generation, *Proceedings of the Fifth International Workshop on Natural Language Generation*, Dawson, Pennsylvania, 31-38.