Paraphrastic grammars

Claire Gardent, Marilisa Amoia, Evelyne Jacquey

Paraphrastic Grammars

Claire Gardent
CNRS-LORIA, Nancy
France
Claire.Gardent@loria.fr

Marilisa Amoia
Computational Linguistics
University of Saarbruecken
Germany
amoia@coli.uni-sb.de

Evelyne Jacquey
CNRS-ATILF, Nancy
France
Evelyne.Jacquey@atilf.fr

Abstract

Arguably, grammars which associate natural language expressions not only with a syntactic but also with a semantic representation, should do so in a way that capture paraphrasing relations between sentences whose core semantics are equivalent. Yet existing semantic grammars fail to do so. In this paper, we describe an ongoing project whose aim is the production of a "paraphrastic grammar" that is, a grammar which associates paraphrases with identical semantic representations. We begin by proposing a typology of paraphrases. We then show how this typology can be used to simultaneously guide the development of a grammar and of a testsuite designed to support the evaluation of this grammar.

1 Introduction

A salient feature of natural language is that it allows paraphrases that is, it allows different verbalisations of the same content. Thus although the various verbalisations in (1) may have different pragmatic or communicative values (with respect for instance to topicalisation, presuppositions or focus/ground partitioning), they all share a core semantic content, the content approximated by a traditional montagovian compositional semantics.

(1) a. La croisière coûte cher.
Lit. *the cruse is expensive*
b. Le coût de la croisière est élevé.
Lit. *the cost of the cruse is high*
c. La croisière a un coût élevé
Lit. *the cruse has a high cost*

Linguists have long noticed the pervasiveness of paraphrases in natural language and attempted to caracterise it. Thus for instance Chomsky's "transformations" capture the relation between one core meaning (a deep structure in Chomsky's terms) and several surface realisations (for instance, between the passive and the active form of the same sentence) while (Mel'čuk, 1988) presents sixty paraphrastic rules designed to account for paraphrastic relations between sentences.

More recently, work in information extraction (IE) and question answering (QA) has triggered a renewed research interest in paraphrases as IE and QA systems typically need to be able to recognise various verbalisations of the content. Because of the large, open domain corpora these systems deal with, coverage and robustness are key issues and much on the work on paraphrases in that domain is based on automatic learning techniques. For instance, (Lin and Pantel, 2001) acquire two-argument templates (inference rules) from corpora using an extended version of the distributional analysis in which paths in dependency trees that have similar arguments are taken to be close in meaning. Similarly, (Barzilay and Lee, 2003) and (Shinyanma et al., 2002) learn sentence level paraphrase templates from a corpus of news articles stemming from different news source. And (Glickman and Dagan, 2003) use clustering and similarity measures to identify similar contexts in a single corpus and extract verbal paraphrases from these contexts.

Such machine learning approaches have known pros and cons. On the one hand, they produce large scale resources at little man labour cost. On the other hand, the degree of descriptive abstraction offered by the list of inference or paraphrase rules they output is low.

We chose to investigate an alternative research direction by aiming to develop a "paraphrastic grammar" that is, a grammar which captures the paraphrastic relations between linguistic structures[1]. Based on a computational grammar that associates natural language expressions with both a syntactic and a semantic representation, a paraphrastic gram-

[1] As we shall briefly discuss in section 4, the grammar is developed with the help of a meta-grammar (Candito, 1999) thus ensuring an additional level of abstraction. The metagrammar is an abstract specification of the linguistic properties (phrase structure, valency, realisation of grammatical functions etc.) encoded in the grammar basic units. This specification is then compiled to automatically produce a specific grammar.

mar is a grammar that moreover associates paraphrases with the same semantic representation. That is, contrary to machine learning based approaches which relate paraphrases via sentence patterns, the paraphrastic grammar approach relates paraphrases via a common semantic representation. In this way, the paraphrastic approach provides an interesting alternative basis for generation from conceptual representations and for the inference-based, deep semantic processing of the kind that is ultimately needed for high quality question answering.

Specifically, we aim at developing a paraphrastic grammar for French, based on the Tree Adjoining Grammar (TAG) developed for this language by Anne Abeillé (Abeillé, 2002).

The paper is structured as follows. We start by proposing a typology of the paraphrastic means made available by natural language. We then show how this typology can be used to develop a testsuite for developing and evaluating a paraphrastic grammar. Finally, we highlight some of the issues arising when developing a paraphrastic grammar.

2 Classifying paraphrases

A paraphrastic grammar should capture the various means made available by natural language to support paraphrasing. But what are those means? We distinguish here between three main classes namely, parallel, shuffling and definitional paraphrastic means.

Parallel paraphrastic means. A parallel paraphrase can hold either between two non predicative lexical units (words or multi word expressions) modulo negation or between two predicative units of identical arity. If it holds between predicative units, the mapping linking grammatical functions (subject, objects, etc.) and thematic roles (agent, theme, etc.) must be the same. Depending on whether or not negation is involved, semantic equivalence will futhermore obtain either through synonymy or through antonymy.

As illustrated in Figure 1, **synonymy** can be further divided in a number of cases depending on various morphological and syntactic criteria. The classification criteria used involve :

- Syntactic category: Do the synonyms have the same syntactic category?
- Morphological relatedness: Do the synonyms contain words that are morphologically related?
- Form: Are the synonyms simple lexical units or multi word expressions?

As for **antonymy**, we distinguish between trans and intracategorial antonymy:

(2) Jean est lent/Jean n'est pas rapide.
Jean is slow/Jean is not fast.
lent/rapide, intracategorial
Jean a cesser de fumer/Jean ne fume plus.
Jean has stopped smoking/Jean smokes no more.
cesse de/ne ... plus, transcategorial

Shuffling paraphrastic means. When a semantic equivalence holds between predicative units with distinct grammatical functions/thematic role linking, we speak of **shuffling paraphrases**. Such paraphrases can be realised either by means of argument preserving alternations (in the sense of Beth Levin, cf. (4)) or using a converse construction (cf. 3)[2].

(3) a Jean donne un livre à Marie.
Jean gives a book to Marie.
Marie reçoit un livre de Jean
Jean receives a book from Marie.

b Jean est le parent de Marie.
Jean is the parent of Marie.
Marie est l'enfant de Jean.
Marie is the child of Jean.

(4) a. Cette clé ouvre le coffre fort
This key opens the safe.
Le coffre fort s'ouvre avec cette clé
The safe opens with this key.

b. Jean mange une pomme
Jean eats an apple.
une pomme est mangée par Jean
An apple is eaten by Jean.
Il a été mangé une pomme par Jean.
There has been an apple eaten by Jean.

c. L'eau remplit la cruche
The water fills the jug .
La cruche se remplit d'eau
The jug fills with water.
On remplit la cruche d'eau
One fills the jug with water.

d. Le laboratoire fusionne avec l'entreprise
The laboratory merges with the firm.
le laboratoire et l'entreprise fusionnent
The laboratory and the firm merge.

e. Jean frappe le mur avec un baton
Jean hit the wall with a stick.

[2]Obviously, the english translations do not reflect the acceptability of the french equivalent.

Same synt. categories	Same morph. family	Form	Example
yes	no	word/word	policier, flic
yes	yes	word/mwe	conseiller, donner conseil
yes	no	word/mwe	s'exprimer sur, donner son avis sur
yes	no	mwe/mwe	donner carte blanche à, laisser tout pouvoir
no	yes	word/word	construire, construction
no	no	word/word	candidature à, briguer

Figure 1: Synonymy

Jean frappe le baton sur le mur.
Jean hit the stick on the wall.

f. Je fournis des livres à Jean
I provide books to Jean.
Je fournis Jean en livre
I provide Jean with books.

Definitional paraphrastic means. Third, we call "defi nitional paraphrases" semantic equivalences that hold between a lexical unit and a phrase consisting of more than one lexical unit. The phrase in this case, defi nes the meaning of the lexical unit. Since defi nitions are notoriously diffi cult to decide upon, we restrict ourselves here to such defi nitions as can be given by derivational morphology that is, defi nitions based on a word that is morphologically linked to the defi niendum (cf. 5).

(5) a. Le conducteur de la BMW est chauve
The driver of the BMW is bald.
La personne qui conduit la BMW est chauve
The person who drives the BMW is bald.

b. Cet outil est paramétrable
This tool is parameterisable.
Cet outil peut être paramétré
This tool can be parameterised.

3 Developing a paraphrase testsuite

Based on the above typology, we can systematically construct a testsuite for developing and evaluating a paraphrastic grammar. Indeed, when developing a grammar, it is necessary to have some means of assessing both the coverage of the grammar (does it generate all the sentences of the described language?) and its degree of overgeneration (does it generate only the sentences of the described language?) While corpus driven efforts along the PARSEVAL lines (Black et al., 1991) are good at giving some measure of a grammar coverage, they are not suitable for fi ner grained analysis and in particular, for progress evaluation, regression testing and comparative report generation. Another known method consists in developing and using a test suite that is, a set of negative and positive items against which the grammar can be systematically tested. For english, there is for instance the 15 year old Hewlett-Packard test suite, a simple text fi le listing test sentences and grouping them according to linguistics phenomena (Flickinger et al., 1987); and more recently, the much more sophisticated TSNLP (Test Suite for Natural Language Processing) which includes some 9500 test items for English, French and German, each of them being annotated with syntactic and application related information (Oepen and Flickinger, 1998).

Yet because they do not take into account the semantic dimension, none of these tools are adequate for evaluating the paraphrastic power of a grammar. To remedy this, we propose to develop a paraphrase test suite based on the paraphrase typology described in the previous section. In such a testsuite, test items pair a semantic representation with a set of paraphrases verbalising this semantics. The construction and annotation of the paraphrases reflects the paraphrase typology. In a fi rst phase, we concentrate on simple, non-recursive predicate/argument structure. Given such a structure, the construction and annotation of a test item proceeds as follows.

First, a "canonical verbalisation" is produced in which the predicate is realised by the "canonical verb" for the given concept[3] and the arguments by the canonical nouns.

Next variants are produced by systematically trying to create parallel, shuffling and defi nitional paraphrases. Each of the variant is furthermore annotated with labels caracterising the type of paraphrasing involved. Here is an example. Suppose the input semantics is:

apply(e), agent(e,jean), theme(e,job), failure(e)

for which the canonical verbalisation is:

(6) Jean a candidaté sans succès sur le poste
Jean has applied in vain for the job.

[3]Like in a thesaurus, we assume that amongst a set of synonyms, one lexical unit is "canonical" and the others not. The canonical unit is sometimes called a *descriptor*.

The parallel synonyms[4] that can be used are the following:[5]

candidater	candidature	+pred-N
	poser sa candidature	+pred-vsupV
	briguer	+pred-V
sans succès	échouer	+mod-V
	être sans succès	+mod-beAdv
	ne pas être retenu	+mod-Vanton

For shuffling synonymy, two alternations are available: the active/passive alternation for "poser" and the active/locative one for "échouer". There is no converse construction. Neither is there any definition given by derivational morphology for any of the terms occurring in the canonical verbalisation. Based on these facts, the following variants and annotations can be constructed.

(7) a. Jean a brigué le poste sans succès
Jean has asked for the job in vain.
+pred-Vsyn

b. Jean a posé sa candidature sur le poste sans succès
Jean has submitted his application for the job in vain.
+pred-vsupN

c. La candidature posée par Jean sur le poste a été sans succès
The application submitted by Jean for the job was in vain.
+pred-partAdj, +mod-beAdv

d. La candidature posée par Jean sur le poste a échoué
The application submitted by Jean for the job failed.
+pred-partAdj, +mod-V

e. La candidature de Jean sur le poste a été sans succès
Jean's application for the job was in vain.
+pred-N, +mod-beAdv

f. La candidature de Jean sur le poste n'a pas été retenue
Jean's application for the job was not successful.
+pred-N, +mod-Vanton

g. La candidature de Jean sur le poste a échoué
Jean's application for the job failed.
+pred-N, +mod-V

h. Jean a échoué dans sa candidature sur le poste.
Jean failed in his application for the job.
+pred-N, +mod-V-altLoc

Thus the typology of paraphrastic means help guide the construction of the various paraphrases contained in a single item. There remains the question of how to choose the particular items of the testsuite. In other words: which semantic representations should we use to populate the test suite and on the basis of which criteria? The basic aim here is to cover the various types of possible semantic combinations and the constraints they are subject to at the syntactic (realisation) level. If, as Beth Levin argues, syntax is a reflex of semantic properties, then different semantic contents should be subject to varying syntactic constraints and the test suite ought to cover these various types of interactions. Accordingly test items are constructed whose main predicate vary along the following dimensions :

(1) WordNet Verb Family; (2) Aspect; (3) Arité

That is, items are constructed for each wordNet family (the french WordNet counts roughly 170 such families). Within a given family, we attempt to find examples with distinct aspectual categories (state, accomplishment and process). Finally, given a WN family and an aspectual category, items will vary with respect to the arity of the main predicate and the types of their arguments e.g., predicates of arity one (run, cost, sleep), of arity two with non propositional arguments (eat, hit, dug), of arity two with a propositional argument (say, promise etc.), etc.

4 A paraphrastic grammar

"Semantic grammars" already exist which describe not only the syntax but also the semantics of natural language. Thus for instance, (Copestake and Flickinger, 2000; Copestake et al., 2001) describes a Head Driven Phrase Structure Grammar (HPSG) which supports the parallel construction of a phrase structure (or derived) tree and of a semantic representation and (Dalrymple, 1999) show how to equip

[4] As has been abundantly argued by linguists, real synonyms are extremely rare. By synonyms, we in fact refer here to the notion of quasi-synonyms used for instance in WordNet that is, words that are interchangeable in a restricted set of contexts.

[5] The labels are the ones used for annotation. They caracterise variations with respect to the canonical realisation. For instance, +pref-N indicates that the main predicate (realised by a verb in the canonical verbalisation) is realised as a noun.

Lexical Functional grammar (LFG) with a glue semantics.

These grammars are both effi cient and large scale in that they cover an important fragment of the natural language they describe and can be processed by parsers and generators in almost real time. For instance, the LFG grammar parses sentences from the Wall Street Journal and the ERG HPSG grammar will produce semantic representations for about 83 per cent of the utterances in a corpus of some 10 000 utterances varying in length between one and thirty words. Parsing times vary between a few ms for short sentences and several tens of seconds for longer ones.

Nonetheless, from a semantics viewpoint, these grammars fail to yield a clear account of the paraphrastic relation. Here is a simple example illustrating this shortcoming. Suppose we parse the following paraphrases where a lexical defi nition (*driver* ≡ *person who drives*) is involved:

(8) a. The person who drives the car is mad.
b. The driver of the car is mad.

When given these sentences, the LKB system based on the ERG HPSG grammar returns semantic representations which can be sketched as follows[6]:

(9) a. the(x, person(x) ∧ the(y, car(y) ∧ drive(e,x,y) ∧ mad(x)))
a. the(y, car(y) ∧ the(x, driver(x,y) ∧ of(x,y)) ∧ mad(x))

In other words, the grammar associates with these paraphrases semantic representations which are very different. It could be argued of course that although these representations are syntactically distinct, they can be inferred, given the appropriate knowledge, to be semantically equivalent. But a solution that avoids placing such extra burden on the inferencing component is obviously better. In short, one important shortcoming of existing large scale semantic grammars is that they do not assign semantically equivalent sentences, the same semantic representation.

By contrast, we propose to develop a grammar which whereever possible assigns identical semantic representations to paraphrases and whose development is based both on semantic and syntactic considerations.

[6]These semantic representations have been simplified for better readibility. The real representations output by the LKB are the following:
prpstn(def(x,person(x)∧prpstn(def(y,car(y),
drive(e1,v1,x,y,v2),v3)), mad(e2,x,v4),v5)
prpstn(def(x,person(x)∧prpstn(def(y,car(y),
drive(e1,v1,x,y,v2),v3)), mad(e2,x,v4),v5)
prpstn(def(y,car(y)∧prpstn(def(x, driver(x,y) ∧ of(e1,x,y,v1),
mad(e2,x,v2,v3)))))

4.1 Linguistic framework

Our grammar is couched within the Feature-Based Tree Adjoining grammar (FTAG) formalism. An FTAG consists of a set of (auxiliary or initial) elementary trees and two tree composition operations: substitution and adjunction. Substitution is the standard tree operation used in phrase structure grammars while adjunction is an operation which inserts an auxiliary tree into a derived tree. To account for the effect of these insertions, two feature structures (called **top** and **bottom**) are associated with each tree node in FTAG. The top feature structure encodes information that needs to be percolated up the tree should an adjunction take place. In contrast, the bottom feature structure encodes information that remains local to the node at which adjunction takes place.

The language chosen for semantic representation is a flat semantics along the line of (Bos, 1995; Copestake et al., 1999; Copestake et al., 2001). However because we are here focusing on paraphrases rather than fi ne grained semantic distinctions, the underspecifi cation and the description of the scope relations permitted by these semantics will here be largely ignored and flat semantics will be principally used as a convenient way of describing predicate/arguments and modifi ers/modifi ed relationships. Thus the semantic representations we assume are simply set of literals of the form $P^n(x_1, \ldots, x_n)$ where P^n is a predicate of arity n and x_i is either a constant or a unifi cation variable whose value will be instantiated during processing.

Semantic construction proceeds from the derived tree (Gardent and Kallmeyer, 2003) rather than – as is more common in TAG – from the derivation tree. This is done by associating each elementary tree with a semantic representation and by decorating relevant tree nodes with unifi cation variables and constants occuring in associated semantic representation. The association between tree nodes and unifi cation variables encodes the syntax/semantics interface – it specifi es which node in the tree provides the value for which variable in the fi nal semantic representation.

As trees combine during derivation, (i) variables are unifi ed – both in the tree and in the associated semantic representation – and (ii) the semantics of the derived tree is constructed from the conjunction of the semantics of the combined trees. A simple example will illustrate this.

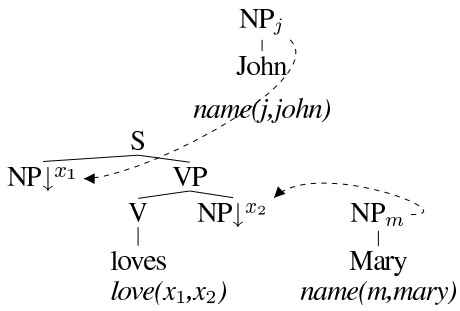

Figure 2: "John loves Mary"

Suppose the elementary trees for "John", "loves" and "Mary" are as given in Fig. 2 where a downarrow (↓) indicates a substitution node and C^x/C_x abbreviate a node with category C and a top/bottom feature structure including the feature-value pair $\{$ **index :** $x\}$. On substitution, the root node of the tree being substituted in is unifi ed with the node at which substitution takes place. Further, when derivation ends, the top and bottom feature structures of each node in the derived tree are unifi ed. Thus in this case, x_1 is unifi ed with j and x_2 with m. Hence, the resulting semantics is:

$$love(j, m), name(j, john), name(m, mary)$$

4.2 The signature of the semantic representation language

Let us now come back to the paraphrases given in example 1. To produce an identical semantic representation of these three sentences, we fi rst need to ensure that synonyms be assigned the same concept. That is, we need to fi x a concept inventory and to use this inventory in a consistent way in particular, by assigning synonyms the same concept.

For non predicative units, we use WordNet synset numbers or when working within a restricted domain with a well defi ned thesaurus, the descriptors of that thesaurus.

To represent the semantics of predicative units, we use FrameNet inventory of frames and frame elements (C.Johnson et al., 2002). FrameNet is an online lexical resource for English based on the principles of Frame Semantics. In this approach, a word evokes a *frame* i.e., a simple or a complex event, and each frame is associated with a number of *frame elements* that is, a number of participants fulfi lling a given role in the frame. Finally each frame is associated with a set of target words, the words that evoke that frame.

Thus FrameNet associates synonyms with an identical concept namely, the frame evoked by those synonyms. We make use of this feature and instead of choosing our own semantic predicates and relations, draw on FrameNet frames and frame elements. For instance, the paraphrases in example 1 are taken to evoke the FrameNet COMMERCE frame and to instantiate two of its frame elements namely, GOODS and MONEY. The semantic representation they will be assigned will therefore be the following:

commerce(e,g,m), cruise(g), goods(e,g), high(m), money(e,m)

4.3 Capturing paraphrastic relations

Given the basic signature provided by FrameNet (and any extension of it that will prove necessary to account for the data), the grammar must then specify a compositional semantics which will derive identical representations for the types of paraphrases captured by our typology. In essence, this implies assigning the same semantic representations to synonyms, converses and alternations. Concretely, this involves two different subtasks : fi rst, a modeling of the synonymic relation between syntactically divergent constructs (e.g., between a predicative noun, a support verb construction and a verb) and second, the identifi cation of the synonymic sets (which are the words and multi word expressions that stand in a parallel, shuffling or defi nitional paraphrastic relation?).

Modeling intercategorial synonymic links. A fi rst investigation of Anne Abeillé's TAG for French suggests that modeling the synonymic relations across syntactic constructs is reasonably straightforward. For instance, as Figures 3, 4 and 5 show, the FTAG trees assigned on syntactic grounds by Anne Abeillé FTAG to predicative nouns, support verb constructions and transitive verbs can be equiped with a flat semantics in such a way as to assign the three sentences in 1 a unique semantic representation namely the one given above. Generally, the problem is not so much to state the correspondances between synonymic but syntactically different constructs as to do this in a general way while not overgeneralising. To address this problem, we are currently working on developing a metagrammar in the sense of (Candito, 1999). This metagrammar allows us to factorise both syntactic and semantic information. Syntactic information is factorised in the usual way. For instance, there will be a class NOVN1 which groups together all the initial trees representing the possible syntactic confi gurations in which a transitive verb with

two nominal arguments can occur. But additionnally there will be semantic classes such as, "binary_predicate_of_semantic_type_X" which will be associated with the relevant syntactic classes for instance, NOVN1 (the class of transitive verbs with nominal arguments), BINARY_NPRED (the class of binary predicative nouns), NOVSUPNN1 , the class of support verb constructions taking two nominal arguments. By further associating semantic units (e.g., "cost") with the appropriate semantic classes (e.g., "binary_predicate_of_semantic_type_X"), we can in this way capture both intra and intercategorial paraphrasing links in a general way.

Constructing paraphrastic sets. Depending on the type of paraphrastic means involved, constructing a paraphrastic set (the set of all lexical items related by a paraphrastic link be it parallel, shuffling or defi nitional) is more or less easy as resources for that specifi c means may or may not be readily available.

Cases of intracategorial synonymy are relatively straigthtforward as several electronic synonym dictionnaries for french are available (Ploux, 1997). Multi word expressions however remain a problem as they are often not or only partially included in such dictionnaries. For these or for a specifi c domain, basic synonymic dictionaries can be complemented using learning methods based on distributional similarity (Pereira et al., 1993; Lin, 1998). techniques.

For intercategorial synonymy involving a derivational morphology link, some resources are available which however are only partial in that they only store morphological families that is, sets of items that are morphologically related. Lexical semantics information still need to be included.

Intercategorial synonymy not involving a derivational morphology link has been little studied and resources are lacking. However as for other types of synonymy, distributional analysis and clustering techniques can be used to develop such resources.

For shuffling paraphrases, french alternations are partially described in (Saint-Dizier, 1999) and a resource is available which describes alternation and the mapping verbs/alternations for roughly 1 700 verbs. For complementing this database and for converse constructions, the LADL tables (Gross, 1975) can furthermore be resorted to, which list detailed syntactico-semantic descriptions for 5 000 verbs and 25 000 verbal expressions. In particular, (Gross, 1989) lists the converses of some 3 500 predicative nouns.

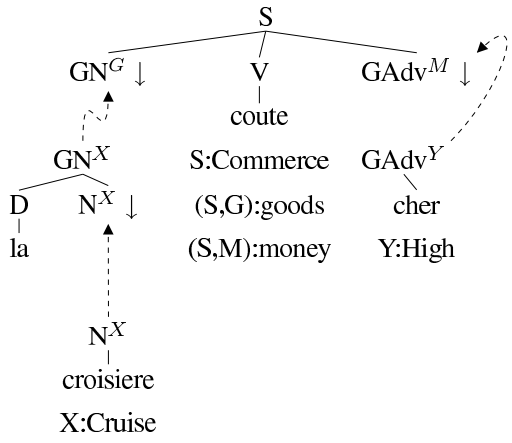

Figure 3: La croisière coûte cher

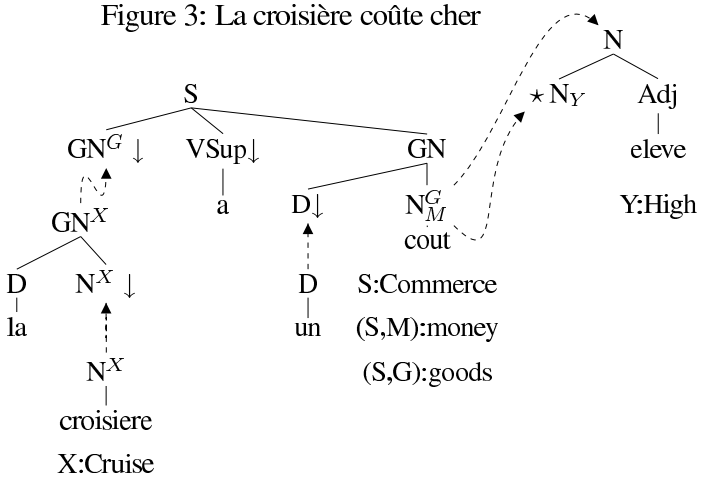

Figure 4: La croisière a un coût élevé

5 Conclusion

Besides the development and evaluation of a core paraphrastic testsuite and grammar for French, we plan to investigate two main issues. First, how precisely should a metagrammar be structured to best describe a paraphrastic grammar? And second: is it possible to extract from the kind of inference rules automatically derived in machine learning approach, information that can be used to specify this metagrammar?

6 Acknowledgments.

This paper is based upon work suppported in part by the project "Des connaissances à leurs réalisation en langue" within the CNRS TCAN program.

References

A. Abeillé. 2002. *Une Grammaire Electronique du Franais*. CNRS Editions.

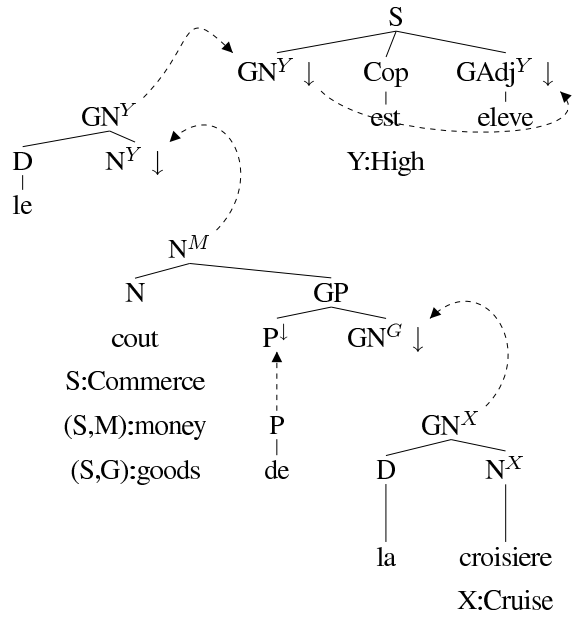

Figure 5: Le coût de la croisière est élevé

R. Barzilay and L. Lee. 2003. Learning to paraphrase: an unsupervised approahc using mutliple-sequence alignment. In *Proceedings of NAACL-HLT*.

A. Black, S. Abney, D. Flickinger, C. Gdaniec, R. Grishman, P. Harrison, D. Hindel, R. INgria, F. Jelinek, F. Klaavans, M. Liberman, M. Marcus, S. Roukos, B. Santorini, and T. Strzalkowski. 1991. A procedure for quantitatively comparing the syntactic coverage of english grammars. In *Proceedings of teh 4th DARPA Speech and Natural Language Workshop*.

J. Bos. 1995. Predicate logic unplugged. In Paul Dekker and Martin Stokhof, editors, *Proceedings of the 10th Amsterdam Colloquium*, pages 133–142.

M.H Candito. 1999. Un outil multilingue de generation de ltag : application au francais et a l'italien. *TAL*, 40(1).

C.Johnson, C. Fillmore, M. Petruckand C. Baker, M. Ellsworth, and J. Ruppenhofer. 2002. Framenet: Theory and practice. Technical report, Berkeley.

Ann Copestake and Dan Flickinger. 2000. An open source grammar development environment and broad-coverage English grammar using HPSG. In *Proceedings of the 2nd International Conference on Language Resources and Evaluation*, Athens, Greece.

A. Copestake, D. Flickinger, I. Sag, and C. Pollard. 1999. Minimal Recursion Semantics. An Introduction. Manuscript, Stanford University.

A. Copestake, A. Lascarides, and D. Flickinger. 2001. An algebra for semantic construction in constraint-based grammars. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, Toulouse, France.

M. Dalrymple. 1999. *Semantics and syntax in lexical functional grammar*. MIT Press.

D. Flickinger, J. Nerbonne, I. Sag, and T. Wasow. 1987. Towards evaluation of nlp systems. Technical report, Hewlett-Packard Laboratories.

C. Gardent and L. Kallmeyer. 2003. Semantic construction in ftag. In *Proceedings of EACL*, Budapest, Hungary.

O. Glickman and I. Dagan. 2003. Identifying lexical paraphrases from a single corpus: a case study for verbs. In *Proceedings of Recent Advances in Natural Language Processing*.

M. Gross. 1975. *Méthodes en syntase*. Masson, Paris.

G. Gross. 1989. *Les constructions converses du francais*. CNRS Editions.

Dekang Lin and Patrick Pantel. 2001. Discovery of inference rules for question answering. *Natural Language Engineering*.

D. Lin. 1998. Automatic retrieval and clustering of similar words. In *Proceedings of ACL/COLING*, pages 768–774.

I. Mel'čuk. 1988. Paraphrase et lexique dans la thorie linguistique sens-texte. *Lexique*, 6:13–54.

S. Oepen and D. Flickinger. 1998. Towards systematic grammar profi ling. test suite technology 10 years after. *Computer Speech and Language*, 12:411–435.

F. Pereira, N. Tishby, and L. Lee. 1993. Distributional clustering of english words. In *Proceedings of the ACL*, pages 183–190.

S. Ploux. 1997. Modlisation et traitement informatique de la synonymi. *Linguisticae Investigationes*, XXI(1).

P. Saint-Dizier, 1999. *Alternations and Verb Semantic Classes for French: analysis and class formation*, chapter 5. Kluwer.

Y. Shinyanma, S. Sekine, K. Sudo, and R. Grishman. 2002. Automatic paraphrase acquisition from news articles. In *Proceedings of HLT*.