

# COMP IN (PARALLEL) GRAMMAR WRITING

Martin Forst

Universität Stuttgart  
Institut für Maschinelle Sprachverarbeitung

**Proceedings of the LFG06 Conference**

Universität Konstanz

Miriam Butt and Tracy Holloway King (Editors)

2006

CSLI Publications

<http://csli-publications.stanford.edu/>

## Abstract

This paper is a grammar writer's reaction to the 'COMP debate', which has been going on in the LFG community for more than a decade now. Taking as a starting point the work by Dalrymple and Lødrup (2000), Alsina et al. (2005) and Berman (2006), I first consider the question with respect to a German large-coverage LFG. I show that, in addition to the reasons put forth by Alsina et al. (2005) and Berman (2006), there are further reasons to reinterpret as  $OBL_{\theta S}$  (or  $OBJ_{\theta S}$ ) the arguments that Dalrymple and Lødrup (2000) analyze as COMPs in German, a language which they consider as 'mixed'. These have to do with COMPs subcategorized for by nouns and, to a lesser extent, with past participles of OBJ experiencer psych-verbs. I then present some data from Spanish, a 'non-mixed' language, and show that the distinction introduced in the reinterpretation of COMPs of German nouns has a correlate in Spanish. Furthermore, I point out how the reinterpretation of COMP can increase parallelism between grammars, an argument that needs to be taken with caution, of course, but in my opinion, does have its place in parallel grammar development. The final section explains why the linguistically more adequate analysis without COMP is also more attractive from the point of view of grammar engineering or, in other words, why the enhanced descriptive elegance of a grammar leads to improved efficiency in its processing. I report an 11% gain in processing time with a revised grammar compared to an equivalent grammar that makes use of COMP.

## 1 Introduction

The status of the grammatical function COMP (and – to a lesser extent – XCOMP) has been the subject of a considerable amount of work in theoretical LFG. However, the implementational efforts for various languages realized in the *ParGram* initiative do not reflect any of the results of this work so far, probably because grammar developers avoid the major effort of adapting their grammars as long as the controversy does not converge towards a consensus. This paper is an attempt to contribute to a possible consensus and to show that implemented grammars do benefit from insights from theoretical work, as a better understanding of the generalizations at work in the languages considered allows for improved lexicon acquisition and more general, and hence more efficient, grammars.

Before considering the linguistic data themselves and their modelling in the implemented grammars, let us recall the major steps of the 'COMP debate': In a contribution to the LFG List, Alsina et al. (1996) suggest reinterpreting COMPs as OBJs, arguing that the difference in category at the c-structure level between a nominal OBJ and an argument clause should not be reflected by a difference in grammatical function at the f-structure level if there were no further reasons to differentiate OBJs and COMPs. Dalrymple and Lødrup (2000) take up this argument and show that it holds for some argument clauses, but not for all. They thus propose to reinterpret some COMPs as OBJs, but keep COMP in the inventory of grammatical functions, even if, according to their terminology, COMP only exists in 'mixed' languages, whereas it does not in 'non-mixed' languages. Alsina et al. (2005), finally, revise their initial proposal of reinterpreting all COMPs as OBJs and suggest instead to reinterpret COMPs as OBJs,  $OBJ_{\theta S}$  or  $OBL_{\theta S}$ , depending on the subcategorizing element. One central argument of theirs is the alternation of non-OBJ argument clauses with different clitics in Catalan; another one is the parallelism between Catalan ('mixed') and Spanish ('non-mixed') translational equivalences. Interestingly, Berman (2006) comes to a similar conclusion, although she bases her argumentation on German ('mixed') facts only.

The remainder of this paper is organized as follows: In section 2, I show that, in addition to the reasons put forth by Alsina et al. (2005) and Berman (2006), there are further reasons to reinterpret as  $OBL_{\theta S}$  (or  $OBJ_{\theta S}$ ) the arguments that Dalrymple and Lødrup (2000) analyze as COMPs in German, a language which they consider as 'mixed'. These have to do with COMPs subcategorized for by nouns and, to a lesser extent, with past participles of OBJ experiencer psych-verbs. In section 3, I then present some data from Spanish, a 'non-mixed' language, and show that the distinction introduced in the reinterpretation of COMPs of German nouns has a correlate in Spanish. Furthermore, I point out how the reinterpretation

of COMP can increase parallelism between grammars, an argument that needs to be taken with caution, of course, but in my opinion, does have its place in parallel grammar development. Finally, section 4 explains why the linguistically more adequate analysis without COMP is also more attractive from the point of view of grammar engineering or, in other words, why the enhanced descriptive elegance of the respective grammars leads to improved efficiency in their processing. This claim is sustained by the result of a small experiment with two grammar versions, one with and one without COMP.

Finally, it should be noted that my arguments with respect to the reinterpretation of COMP also apply to the arguments called VCOMPs in our grammar. These are infinitival arguments that are anaphorically controlled, i.e. arguments of equi verbs. I do not advocate, however, the reinterpretation of XCOMPs, which, in the German *ParGram* LFG, are functionally controlled arguments of modal, raising and AcI (*accusativus cum infinitivo*) verbs. Their behaviour is clearly different from the behaviour of VCOMPs with respect to passivization, the alternation with DPs<sup>1</sup> and control, so that I prefer maintaining XCOMP as a grammatical function, as long as no linguistically and technically adequate alternative is available.

## 2 The status of COMP in German and English

In our subcategorization lexicons for verbs and adjectives, we observe that almost all COMPs alternate with either OBJs or OBL<sub>θ</sub>s. (COMPs that alternate with OBJ<sub>θ</sub>s seem to be rare.) This redundancy seems undesirable to me, both for conceptual and for practical reasons; I will thus propose a reinterpretation of some COMPs as OBJs, as suggested by Dalrymple and Lødrup (2000), and then reinterpret the remaining COMPs as OBL<sub>θ</sub>s (and potentially OBJ<sub>θ</sub>s), along the lines of Alsina et al. (2005) and Berman (2006).

### 2.1 Uncontroversial OBJ clauses of verbs

In the theoretical literature, there seems to be a consensus now that certain COMPs should be reinterpreted as OBJs. The main criteria for distinguishing OBJ clauses from non-OBJ clauses are their alternation with DPs, their ability of being fronted and of being promoted to SUBJ status in passivized sentences. I will briefly go through these criteria again, although they have been discussed in the literature mentioned, because most *ParGram* grammars do not yet distinguish OBJ clauses from non-OBJ clauses and thus make wrong predictions with respect to the behaviour of either the OBJ clauses or the non-OBJ clauses.

#### 2.1.1 Alternation with DPs

OBJ clauses subcategorized for by verbs alternate with DPs, as can be seen in (1) and (2). Non-OBJ clauses do not (see (7) and (8)). In the German *ParGram* LFG, as in most *ParGram* grammars, this alternation is stipulated through the presence of two unrelated subcategorization frames in the entry of the verbs concerned.

(1) *I believe [that the earth is round] / it / that.*

(2) *Ich glaube [, dass die Erde rund ist] / es / das.*  
 I believe that the earth round is / it / that.  
 ‘I believe that the earth is round / it / that.’

---

<sup>1</sup>The distinction between DPs and NPs is without importance for our argumentation. We use the term *DP* throughout this paper because there is a category DP in the German *ParGram* LFG. For grammars that do not have such a category or for readers that have reservations towards the notion of DP, the adequate term would be *NP*.

### 2.1.2 Fronting

OBJ clauses subcategorized for by verbs can be fronted, as in (3) and (4), whereas non-OBJ clauses cannot.

(3) *[That the earth is round] / That I believe.*

(4) *[Dass die Erde rund ist,] / Das wurde nicht geglaubt.*  
That the earth round is / That was not believed.  
'That the earth is round / That was not believed.'

### 2.1.3 Passivization

OBJ clauses subcategorized for by verbs can be promoted to SUBJ status in passivized sentences, as can be seen in (5) and (6). Non-OBJ clauses do not participate in passivization in the same way.

(5) *[That the earth is round] / That was not generally accepted.*

(6) *[Dass die Erde rund ist,] / Das glaube wurde nicht allgemein akzeptiert.*  
That the earth round was / That was not generally accepted.  
'That the earth is round / That was not generally accepted.'

## 2.2 Potential OBL<sub>θ</sub> clauses of verbs

Argument clauses that are neither SUBJ nor OBJ are OBJ<sub>θ</sub> or OBL<sub>θ</sub> according to Alsina et al. (2005). In German (and English), OBJ<sub>θ</sub> clauses seem to be rare. For the sake of simplicity, I thus talk about OBL<sub>θ</sub> clauses here, although OBJ<sub>θ</sub> clauses are expected to behave similarly.

### 2.2.1 Alternation with PPs, not DPs

OBL<sub>θ</sub> clauses subcategorized for by verbs do not alternate with DPs, but with PPs, as can be seen in (7) and (8). In most *ParGram* grammars, this alternation is stipulated through the presence of two unrelated subcategorization frames in the entry of the verbs concerned.

(7) *The secretary has already insisted [that I have to fill in the form] / \*it / [on it].*

(8) *Die Sekretärin passt auf [, dass ich das Formular ausfülle].*  
The secretary pays attention that I the form fill in.  
'The secretary is attentive that I fill in the form.'

### 2.2.2 Fronting

In English, OBL<sub>θ</sub> clauses can only be fronted with a stranded preposition appearing after the verb. In German, they can only be fronted together with the corresponding pronominal adverb. Both the stranded preposition and the pronominal adverb indicate the type of OBL<sub>θ</sub> function the fronted argument clause has; without this indication, the OBL<sub>θ</sub> clauses, unlike their OBJ counterparts, cannot be fronted.

(9) *[That I have to fill in the form] the secretary has already insisted \*(on).*

(10) *\*(Darauf,) [dass ich das Formular ausfülle,] passt die Sekretärin auf.*  
On that that I the form fill in pays the secretary attention.  
'The secretary is attentive that I fill in the form.'

As the German *ParGram* LFG, as it is, does not distinguish OBJ clauses from OBL<sub>θ</sub> clauses, it wrongly parses (9). The non-distinction of OBJ clauses and OBL<sub>θ</sub> clauses thus causes overgeneration in this case.

### 2.2.3 Passivization

In English, passivization is only possible with a stranded preposition appearing after the verb, and in German, the argument clause must be preceded by the corresponding pronominal adverb. For English, the explanation is that not only *Objs* are promoted; in the German example, the argument clause is clearly not the SUBJ of the sentence (since PPs never are SUBJs), so that the construction has to be analyzed as an impersonal passive.

(11) [*That I have to fill in the form*] *has already been insisted* \*(*on*).

(12) \*(*Darauf,*) [*dass ich das Formular ausfülle,*] *wird aufgepasst.*

On that that I the form fill in is paid attention.

‘They are / Someone is attentive that I fill in the form. (impersonal passive)’

Again, the German *ParGram* LFG overgenerates due to the non-distinction of OBJ clauses and OBL<sub>θ</sub> clauses, by wrongly parsing the unacceptable version of (12).

## 2.3 OBJ clauses of adjectives

Although adjectives are often believed not to take OBJs, a small number of German adjectives, like *gewohnt* (‘used to’) and *wert* (‘worth’), do.

### 2.3.1 Alternation with DPs

Interestingly, the OBJ clauses and infinitives subcategorized for by adjectives alternate with DPs, just like OBJ clauses and infinitives subcategorized for by verbs. But again, just like in the lexical entries of verbs, this alternation is stipulated by two seemingly unrelated subcategorization frames.

(13) a. *Wir sind bei diesen Themen ja gewohnt* [, *dass die Damen unter sich sind*].  
We are with these topics indeed used that the ladies among themselves are.

‘With respect to these topics, we are indeed used to the fact that the ladies stick to themselves.’<sup>2</sup>

b. *Wir sind es / das bei diesen Themen ja gewohnt.*

We are it / that with these topics indeed used.

‘With respect to these topics, we are indeed used to it / that.’

(14) a. *Die Begründung ist ?(es) wert* [, *im Wortlaut wiedergegeben zu werden*]:  
The justification is it worth in the wording reproduced to be:

‘The justification is worth being reproduced in its exact wording.’

b. *Die Begründung ist es / das wert.*

The justification is it / that worth.

‘The justification is worth it / that.’

---

<sup>2</sup>This example, as most of the following examples, is an edited version of a corpus sentence. The corpora consulted were the TIGER Corpus, the Huge German Corpus (HGC) and the Europarl Corpus.

### 2.3.2 Fronting

Just like OBJ clauses and infinitives subcategorized for by verbs, OBJ clauses and infinitives subcategorized for by adjectives can be fronted. The German *ParGram* LFG, however, does not provide the necessary functional uncertainty path in the annotation of the fronted clausal or infinitival constituent, so that it cannot parse (15) and (16). Here, the grammar thus undergenerates.

- (15) *[Dass die Damen unter sich sind,] sind wir bei diesen Themen ja gewohnt.*  
That the ladies among themselves are are we with these topics indeed used.

‘With respect to these topics, we are indeed used to the fact that the ladies stick to themselves.’

- (16) *[Im Wortlaut wiedergegeben zu werden,] ist die Begründung nicht wert.*  
In the wording reproduced to be is the justification not worth.

‘The justification is not worth being reproduced in its exact wording.’

## 2.4 OBL<sub>θ</sub> clauses of adjectives

In my view, most clausal and infinitival arguments subcategorized for by adjectives are OBL<sub>θ</sub>s. This is confirmed by the criteria that I have applied to OBL<sub>θ</sub> clauses subcategorized for by verbs above.

### 2.4.1 Alternation with PPs, not DPs

OBL<sub>θ</sub> clauses subcategorized for by adjectives alternate with PPs, not with DPs.

- (17) *Ich bin froh, dass es alle geschafft haben.*  
I am glad that it all made have.

‘I am glad that they all made it.’

- (18) *Ich bin \*es / \*das / darüber froh.*  
I am \*it / \*that / about that glad.

‘I am glad about that.’

### 2.4.2 Fronting

OBL<sub>θ</sub> clauses subcategorized for by adjectives cannot be fronted without the corresponding pronominal adverb, whereas OBJ clauses subcategorized for by adjectives can, as we have seen above.

- (19) *\*(Darüber,) [dass es alle geschafft haben,] bin ich froh.*  
About that that it all made have am I glad.

‘I am glad that they all made it.’

## 2.5 Why can OBL<sub>θ</sub> clauses not be fronted (or, at least, only exceptionally)?

We have seen above that OBL<sub>θ</sub> clauses cannot be fronted with the remainder of the sentence staying unchanged. Dalrymple and Lødrup (2000) take this observation as an argument for the existence of a distinct grammatical function COMP, which cannot be fronted in German and English. (They note, however, that in earlier stages of the German language, non-OBJ argument clauses could be topicalized.)

Berman (2006), who is in favour of the reinterpretation of COMP, gives a relatively complicated explanation for the fact that, in modern German, non-OBJ argument clauses cannot appear in SpecCP and she makes claims with respect to the ability of non-OBJ argument clauses to appear in topicalized partial VPs that, to me, seem to complicate the picture artificially.

What I believe is active in English and modern German is a constraint on the linear order of the subcat-frame-evoking element and the OBL<sub>θ</sub> (or OBJ<sub>θ</sub>) clause, which states that a (morphologically unmarked) *that/dass* clause can only function as an OBL<sub>θ</sub> (or OBJ<sub>θ</sub>) if it appears to the right of the verb, adjective or noun that subcategorizes for it. OBL<sub>θ</sub> PPs can be fronted without problems because this constraint simply does not apply to them. With respect to argument clauses, this constraint explains the relevant data,<sup>3</sup> and I think this is a plausible constraint, since OBL<sub>θ</sub> is a more marked grammatical function than SUBJ and OBJ, and morphologically unmarked constituents such as clauses can only be interpreted as such if the subcat-frame-evoking element prepares the hearer to do so.

In older German, this constraint apparently was weaker than today, but even in modern corpora we can find examples where, like in (20), a fronted *dass* clause or a fronted infinitival VP functions as an OBL<sub>θ</sub>.

- (20) *Sie zu ächten und zu verabscheuen gibt es gute Gründe;*  
 Them to ostracise and to loathe gives it good reasons;  
 ‘There are good reasons to ostracise and loathe them;’

## 2.6 COMPS subcategorized for by nouns

As for nouns, none of those from our subcategorization lexicon that subcategorize for a COMP can alternatively subcategorize for an OBJ, which is not surprising, as nouns are known to be intransitive. However, a large proportion of these nouns can alternatively subcategorize for an OBL<sub>θ</sub>. I will show that the COMPS subcategorized for by these nouns can safely be reinterpreted as OBL<sub>θ</sub>s, in the very same way as many COMPS subcategorized for by verbs and adjectives, and the same restrictions on unbounded dependencies apply for all clausal OBL<sub>θ</sub>s, as example (21) illustrates.

- (21) a. *Es gibt keinen Zweifel (daran), dass hier eine höhere Summe stehen sollte.*  
 It gives no doubt at this that here a higher sum stand should.
- b. *\*(Daran,) dass hier eine höhere Summe stehen sollte, gibt es keinen Zweifel.*  
 At this that here a higher sum stand should gives it no doubt.  
 ‘There is no doubt that there should be a higher sum here.’

But what about the COMPS that cannot be reinterpreted as OBL<sub>θ</sub>s, like the one in (22)?

- (22) a. *Es gibt den Vorwurf (\*dafür/dazu...), dass sich die DDR-Journalisten*  
 It gives the reproach that themselves the GDR journalists  
*moralisch diskreditiert hätten.*  
 morally discredited had.  
 ‘There is the reproach that the GDR journalists had discredited themselves morally.’
- b. *\*(Dafür/Dazu...) Dass sich die DDR-Journalisten moralisch diskreditiert*  
 That themselves the GDR journalists morally discredited  
*hätten, gibt es den Vorwurf.*  
 had gives it the reproach.

<sup>3</sup>Example (25) in Berman (2006) is not relevant in my view, since the *dass* clause there is a SUBJ, and SUBJs are known to appear in topicalized partial VPs only with a very small number of verbs.

I propose to treat these as a kind of apposition or adjunct rather than an argument, a solution already hinted at in Dalrymple and Lødrup (2000). This treatment is motivated by semantic considerations, but also by the fact that none of these COMPs is obligatory, whereas at least some of the  $OBL_{\theta}$  clauses subcategorized for by nouns are, and that the restrictions on unbounded dependencies that apply to appositive clauses are more strict than the ones that apply to  $OBL_{\theta}$  clauses.

Interestingly, the nouns that can take clausal appositions are the very same ones that can subcategorize for a clausal SUBJ when used predicatively. This is illustrated in (23), which contains the same *dass* clause and the same noun, namely *Vorwurf* ('reproach'), as (22).

- (23) *Dass sich die DDR-Journalisten moralisch diskreditiert hätten, ist ein schwerer Vorwurf.*  
 That themselves the GDR journalists morally discredited had is a serious reproach.  
 'That the GDR journalists had discredited themselves morally is a serious reproach.'

Nouns that subcategorize for  $OBL_{\theta}$  clauses do not show this behaviour, as (24) illustrates.

- (24) \**Dass hier eine höhere Summe stehen sollte, ist ihr Zweifel.*  
 That here a higher sum stand should is her doubt.

Finally, the distinction between  $OBL_{\theta}$  clauses subcategorized for by nouns and appositive clauses which accompany non-predicatively used nouns and which correspond to SUBJ clauses when the noun is used predicatively also allows us to analyze examples like (25) properly. Here, the noun *Beweis*, which is predicatively used, subcategorizes for a clausal SUBJ, which is instantiated by the first *dass* clause, and for a clausal  $OBL_{\theta}$ , which is the latter *dass* clause.

- (25) *Dass inzwischen neun Prozent als politisch Verfolgte anerkannt werden, ist für Kanther Beweis, dass das neue Recht Schutz garantiert, ...*  
 That now nine percent as politically persecuted recognized are is for Kanther proof that the new legislation protection guarantees, ...  
 'That nine percent are now recognized as political refugees proves, for Kanther, that the new legislation guarantees protection, ...'

At the moment, it is not at all recorded in our subcategorization lexicon which ones are the nouns that can take clausal SUBJs. However, thanks to the knowledge about the relationship between appositive clauses of non-predicatively used nouns and clausal SUBJs of predicatively used nouns, it should be easy to acquire this knowledge by revisiting all lexical entries of nouns that subcategorize for COMPs at the moment.

## 2.7 Participles of OBJ experiencer psych-verbs

Further evidence for the ability of CPs to function as  $OBL_{\theta}$ s comes from the subcategorization behaviour of the participles of OBJ experiencer psych-verbs (e.g., *beruhigt* 'reassured', *beunruhigt* 'worried', *geervt* 'annoyed', *schockiert* 'shocked', *überrascht* 'surprised'). These participles are special because they seem to subcategorize for a COMP although the corresponding active forms clearly do not. As a temporary solution in order to analyze sentences like (26), where such a participles occurs, we entered them as 'lexicalized' participles in our lexicon. However, apart from their subcategorization behaviour, nothing indicates that they are lexicalized in any way.



- (26) *Ich bin schockiert [, dass sich Bernard so positioniert hat.]*  
 I am shocked that himself Bernard so positioned has.  
 ‘I am shocked that Bernard positioned himself this way.’

By reinterpreting certain COMPs as OBL<sub>θ</sub>S and, hence, potentially as OBL-AGs, I will be able to account for the subcategorization behaviour of these participles with the standard lexical rule for passive.

### 3 COMP cross-linguistically

Unlike the other core grammatical functions, which seem to be present in all languages, COMP seems to be used only by the so-called ‘mixed’ languages (Dalrymple and Lødrup 2000). To me, this assumption seems somehow surprising and, moreover, it forces us to assume non-parallel analyses for translational equivalents that only differ in the presence or absence of a preposition. (See, e.g., Alsina et al. (2005) for translational equivalents from Spanish and Catalan or the examples below for translational equivalents from Spanish and French.) Since there seems to be consensus as to the non-use of COMP in ‘non-mixed’ languages, the question that needs to be clarified before COMP is abandoned as a grammatical function is whether the COMPs in ‘mixed’ languages can reasonably be reinterpreted as something else. In section 2, I have argued that they can in German and English; in the following, I will show that they can in French, yet another ‘mixed’ language, that French (just like Catalan) provides another argument for doing so, and that parallelism between closely related languages that differ with respect to their ‘mixedness’ is greatly improved.

#### 3.1 OBJ clauses in French (a ‘mixed’ language) and Spanish (a ‘non-mixed’ language)

Here, I will briefly show that the distinction between OBJ clauses and non-OBJ clauses makes sense in French and Spanish and that the criteria for the distinction used in German and English can be applied in these two languages as well. French and Spanish (just like Catalan and Spanish in Alsina et al. (2005)) are an interesting language pair because they are closely related, both historically and typologically, but, according to Dalrymple and Lødrup (2000), French is a ‘mixed’ language, whereas Spanish is a ‘non-mixed’ language.

##### 3.1.1 Alternation with direct object clitic

Both in French and in Spanish, OBJ clauses alternate with OBJ clitics.

- (27) a. *Les gens ne croyaient pas que la terre était ronde.*  
 The people NE believed not that the earth was round.  
 ‘People did not believe that the earth was round.’  
 b. *Les gens ne le croyaient pas.*  
 The people NE it believed not.  
 ‘People did not believe it.’
- (28) a. *La gente no creía que la tierra era redonda.*  
 The people not believed that the earth was round.  
 ‘People did not believe that the earth was round.’  
 b. *La gente no lo creía.*  
 The people not it believed.  
 ‘People did not believed it.’

### 3.1.2 Fronting

When fronted, OBJ clauses cooccur with a resumptive OBJ clitic in both French and Spanish.

(29) *Que la terre était ronde, les gens ne le croyaient pas.*  
That the earth was round the people NE it believed not.  
'That the earth was round, people did not believe.'

(30) *Que la tierra era redonda, la gente no lo creía.*  
That the earth was round the people not it believed.  
'That the earth was round, people did not believe.'

### 3.1.3 Passivization

In both French and Spanish, OBJ clauses can be promoted to SUBJ status in passivized sentences.

(31) *Que la terre était ronde n' était pas généralement accepté.*  
That the earth is round NE was not generally accepted.  
'That the earth is round was not generally accepted.'

(32) *Que la tierra era redonda no era generalmente aceptado.*  
That the earth was round not was generally accepted.  
'That the earth was round was not generally accepted.'

## 3.2 OBL<sub>θ</sub> clauses in French (a 'mixed' language) and Spanish (a 'non-mixed' language)

In Spanish, *que* clauses can be preceded by prepositions that indicate their status as OBL<sub>θ</sub>s. In French, *que* clauses cannot be directly preceded by prepositions. Just like in Catalan (Alsina et al. 2005), there are good reasons, however, to suppose that many *que* clauses are OBL<sub>θ</sub>s.

### 3.2.1 Alternation with *both* adverbial clitics (French) or PPs (Spanish) respectively

The most important reason is that French non-OBJ clauses alternate with the two adverbial clitics available in the language, depending on the type of OBL<sub>θ</sub> the verb (or adjective or noun) subcategorizes for. If these non-OBJ clauses were COMPs, as proposed in Dalrymple and Lødrup (2000), we could not explain why the argument clause of *insister* in (33) alternates with the clitic *y*, whereas the argument clause of *réjouir* in (35) alternates with the clitic *en*. In Spanish, the non-OBJ clauses all alternate with PPs.

(33) a. *La secrétaire a déjà insisté que je dois remplir le formulaire.*  
The secretary has already insisted that I must fill in the form.  
'The secretary has already insisted that I have to fill in the form.'

b. *La secrétaire y a déjà insisté.*  
The secretary Y has already insisted.  
'The secretary has already insisted on it.'

(34) a. *La secretaria ya ha insistido en que tengo que llenar el formulario.*  
The secretary already has insisted in that I have to fill in the form.  
'The secretary has already insisted that I have to fill in the form.'

- b. *La secretaria ya ha insistido en eso.*  
The secretary already has insisted in that.  
'The secretary has already insisted on that.'
- (35) a. *Je me réjouis beaucoup que mes parents viennent pour Noël.*  
I myself am glad much that my parents come for Christmas.  
'I am very glad that my parents are coming for Christmas.'
- b. *Je m' en réjouis beaucoup.*  
I myself EN am glad much.  
'I am very glad about that.'
- (36) a. *Me alegro mucho de que mis padres vengan para Navidad.*  
Myself am glad much about that my parents come for Christmas.  
'I am very glad that my parents are coming for Christmas.'
- b. *Me alegro mucho de eso.*  
Myself am glad much about that.  
'I am very glad about that.'

### 3.2.2 Fronting

In French, non-OBJ clauses can be fronted, but must then cooccur with the corresponding adverbial clitic, which is *y* in (37) and *en* in (39). In Spanish, non-OBJ clauses can only be fronted together with the preposition that precedes them.

- (37) *Que je dois remplir le formulaire, la secrétaire y a déjà insisté.*  
That I must fill in the form the secretary Y has already insisted.  
'That I have to fill in the form, the secretary has already insisted on.'
- (38) *En que tengo que llenar el formulario la secretaria ya ha insistido.*  
In that I have to fill in the form the secretary already has insisted.  
'That I have to fill in the form, the secretary has already insisted on.'
- (39) *Que mes parents viennent pour Noël, je m' en réjouis beaucoup.*  
That my parents come for Christmas I myself EN am glad much.  
'That my parents are coming for Christmas I am very glad about.'
- (40) *De que mis padres vengan para Navidad me alegro mucho.*  
About that my parents come for Christmas myself am glad much.  
'That my parents are coming for Christmas I am very glad about.'

### 3.2.3 Passivization

Non-OBJ clauses cannot be promoted to SUBJ status in either French or Spanish. I just give a French example here because only in the 'mixed' language French is there a danger of overgeneration due to the non-distinction of OBJ and non-OBJ clauses.

- (41) *\*Que je dois remplir le formulaire a déjà été insisté.*  
That I must fill in the form has already been insisted.  
'That I have to fill in the form has already been insisted on.'

### 3.3 COMPS subcategorized for by nouns

Let us now consider COMPs that seem to be subcategorized for by nouns in a crosslinguistic perspective. I have argued above that *dass* clauses like the one in (42) are OBL<sub>θ</sub>s, whereas clauses like the one in (44) are appositions. I will argue that the same holds true for the *que* clauses in (43) and (45) respectively. My main arguments are that, in Spanish, OBL<sub>θ</sub> *que* clauses can be preceded by basically any preposition that can introduce OBL<sub>θ</sub>s, whereas clausal appositions are always introduced by the preposition *de*, and that basically the same restrictions as to unbounded dependencies apply to *que* clauses as to *dass* clauses. (See subsection 2.6.)

- (42) ... [*DP das Vertrauen, dass es auch in Zukunft ein Land Bosnien-Herzegowina*  
...the confidence that it also in future a country Bosnia-Herzegowina  
*gibt*] ...  
gives ...  
'... confidence that the country of Bosnia-Herzegowina will continue to exist in the future ...'
- (43) ... [*DP la confianza en que en el futuro exista también un país como B-H*]  
...the confidence in that in the future exist also a country like B-H  
...  
...  
'... confidence that the country of Bosnia-Herzegowina will continue to exist in the future ...'
- (44) [*DP Die Tatsache, dass diese Misshandlung durch andere Muslime ausgeführt wurde,*]  
The fact that this mistreatment by other Muslims carried out was  
...  
...  
'The fact that this abuse was perpetrated by other Muslims ...'
- (45) [*DP El hecho de que los malos tratos fueran infligidos por otros musulmanes*]  
The fact of that the bad treatments were inflicted by other Muslims  
...  
...  
'The fact that this abuse was perpetrated by other Muslims ...'

In this context, it is interesting to note that in the French translation of the two sentences above, which are from the Europarl Corpus, we find a construction consisting of the preposition *en*, the pronoun *ce* and the *que* clause in the case of the OBL<sub>θ</sub> clause, whereas the appositive *que* clause directly follows the noun *fait*, on which it depends. This does not mean that all OBL<sub>θ</sub> clauses subcategorized for by nouns are preceded by a preposition and the pronoun *ce* in French, but only OBL<sub>θ</sub> clauses can be constructed this way. Appositive clauses always directly follow their governing noun in French.

- (46) ... *la confiance en ce qu' à l' avenir, la Bosnie-Herzégovine demeure aussi*  
...the confidence in it that to the future the Bosnia-Herzegowina stays also  
*un pays* ...  
a country ...  
'... confidence that the country of Bosnia-Herzegowina will continue to exist in the future ...'

- (47) *Le fait que ces actes de violence aient été perpétrés par d'autres musulmans ...*  
 The fact that these acts of violence have been perpetrated by Article.Indefinite.PI other Muslims ...  
 'The fact that this abuse was perpetrated by other Muslims ...'

A further interesting observation is that the generalization stating that nouns that can head an appositive clause when used non-predicatively are the ones that can take a clausal SUBJ when used predicatively carries over to French and Spanish.

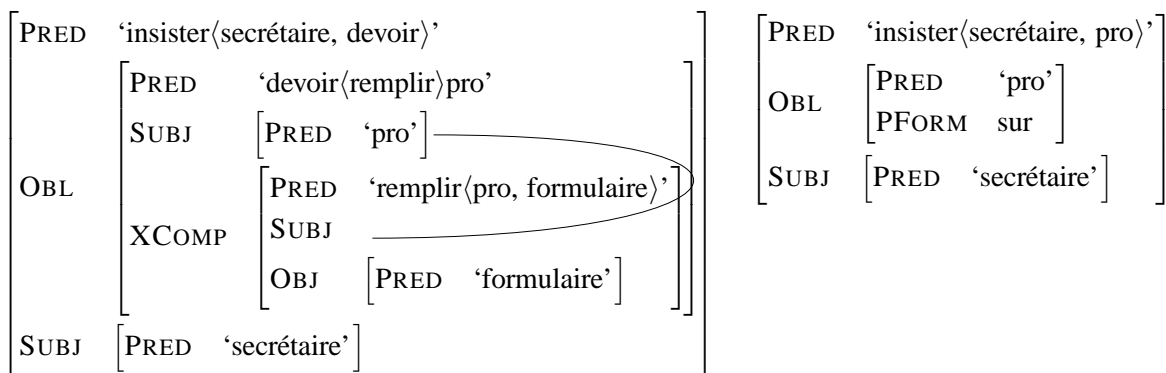
- (48) *[Que ces actes de violence aient été perpétrés par d'autres musulmans] est un fait.*  
 That these acts of violence have been perpetrated by Article.Indefinite.PI other Muslims is a fact.  
 'That this abuse was perpetrated by other Muslims is a fact.'
- (49) *[Que los malos tratos fueron infligidos por otros musulmanes] es un hecho.*  
 That the bad treatments were inflicted by other Muslims is a fact.  
 'That this abuse was perpetrated by other Muslims is a fact.'

### 3.4 Parallelism

The *ParGram* grammars are regularly checked for parallelism among them, parallelism referring mainly to f-structures as the level of representation that is used for applications that build on top of the parser output. Whenever translational equivalents in two *ParGram* languages are structurally similar, the f-structures associated with these translational equivalents are supposed to differ only in the values of the PRED features and perhaps minor morphosyntactic features.

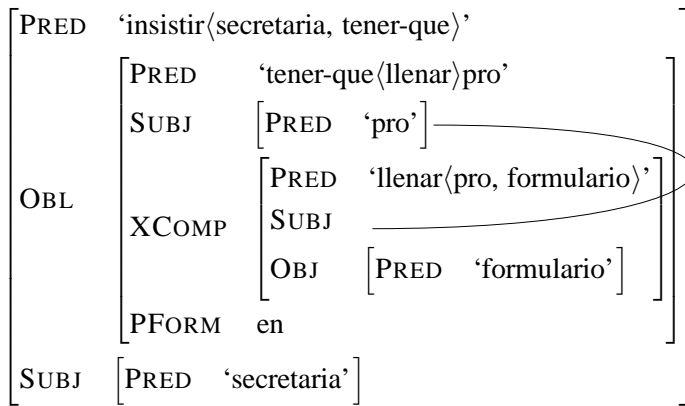
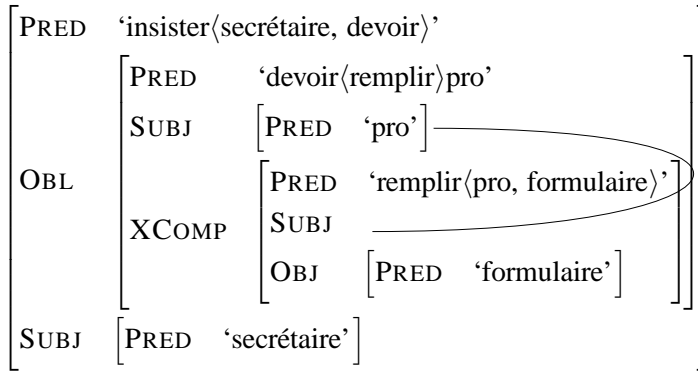
#### 3.4.1 Parallelism within a ('mixed') language

Although parallelism is generally viewed as a criterion for analyses across languages, it can also be applied as a criterion for analyses of related sentences within a language. In 'mixed' languages, the criterion of parallelism is interesting with respect to the alternation of argument clauses with DPs or PPs. The following two f-structures, associated with (33a) and (33b) after the reinterpretation of COMP as OBJ, OBJ<sub>θ</sub> or OBL<sub>θ</sub>, are parallel with respect to the grammatical functions subcategorized for by *insister*, whereas the f-structures currently produced by the French *ParGram* LFG are not.



### 3.4.2 Parallelism across languages

Parallelism across languages, in particular between ‘mixed’ and ‘non-mixed’ languages, also greatly benefits from the reinterpretation of COMP. The two following f-structures, associated to (33a) and (34a), which are translational equivalents in French and Spanish, are parallel. If COMP were maintained as a grammatical function in ‘mixed’ languages, they would diverge.



## 4 Engineering advantages

### 4.1 Simplification of subcategorization lexicons

In section 2, I mentioned the huge redundancy that exists in our subcategorization lexicons for verbs and adjectives. I believe that this redundancy is harmful in several ways, not only conceptually but also in terms of grammar efficiency. In addition to the grammatical functions a verb or an adjective can take, our subcategorization lexicons encode what categories can realize a given function. For example, thematic SUBJs can maximally be realized as DPs, *dass* CPs, declarative verb-second CPs, interrogative CPs or infinitival VPs. Although this is not yet done in practice, underspecification could be used in cases where all five categories are possible as the SUBJ of a lexical element. This possibility is not available, however, for non-SUBJ functions if DPs and PPs are analyzed as OBJs (or OBJ<sub>θ</sub>S) and OBL<sub>θ</sub>S respectively and CPs and VPs are analyzed as COMPs and VCOMPs respectively. The reinterpretation of COMPs and VCOMPs as OBJs and OBL<sub>θ</sub>S would allow the use of underspecification with respect to category for all grammatical functions and, hence, open up the way for a great simplification of our subcategorization lexicons. Apart from the conceptual advantage this represents, in my opinion, it is reasonable to expect a substantive gain in efficiency from this simplification, since it considerably reduces the number of disjuncts in the lexical entries of verbs and adjectives that have to be tested by the parser which processes the grammar.

The two following examples illustrate this point: *akzeptieren* ('to accept') has the following lexical entry in the original verb subcategorization lexicon of our grammar.

```
akzeptieren !V-S xle
  {@(DPnom-DPacc %stem) @(AUX-HABEN)
  |@(DPnom-Sdass_corr %stem) @(AUX-HABEN)
  |@(DPnom-Sv2_corr %stem) @(AUX-HABEN)
  |@(DPnom-Swh_corr %stem) @(AUX-HABEN)
  |@(DPnom-VPzuinf_corr %stem) @(AUX-HABEN)
  }; ETC.
```

The templates ending in `_corr` allow for a clausal or infinitival argument both with and without the correlative pronoun *es*. Since the functional interpretation of the clausal or infinitival argument changes, depending on the presence or absence of the correlative element, each of these templates involves a two-way disjunction, so that there are actually nine disjuncts in the lexical entry.

This number could be reduced to three, if we made maximal usage of underspecification in the lexical entry. This means that we would not specify the possible categorial realizations of a grammatical function if all categorial realizations permitted by the grammar are possible. We would then have something like the following:

```
akzeptieren !V-S xle
  {@(SUBJ_DPnom-OBJ %stem) @(AUX-HABEN)
  |@(SUBJ_DPnom-COMP %stem) @(AUX-HABEN)
  |@(SUBJ_DPnom-VCOMP %stem) @(AUX-HABEN)
  }; ETC.
```

A further reduction is not possible because each disjunct evokes a functionally distinct subcategorization frame. If, however, COMP and VCOMP are reinterpreted as OBJ in the case of *akzeptieren*, we could further simplify the lexical entry as follows:

```
akzeptieren !V-S xle @(SUBJ_DPnom-OBJ %stem) @(AUX-HABEN); ETC.
```

My second example is a verb whose COMP, in my view, is actually an  $OBL_{\theta}$ , namely *drohen*. Its lexical entry in the original verb subcategorization lexicon looks as follows:

```
drohen !V-S xle
  {@(DPnom-PP %stem mit dat) @(AUX-HABEN)
  |@(DPnom-PPSdass %stem mit dat) @(AUX-HABEN)
  |@(DPnom-PPSv2 %stem mit dat) @(AUX-HABEN)
  |@(DPnom-PPVPzuinf %stem mit dat) @(AUX-HABEN)
  |@(DPnom-Sdass %stem) @(AUX-HABEN)
  |@(DPnom-Sv2 %stem) @(AUX-HABEN)
  |@(DPnom-VPzuinf %stem) @(AUX-HABEN)
  |...
  }; ETC.
```

These seven disjuncts can be reduced to three if maximal usage of underspecification is made.

```

drohen !V-S xle
  {@(DPnom-OBL_noInt %stem mit dat) @(AUX-HABEN)
  |@(DPnom-COMP_noInt %stem) @(AUX-HABEN)
  |@(DPnom-VCOMP %stem) @(AUX-HABEN)
  |...
  }; ETC.

```

But again, further simplification is made impossible by the distinction of  $OBL_{\theta}$ , COMP and VCOMP. Only by reinterpreting the COMP and the VCOMP of *drohen* as  $OBL_{\theta}$ s can we further simplify this lexical entry.

```

drohen !V-S xle
  {@(DPnom-OBL_noInt %stem mit dat) @(AUX-HABEN)
  |...
  }; ETC.

```

## 4.2 Simplified and more regular functional uncertainty paths

In the German *ParGram* LFG as it is, i.e. with COMPS and VCOMPS, there are functional uncertainty paths in the annotation of both topicalized and extraposed CPs and VPs that lead to both over- and undergeneration, as explained in section 2. Moreover, the functional uncertainty path in the annotation of extraposed CPs and VPs involves a high number of disjuncts due to the fact that extraposed CPs and VPs that are not preceded by a correlative pronoun or pronominal adverb are analyzed as SUBJS, COMPS or VCOMPS respectively, whereas those that are preceded by a correlative element are analyzed as APP-CLAUSES of SUBJS, OBJs or  $OBL_{\theta}$ s. With COMPS and VCOMPS being reinterpreted as OBJs or  $OBL_{\theta}$ s, all extraposed CPs and VPs would be analyzed as (APP-CLAUSES<sup>4</sup> of) SUBJS, OBJs or  $OBL_{\theta}$ s. The revised functional uncertainty path in the f-annotation of extraposed CPs and VPs then involves fewer disjuncts and exhibits more regularity than the original functional uncertainty path, as is illustrated here.

```

... "Nachfeld"
CPdep[std]: { (^ SUBJ (APP-CLAUSE)) = !
              | (^ VP-PATH { COMP | { OBJ | OBL } APP-CLAUSE } = !
              | (^ DP-PATH COMP) = !
              | ...
              }

... "Nachfeld"
CPdep[std]: { (^ { SUBJ | VP-PATH { OBJ | OBL } } (APP-CLAUSE)) = !
              | (^ DP-PATH { OBL (APP-CLAUSE) | APP } ) = !
              | ...
              }

```

## 4.3 Simplified and more regular application of the lexical rule(s) for passive

In the original grammar, there are three templates that implement lexical rules for passive: PASSIVE-OBJ-TO-SUBJ, PASSIVE-COMP-TO-SUBJ and PASSIVE-VCOMP-TO-SUBJ. The first one can only promote nominal objects to subjects and applies to all subcategorization frames that

---

<sup>4</sup>Although I believe that the function APP-CLAUSE should be removed in order to simplify the functional uncertainty path under consideration even further, I think that this issue should be kept separate from the status of COMP and VCOMP.



involve a thematic SUBJ and a thematic OBJ; the second one can promote COMPs to SUBJ status, but, for reasons that have no independent motivation in the grammar, applies only to subcategorization frames that involve a COMP, but no OBJ, and the same applies to the last template with respect to VCOMPs.

Once COMPs and VCOMPs are reinterpreted as OBJs, OBJ<sub>θ</sub>s or OBL<sub>θ</sub>s, it is sufficient to keep the template PASSIVE-OBJ-TO-SUBJ, which allows for the promotion to SUBJ status of any type of OBJ and is systematically applied to all subcategorization frames that involve a thematic SUBJ and a thematic OBJ, which can be clausal or infinitival in this case. No longer are there lexical rules that apply to subcategorization frames in an unsystematic way.

#### 4.4 Improved acquisition of subcategorization information from corpora

In the context of COMPs of nouns, I have stated above that a distinction is to be made between clauses that are actually OBL<sub>θ</sub>s of nouns and clauses that function as appositions to nouns and, more importantly, that this distinction was related to the subcategorization behaviour of nouns when they are used predicatively. Two properties in the subcategorization behaviour of those nouns which, at first glance, seem to be unrelated thus turn out to be one and the same property in fact.

I believe that there are more properties of this kind, which are recorded as separate pieces of information in our subcategorization lexicons, but are in fact related very regularly. Many of them have nothing or little to do with the grammatical function COMP, but the COMP does contribute to blur the picture that we have of subcategorization and on whose basis we develop the theory that underlies the way we record subcategorization behaviour. To name just two examples, the possibility of a correlative *es* to cooccur with an OBJ clause is independent of the exact nature (*dass*, verb-second declarative, interrogative) of this clause, and all verbs that can subcategorize for an OBL<sub>θ</sub> clause without a correlate can equally subcategorize for such a clause with some correlative pronominal adverb. As long as we make use of COMP as a grammatical function, we are highly unlikely to discover this kind of regularity because the constituents are analyzed as having different grammatical functions (COMP vs. OBJ in the case of OBJ clauses (not) preceded by a correlative *es*; COMP vs. OBL<sub>θ</sub> in the case of OBL<sub>θ</sub> clauses (not) preceded by a correlative pronominal adverb).

For the acquisition of a subcategorization lexicon from corpora that aims at completeness and consistency, it is of utmost importance to have a good understanding of all regularities that are at work in subcategorization. No corpus will contain all realizational variants of a given subcategorization frame, but if the theory on which we build the representation in which the subcategorization information is recorded captures regularities, there is hope that, via these generalizations, the acquired subcategorization information also covers most unseen realizational variants.

#### 4.5 Grammar efficiency

In order to verify my claim that the reorganization of the subcategorization lexicons made possible by the reinterpretation of COMP has a positive effect on grammar efficiency, I created two largely equivalent grammar versions and had them analyze 1,956 sentences from section 8,001 through 10,000 of the TIGER Corpus. The versions mainly differ in the verb subcategorization lexicon used. Further, rather minor, changes were made necessary by the reinterpretation of COMP in the new subcategorization lexicon, such as changes in the f-annotation of CPs and VPs and in treatment of the correlative pronoun *es* and correlative pronominal adverbs.

The comparison of the two runs shows that the original grammar version needs 11% more time to parse the 1,956 sentences than the version with the revised subcategorization lexicon. While this is not an enormous gain in efficiency, it does represent an improvement, which, moreover, reduces the number

of timeouts (sentences that cannot be associated with a full parse within a bounded amount of time, set to 100 seconds in both runs) by 13 to 181 out of the 1,956 sentences.

## 5 Conclusions

COMP seems to be redundant as a grammatical function, both for reasons internal to ‘mixed’ languages like German (or Catalan, English, French etc.) and for reasons of parallelism between closely related languages that, in spite of their close relationship, differ as to their alleged ‘mixed’ or ‘non-mixed’ status, as it is the case, e.g., for Catalan and Spanish and for French and Spanish. Furthermore, categorically restricted functions like COMP and VCOMP pose problems for the efficient and technically economic organization of subcategorization lexicons that, at least in principle, treat the functional status of arguments and their possible realizations in terms of syntactic category as disjunct pieces of information.

## References

- Alsina, Àlex, KP Mohanan, and Tara Mohanan. 1996. Untitled submission to the LFG List.
- Alsina, Àlex, KP Mohanan, and Tara Mohanan. 2005. How to get rid of the COMP. In *Proceedings of the 10th International LFG Conference (LFG’05)*, Bergen, Norway. CSLI Publications.
- Berman, Judith. 2006. Functional identification of complement clauses in German and the specification of COMP. In Jane Grimshaw, Joan Maling, Chris Manning, Jane Simpson, and Annie Zaenen (eds.), *Architectures, Rules, and Preferences: A Festschrift for Joan Bresnan*. Stanford, California: CSLI Publications.
- Bodomo, Adams B., and Y. M. Lee. 2003. On the function COMP in Cantonese. In A. B. Bodomo and K. K. Luke (eds.), *Lexical-Functional Grammar Analysis of Chinese*. Journal of Chinese Linguistics Monograph 19.
- Dalrymple, Mary, and Helge Lødrup. 2000. The Grammatical Functions of Complement Clauses. In *Proceedings of the 5th International LFG Conference (LFG’00)*, Berkeley, California. CSLI Publications.
- Lødrup, Helge. 2002. Infinitival complements in Norwegian and the form-function relation. In *Proceedings of the 7th International LFG Conference (LFG’02)*, Athens, Greece. CSLI Publications.
- Lødrup, Helge. 2004. Clausal complementation in Norwegian. *Nordic Journal of Linguistics* 27(1): 61–95.