How Problematic are Clitics for S-TAG Translations?

Mark Dras and Tonia Bleam

Institute for Research in Cognitive Science, University of Pennsylvania
Suite 400A, 3401 Walnut St, Philadelphia PA 19104 USA
{madras,tbleam}@linc.cis.upenn.edu

Abstract

We show that clitics are not as problematic for Synchronous TAG as has been supposed, and give two solutions; and, in doing so, demonstrate that 'unbounded relations', such as it is argued clitics induce between dependency trees, are only an artefact of particular analyses.

1. Introduction

In this paper we investigate Synchronous TAG as defined in Shieber (1994) (hereafter just S-TAG). This formalism has attractive characteristics such as the weak language preservation property (WLPP), whereby the power of the component TAGs is not altered by their synchronisation. A canonical example of the (potential) limitations of S-TAG is translation between languages with pronominal clitics and those without: because of unbounded clitic dependencies, the argument goes, radically different derivation structures are produced for each language, in violation of the isomorphism required by S-TAG. We illustrate the problem using inalienable possession constructions in Spanish, and then present one possible solution using a metagrammar, as in Dras (1999a).

However, this is not the only possible solution; and in examining a variant analysis, this paper demonstrates that the problematic 'unbounded relations' between trees that Shieber mentions are not an innate characteristic of constructions, but rather are artefacts of the analysis. Further, it suggests that the two solutions for the behaviour of clitics presented here reflect a common concept of 'grouping' in grammars.

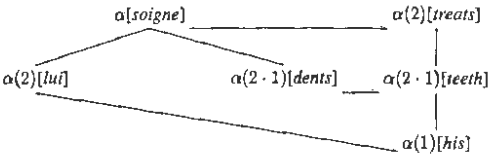

Figure 1: Shieber partial derivation tree pair

2. An Initial Analysis

Shieber (1994) sketches an analysis of clitics (based on a suggestion by Abeillé) giving it as a potential problem for S-TAG, which requires an isomorphism between derivation trees. In this section we discuss Shieber's analysis and show that his class of examples does not, in fact, require non-isomorphic derivation trees. However, such non-isomorphic constructions do exist in other languages and are thus problematic. We go on to argue that the unboundedness in these structures can be handled through the relaxation of the isomorphism requirement via a metagrammar (Dras, 1999a).

2.1. *Shieber's Analysis*

Shieber's example is in (1), with the clitic *lui* indicating possession of the body part by the patient. A partial derivation tree pair for (1) is given in Figure 1, reproduced from Figure 10 of Shieber (1994).

(1) a. Le docteur lui soigne les dents.

b. The doctor treats his teeth.

The trees are clearly not isomorphic. If they represent a fixed relation—i.e. each node is always immediately dominated by its parent, with no possibility of intervening nodes—this could be handled by Shieber's suggestion of 'bounded subderivation', where the fixed relations are treated as single nodes. However,

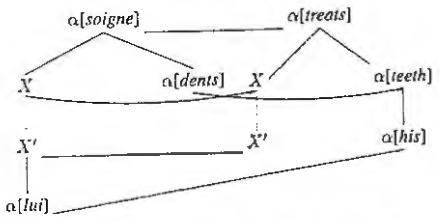

Figure 2: Unbounded relation, variant 1

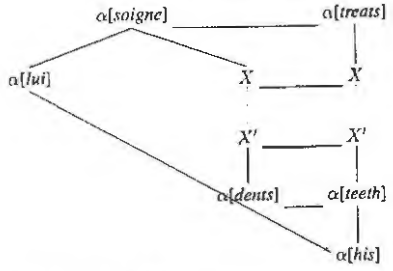

Figure 3: Unbounded relation, variant 2

Shieber also suggests that the "relation between the clitic and the NP which it is semantically related to seems to be potentially unbounded". In terms of tree relations, this suggests that there is unbounded material intervening in the trees between where $\alpha[lui]$ and $\alpha[his]$ attach, hence no possible isomorphism. Given the tree configuration of Figure 1, there are two possible cases where the relation between the trees is unbounded. The first is in Figure 2: the X and X' connected by vertical dots indicate the unbounded material. The derivation represented by Figure 2 is exemplified in (2). In this example, there is an unbounded number of verbs which can be adjoined into $\alpha[soigne]$; $\alpha[lui]$ is adjoined into the lowermost of these nodes (X'). However, expressions such as (2) are unattested in French, since the clitic must occur immediately before *soigner* (and auxiliaries).

(2) a. * Le docteur lui veut pouvoir ... soigner les dents.

b. The doctor wants to be able ... to treat his teeth.

The second possible case is illustrated by Figure 3. This derivation is exemplified by (3), which has an unbounded number of NPs between clitic and body part.

(3) a. * Le docteur lui regarde une copie d'une photo ... des dents.

b. The doctor is looking at a copy of a photo ... of his teeth.

These examples are also ungrammatical in French. Thus, neither possibility for establishing an unbounded relation applies, and hence, contra Shieber's footnote (and accepted folklore) they do not appear to be problematic for isomorphic S-TAG, although they do raise other problems (Abeillé, 1994).

2.2. *A Spanish Example*

Spanish, however, does allow clitic climbing over a potentially unbounded number of 'trigger' verbs (Aissen & Perlmutter 1976). The example in (4) parallels the French example in (2), with clitic *le*, but is acceptable.

(4) El médico le quería poder ... examinar los dientes.

In analysing clitic behaviour in (4), either syntax-dependent or syntax-independent analyses are possible. In a syntax-dependent analysis, there would be a coindexing (in the derived tree) between the clitic and its corresponding NP. In a syntax-independent analysis, the relationship would be handled by some other mechanism which remains to be specified. Our reconstruction of Shieber's analysis is syntax-independent, with $\alpha[lui]$ a single tree.

2.3. *A Metagrammar*

We propose to handle the unboundedness shown in (4), with its derivation tree pair in Figure 4, using a metagrammar (Dras, 1999a). A metagrammar specifies a relation between derivation trees by means of a TAG grammar of derivation structures. A minimal metagrammar for (4) is shown in Figure 5.

The pair $\mathfrak{A}$ does the essential grouping of the clitic and slot for a recursively-addable verb (the X to X' material), mapping to the English substructure. The unbounded intervening material is given by tree pair $\mathfrak{B}$, and clearly there is an isomorphism at the level of the derivation of the derivation (the 'meta-derivation'). This metagrammar is in Rogers' (1994) regular form (it is not possible to adjoin into the

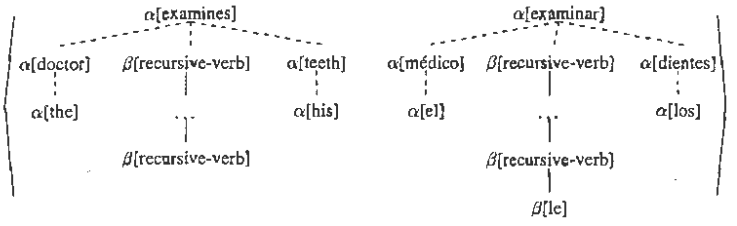

Figure 4: Derivation tree pair

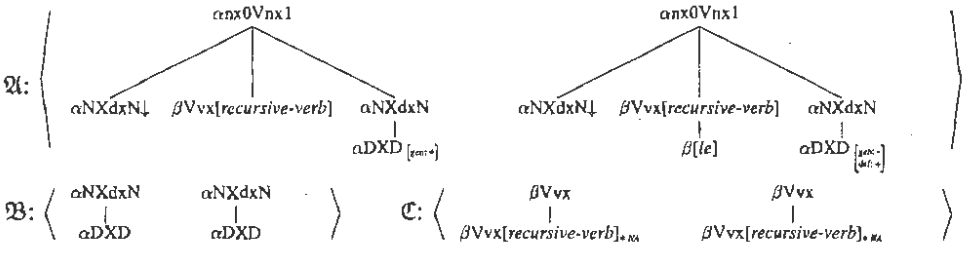

Figure 5: A metagrammar for Figure 4

spine of an auxiliary tree in this metagrammar) and so the results from Dras (1999a) apply: the WLPP holds, and the object-level formalisms still have TAG weak generative capacity.

Note that this analysis is compatible with the spirit of Abeillé (1994). There, the behaviour of the clitic is constrained by an S-TAG which pairs a syntactic and a semantic grammar. The S-TAG there is the earlier, non-isomorphic S-TAG of Shieber & Schabes (1990), so the precise analysis is not of use for investigating isomorphic S-TAG, and moreover its mathematical properties are not well understood. What we have done here, however, is compatible with Abeillé's syntax-semantics idea. There is a parallel between the English side of our grammar and the semantic side of Abeillé's grammar, with the metagrammar pairing the nodes in such a way that the clitic must be interpreted as an inalienable possessor.

3. An Alternative Analysis

Taking an individual Romance syntactic grammar by itself (that is, not constraining it through pairing with another grammar), the analysis above is insufficiently restrictive. For example, if there is a standard bridge verb tree adjoined, as in (5), there is nothing in this analysis preventing the clitic from climbing over

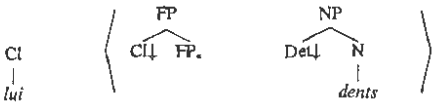

Figure 6: New clitic analysis

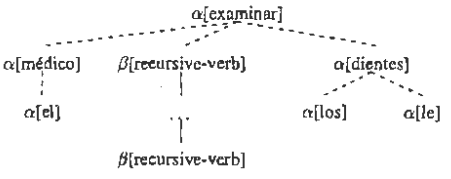

Figure 7: Reanalysed Spanish derivation tree

the bridge verb (*piensa*, thinks).

(5) * Juan le piensa que el médico examinó los dientes.

To account for Spanish clitic climbing, Bleam (1994) adopts a syntax-dependent analysis in which the coindexing between the clitic and the NP is represented by an MCTAG sequence. For us, the important aspect of this analysis is that the clitic is prevented from moving past particular constituents, such as negation and complementizers, and examples like (5) are not generated.

We analyze (4) using the tree sequence shown in Figure 6. The Spanish derivation tree is as

in Figure 7, with the English tree as before.[1,2]

In the Spanish tree, *los* and *le* are inserted into the tree sequence for *dientes*. A bounded relation between the English and Spanish trees is now induced, treating *los* and *le* as a bounded subderivation.[3]

4. Discussion

In the analysis presented in Section 3, the relation between the clitic and its associated NP is local, so we do not need to represent unbounded relations in a metagrammar. In addition, it not only rules out ungrammatical structures that our first approach does not, but also captures the intuition that the clitic is as much a part of the *dientes* structure as *his* is of *teeth*.

Both analyses discussed here draw attention to the fact that the 'unbounded' nature of constructions is not fixed. What constitutes an unbounded relation at the derivation level for a given object-level grammar becomes a bounded relation for a slightly different object-level grammar. To explain this, a notion of 'grouping' is useful here. Grouping is related to the concept of domain of locality: MCTAG group entities by associating trees together in multi-component tree sequences; a metagrammar groups elements by associating nodes in the derivation tree. So the role of grouping elements so that a relation between derivation trees is established can be traded off between the object-level grammar and a metagrammar. As an obvious rule of thumb, grouping should occur in the object-level grammar when justified by linguistic reasons, such as a preference for a syntax-dependent analysis of clitics; a metagrammar can group items that are related in some other way, such as if a syntax-independent (semantic) analysis of clitics were preferred, or in cases such as the structurally-rearranging paraphrases of Dras (1999a).

In sum, we have shown that problematic cases in S-TAG models of Romance-English translation can be resolved by using either a metagrammar or an MCTAG analysis of the clitic-body part relationship; and in doing so, we have demonstrated that unbounded relations between derivation trees in S-TAG are only an artefact of the analysis.

[1]Note that there are some changes in the Spanish derivation tree. There is a new location for the *le* node, and α[*dientes*] is a two-element sequence. In addition, the *examinar* tree is modified slightly as well, now including a functional projection (FP) node. This is necessary for two reasons: to prevent multiple adjunctions at the VP node (clitic and recursive verb); and to account for the effects discussed in Bleam (1994).

[2]Note that a synchronous relation between a TAG and an MCTAG is formally well-defined (Dras, 1999b), working in essentially the same manner as S-TAG, but pairing trees with sequences rather than with trees.

[3]Other alternative TAG-based analyses are possible here also (e.g. Abeillé, 1994; Kulick, 1998; Candito, 1999). However, we have chosen the analysis given here because, as we are examining the relation between English and Romance derivation trees, we would like to have 'minimal tree pairs', to concentrate on the one phenomenon of unboundedness; the other analyses give substantially different derivations.

References

1. Abeillé, A. 1994. Two cases of clitic-noun dependencies in French. *TAG+3*, Paris. To appear: Abeillé & Rambow, eds., *Tree Adj. Gram.: Formal, Computational, and Linguistic Aspects.*
2. Aissen, J. and D. Perlmutter. 1976. Clause Reduction in Spanish. *Proc. of the Second Annual Meeting of Berkeley Linguistics Society.*
3. Bleam, T. 1994. Clitic Climbing and The Power of Tree Adjoining Grammar. *TAG+3*, Paris. To appear: Abeillé & Rambow, eds., *Tree Adj. Gram.: Formal, Computational, and Linguistic Aspects.*
4. Candito, M-H. 1999. *Organisation modulaire et paramétrable de grammaires électroniques lexicalisées.* PhD thesis. Univ. of Paris 7, France.
5. Dras, M. 1999a. A Meta-Level Grammar: Redefining Synchronous TAG for Translation and Paraphrase. *ACL'99.*
6. Dras, M. 1999b. Synchronous Parallelisms Between Different Grammar Formalisms. *MoL6.*
7. Kulick, S. 1998. Clitic climbing in Romance: "Restructuring", causatives, and object-control verbs. *TAG+4*, 88–91.
8. Rogers, J. 1994. Capturing CFLs with Tree Adjoining Grammars. *ACL'94*, 155–162.
9. Shieber, S. & Y. Schabes. 1990. Synchronous Tree-Adjoining Grammars. *COLING-90*, 253–258.
10. Shieber, S. 1994. Restricting the Weak-Generative Capacity of Synchronous Tree-Adjoining Grammars. *Computational Intelligence*, 10(4), 371–386.