

Parsing preferences with Lexicalized Tree Adjoining Grammars : exploiting the derivation tree

Alexandra KINYON

TALANA

Universite Paris 7, case 7003,

2pl Jussieu 75005 Paris France

Alexandra.Kinyon@linguist.jussieu.fr

Abstract

Since Kimball (73) parsing preference principles such as "Right association" (RA) and "Minimal attachment" (MA) are often formulated with respect to constituent trees. We present 3 preference principles based on "derivation trees" within the framework of LTAGs. We argue they remedy some shortcomings of the former approaches and account for widely accepted heuristics (e.g. argument/modifier, idioms...).

Introduction

The inherent characteristics of LTAGs (i.e. lexicalization, adjunction, an extended domain of locality and "mildly-context sensitive" power) makes it attractive to Natural Language Processing : LTAGs are parsable in polynomial time and allow an elegant and psycholinguistically plausible representation of natural language¹. Large coverage grammars were developed for English (Xtag group (95)) and French (Abeille (91)). Unfortunately, "large" grammars yield high ambiguity rates : Doran & al. (94) report 7.46 parses / sentence on a WSJ corpus of 18730 sentences using a wide coverage English grammar. Srinivas & al. (95) formulate domain independent heuristics to rank parses. But this approach is practical, English-oriented, not explicitly linked to psycholinguistic results, and does not fully exploit "derivation"

information. In this paper, we present 3 disambiguation principles which exploit derivation trees.

1. Brief presentation of LTAGs

A LTAG consists of a finite set of elementary trees of finite depth. Each elementary tree must «anchor» one or more lexical item(s). The principal anchor is called «head», other anchors are called «co-heads». All leaves in elementary trees are either «anchor», «foot node» (noted *) or «substitution node» (noted ↓). These trees are of 2 types : auxiliary or initial². A tree has at most 1 foot-node, such a tree is an auxiliary tree. Trees that are not auxiliary are initial. Elementary trees combine with 2 operations : substitution and adjunction. Substitution is compulsory and is used essentially for arguments (subject, verb and noun complements). It consists in replacing in a tree (elementary or not) a node marked for substitution with an initial tree that has a root of same category. Adjunction is optional (although it can be forbidden or made compulsory using specific constraints) and deals essentially with determiners, modifiers, auxiliaries, modals, raising verbs (e.g. seem). It consists in inserting in a tree in place of a node X an auxiliary tree with a root of same category . The descendants of X then become the descendants of the foot node of the auxiliary tree. Contrary to context-free rewriting rules, the history of derivation must be made explicit since the same derived tree can be obtained using different derivations. This is why parsing LTAGs yields a derivation tree, from

¹ e.g. Frank (92) discusses the psycholinguistic relevance of adjunction for Children Language Acquisition, Joshi (90) discusses psycholinguistic results on crossed and serial dependencies.

² Traditionally initial trees are called α , and auxiliary trees β

which a **derived tree** (i.e. constituent tree) can be obtained. (Figure 1)³. Branches in a derivation tree are unordered.

Moreover, linguistic constraints on the well-formedness of elementary trees have been formulated :

- Predicate Argument Cooccurrence Principle : there must be a leaf node for each realized argument of the head of an elementary tree.
- Semantic consistency : No elementary tree is semantically void
- Semantic minimality : an elementary tree corresponds at most to one semantic unit

2. Former results on parsing preferences

A vast literature addresses parsing preferences. Structural approaches introduced 2 principles : RA accounts for the preferred reading of the ambiguous sentence (a) : "yesterday" attaches to "left" and not to "said" (Kimball (73)).

MA accounts for the preferred reading of (b) : "for Sue" attaches to "bought" and not to "flowers" (Frazier & Fodor (78))

(a) *Tom said that Joe left yesterday*

(b) *Tom bought the flowers for Sue*

These structural principles have been criticized though : Among other things, the interaction between these principles is unclear. This type of approach lacks provision for integration with semantics and/or pragmatics (Schubert (84)), does not clearly establish the distinction between arguments and modifiers (Ferreira & Clifton (86)) and is English-biased : evidence against RA has been found for Spanish (Cuetos & Mitchell (88)) and Dutch (Brysbaert & Mitchell (96)).

Some parsing preferences are widely accepted, though :

The idiomatic interpretation of a sentence is favored over its literal interpretation (Gibbs & Nayak (89)).

Arguments are preferred over modifiers (Abney (89), Britt & al. (92)).

Additionally, lexical factors (e.g. frequency of subcategorization for a given verb) have been shown to influence parsing preferences (Hindle & Rooth (93)).

It is striking that these three most consensual types of syntactic preferences turn out to be difficult to formalize by resorting only to "constituent trees", but easy to formalize in terms of LTAGs.

Before explaining our approach, we must underline that the examples⁴ presented later on are not necessarily counter-examples to RA and or MA, but just illustrations : our goal is not to further criticize RA and MA, but to show that problems linked to these "traditional" structural approaches do not automatically condemn all structural approaches.

3 Three preference principles based on derivation trees

For sake of brevity, we will not develop the importance of "lexical factors", but just note that LTAGs are obviously well suited to represent that type of preferences because of strong lexicalization⁵.

To account for the "idiomatic" vs "literal", and for the "argument" vs "modifier" preferences, we formulate three parsing preference principles based on the shape of derivation trees :

1. Prefer the derivation tree with the fewer number of nodes
2. Prefer to attach an α -tree low⁶
3. Prefer the derivation tree with the fewer number of β -tree nodes

Principle 1 takes precedence over principle 2 and principle 2 takes precedence over principle 3.

³ Our examples follow linguistic analyses presented in (Abeillé (91)), except that we substitute sentential complements when no extraction occurs. Thus we use no VP node and no Wh nor NP traces. But this has no incidence on the application of our preference principles.

⁴ These examples are kept simple on purpose, for sake of clarity.

⁵ Also, "lexical preferences" and "structural preferences" are not necessarily antagonistic and can both be used for practical purpose.

⁶ By low we mean "as far as possible from the root".

3.1 What these principles account for

Principle 1 accounts for the preference "idiomatic" over "literal": In LTAGs, all the set elements of an idiomatic expression are present in a single elementary tree. Figure 1 shows the 2 derivation trees obtained when parsing "Yesterday John kicked the bucket". The preferred one (i.e. idiomatic interpretation) has fewer nodes.

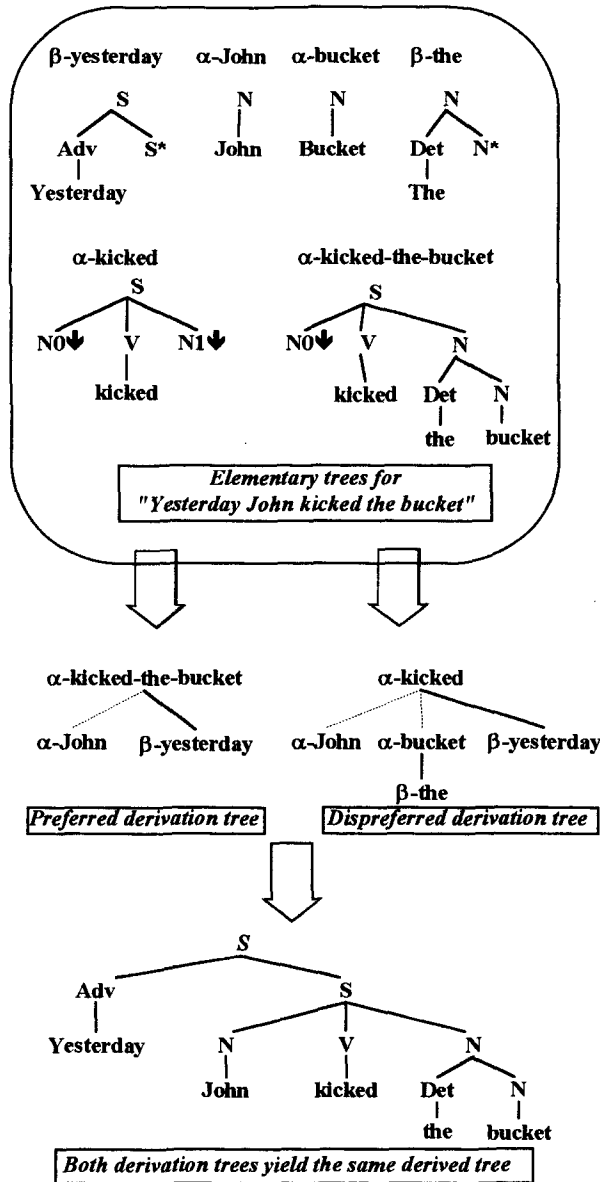


FIGURE 1⁷

Illustration of Principle 1

⁷ In derivation trees, plain lines indicate an adjunction, dotted lines a substitution.

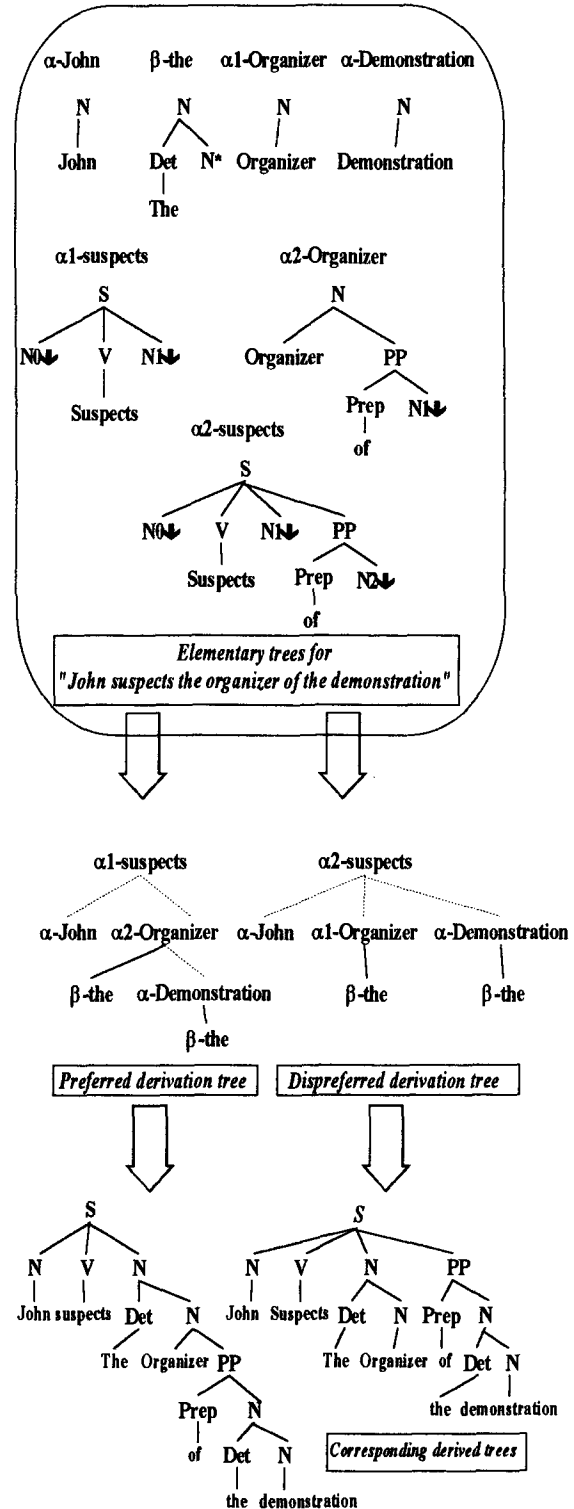


FIGURE 2

Illustration of Principle 2

for French (Abeillé & Candito (99)). We kept the 1074 grammatical ones (i.e. noted "1" in the TSNLP terminology) of category S or augmented to S (excluding coordination) that were accepted. A human picked one or more "correct" derivations for each sentence parsed⁸. Principle 1, and then Principles 1 & 2 were applied on the derivation trees to eliminate some derivations. Table 1 shows the results obtained.

	Before applying principles	After applying principle 1	After applying principles 1 & 2
Total # of sentences	1074	1074	1074
Total # of derivations	3057	2474	2334
# of sentences with at least 1 correct parse	1070 (99.6 %)	1055 (98.2 %)	1054 (98.1 %)
# of ambiguous sentences	537	427	424
# of non ambiguous sentences	537	647	650
# of partially disambiguated sentences	n.a.	89	86
# of parses / sentence	2.85	2.3	2.17

TABLE 1 : results for TSNLP

4.1 Comments on the results

After disambiguating with principles 1 and 2, the proportion of sentences with at least one parse judged correct by a human only marginally decreased while the average number of parses per

⁸ More than one derivation was deemed "correct" when non spurious ambiguity remained in modifier attachment (e.g. He saw the man with a telescope)

sentence went down from 2.85 to 2.17 (i.e. -24 %).

Since "strict modifier attachment" is orthogonal to our concern, a sentence such as (f) still yields 5 derivations, partly because of spurious ambiguity, partly because of adverbial attachment (i.e. "hier" attached to S or to V).

(f) *Il a travaillé hier (He worked yesterday)*

Therefore most sentences aren't disambiguated by principles 1 or 2, especially those anchoring an intransitive verb. For sentences that are affected by at least one of these two principles, the average number of parses per sentence goes down from 6.76 to 2.94 after applying both principles (i.e. -56.5 %). (Table 2).

	Before applying principles	After applying principle 1	After applying principles 1 & 2
# of sentences affected by at least one principle	189	189	189
# of derivations	1279	696	556
# of parses/sentence	6.77	3.68	2.94

TABLE 2 : Results for sentences affected by at least one Principle

4.2 The gap between theory and practice

Surprisingly, Principle 1 was used in only one case to prefer an idiomatic interpretation, but proved very useful in preferring arguments over modifiers : derivation trees with arguments often have fewer nodes because of co-heads. For instance it systematically favored the attachment of "by" phrases as passive with agent.

Principle 2 favored lower attachment of arguments as in (g) but proved useful only in conjunction with Principle 1 : it provided further disambiguation by selecting derivation trees among those with an equally low number of nodes.

Principle 2 says to attach an argument low (e.g. to the direct object of the main verb) rather than high (e.g. to the verb). In (c1), "of the demonstration" attaches to "organizer" rather than to "suspect", while in (c2) "of the crime" can only attach to the verb. Figure 2 shows how principle 2 yields the preferred derivation tree for sentence (c1). Similarly, in sentence (d1) "to whom" attaches to "say" rather than to "give", while in (d2) it attaches to "give" since "think" can not take a PP complement. This agrees with psycholinguistic results such as "filled gap effects" (Crain & Fodor (85)).

- (c1) John suspects the organizer of the demonstration
- (c2) John suspects Bill of the crime
- (d1) To whom does Mary say that John gives flowers.
- (d2) To whom does Mary think that John gives flowers.

Principle 3 prefers arguments over modifiers. Figure 3 shows that principle 3 predicts the preferred derivation tree for (e) : "to be honest" argument of "prefer", ruling out "to be honest" as sentence modifier (i.e. "To be honest, he prefers his daughter").

- (e) John prefers his daughter to be honest.

These three principles aim at attaching arguments as accurately as possible and do not deal with "strict" modifier attachment for the following reasons :

- There is a lack of agreement concerning the validity of preferences principles for "modifier attachment"
- Principle 3, which deals the most with modifier attachment, turned out the least conclusive when confronted to empirical data
- We wanted to evaluate how attaching arguments correctly affects ambiguity, all other factors remaining unchanged.

4 Some results

French sentences from the test suite developed in the TSNLP project (Estival & Lehman (96)) were originally parsed using Xtag with a domain independent wide-coverage grammar

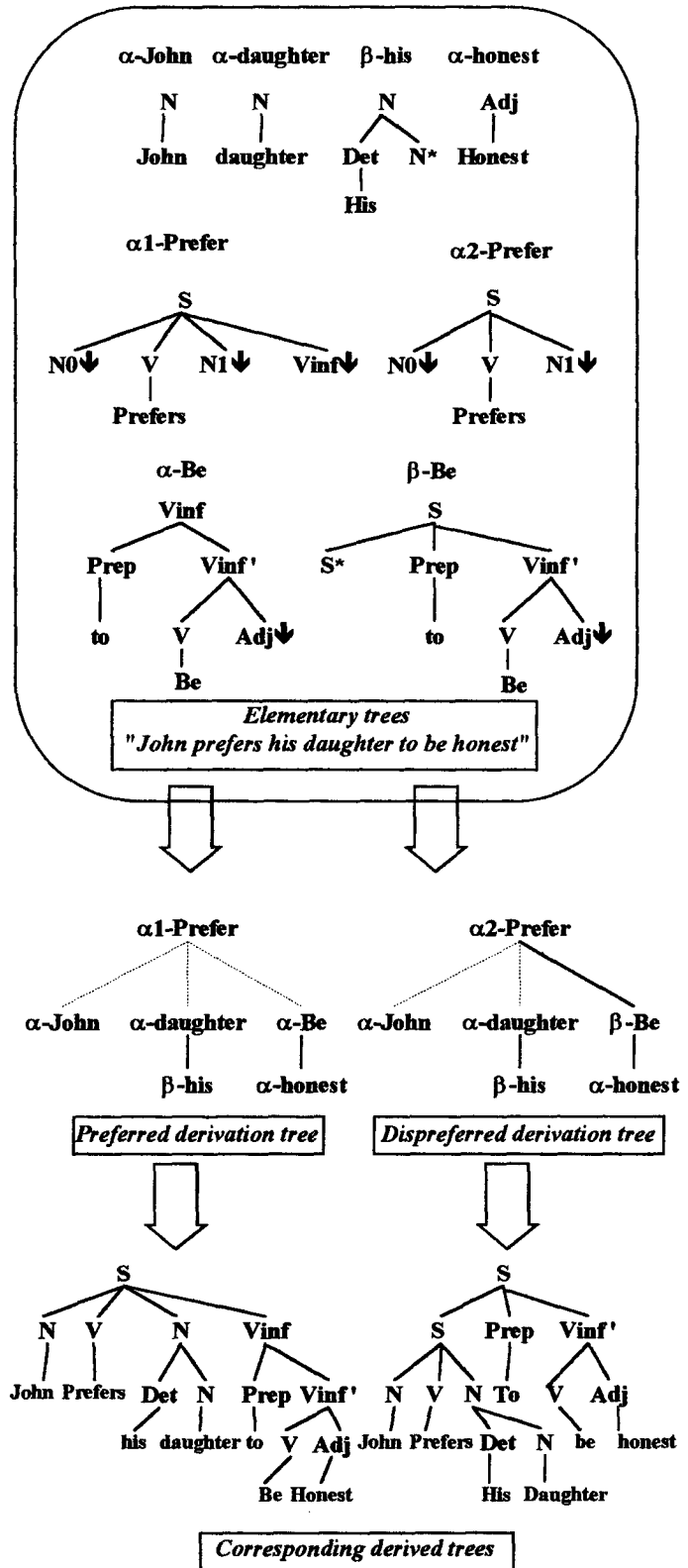


FIGURE 3
Illustration of Principle 3

(g)- *L'ingénieur obtient l'accord de l'entreprise*
(*The engineer obtains the agreement of the company / from the company*)

Principle 3 did not prove as useful as the two others : first, it aims at favoring arguments over modifiers, but these cases were already handled by Principle 1 (again because of co-heads). Second, it consistently made wrong predictions in cases of lexical ambiguity (e.g it favored "être" as a copula rather than as an auxiliary, although the auxiliary is much more common in French.). Therefore we have postponed testing it until further refinement is found.

5 Conclusion

We have presented three application-independent, domain-independent and language-independent disambiguation principles formulated in terms of derivation trees within the framework of LTAGs. But since they are straightforward to implement, these principles can be used for parse ranking applications or integrated into a parser to reduce non determinism. Preliminary results are encouraging as to the soundness of at least two of these principles. Further work will focus on testing these principles on larger corpora (e.g. Le Monde) as well as on other languages, refining them for practical purposes (e.g. addition of frequency information and principles for modifiers attachment). Since it is the first time to our knowledge that parsing preferences are formulated in terms of derivation trees, it would also be interesting to see how this could be adapted to dependency-based parsing.

References

- Abeillé A. (1991) *Une grammaire lexicalisée d'arbres adjoints pour le français*. PhD dissertation.. Université Paris 7.
- Abeillé A., Candito M.H. (1999) *FTAG : A LTAG for French*. In *Tree Adjoining Grammars*. Abeillé, Rambow(eds). CSLI, Stanford.
- Abney S. (1989) *A computational model of human parsing*. *Journal of psycholinguistic Research*, 18, 129-144.
- Britt M, Perfetti C., Garrod S, Rayner K. (1992) *Parsing and discourse : Context effects and their limits*. *Journal of memory and language*, 31, 293-314.
- Brysbaert M., Mitchell D.C. (1996) *Modifier Attachment in sentence parsing : Evidence from Dutch*. *Quarterly journal of experimental psychology*, 49a, 664-695.
- Crain S., Fodor J.D. (1985) *How can grammars help parsers?* In *Natural language parsing ..* 94-127. D. Dowty, L. Karttunen, A. Zwicky (eds). Cambridge University Press.
- Cuetos F., Mitchell D.C. (1988) *Cross linguistic differences in parsing : restrictions on the use of the Late Closure strategy in Spanish*. *Cognition*, 30,73-105.
- Doran C., Egedi D., Hockey B.A., Srinivas B., Zaidel M. (1994) *Xtag System- a wide coverage grammar for English*. COLING'94. Kyoto. Japan.
- Estival D., Lehman S (1997) *TSNLP: des jeux de phrases test pour le TALN*, TAL 38:1, 115-172
- Ferreira F. Clifton C. (1986) *The independence of syntactic processing*. *Journal of Memory and Language*, 25,348-368.
- Frank R. (1992) *Syntactic Locality and Tree Adjoining Grammar : Grammatical Acquisition and Processing Perspectives*. PhD dissertation. University of Pennsylvania.
- Frazier L, Fodor J.D. (1978) *"The sausage machine" : a new two stage parsing model*. *Cognition* 6.
- Gibbs R., Nayak (1989) *Psycholinguistic studies on the syntactic behaviour of idioms*. *Cognitive Psychology*, 21, 100-138.
- Hindle D. Rooth M. (1993) *Structural ambiguity and lexical relations*. *Computational Linguistics*, 19, pp.103-120.
- Joshi A. (1990) *Processing crossed and serial dependencies : an automaton perspective on the psycholinguistic results*. *Language and cognitive processes*, 5:1, 1-27.
- Kimball J. (1973) *Seven principles of surface structure parsing in natural language*. *Cognition* 2.
- Schubert L. (1984). *On parsing preferences*. COLING'84, Stanford. 247-250.
- Srinivas B., Doran C., Kulick S. (1995) *Heuristics and Parse Ranking*. 4th international workshop on Parsing Technologies.. Prag. Czech Republic.
- Xtag group (1995) *A LTAG for English. Technical Report IRCS 95-03*. University of Pennsylvania.