

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Are Structural Principles Useful for Automatic Disambiguation?

Permalink

<https://escholarship.org/uc/item/3jd750xj>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 22(22)

Author

Kinyon, Alexandra

Publication Date

2000

Peer reviewed

Are Structural Principles Useful for Automatic Disambiguation ?

Alexandra Kinyon (Alexandra.Kinyon@linguist.jussieu.fr)
TALANA-Université Paris 7 / CIS Dpt University of Pennsylvania
Case 7003 2 Pl Jussieu 75005 Paris. France

Abstract

In this paper we discuss how structural Preferences can be expressed within an LTAG framework on dependency like structures. We argue that the use of psycholinguistically motivated criteria is useful for building practical parse-ranking applications.

Introduction

On the one hand computational linguists aim at parsing real texts : it is a difficult task, essentially because of spurious ambiguity. The goal is then to find a single preferred overall analysis for each sentence, either by resorting to general principles or to statistical methods, most of the time by focussing on the efficiency of the technique used, rather than on its theoretical relevance. So most of these disambiguation techniques do not take into account theoretical (i.e. cognitive) relevance, especially the incremental nature of human sentence processing.

On the other hand, psycholinguists aim at modeling the very early preferences which people employ in ambiguity resolution during an incremental parse of a sentence, without being concerned with the development of wide-coverage parsing systems or the integration of their principles in wide-coverage grammars. Importantly, these early decisions that people make may or may not eventually be compatible with the overall analysis of the sentence.

In this paper we argue that the two approaches are not antagonistic : wide-coverage disambiguation systems can integrate psycholinguistically motivated principles and yet be efficient. We present structural preferences which are expressed on dependency structures instead of constituent trees, within the framework of Lexicalized Tree Adjoining Grammars (LTAGs). In the first part of this paper, we briefly introduce the LTAG formalism. Then we present the preference principles used, and show that they work well in practice on large data. In a third part, we show why "pure" lexicalist approaches seem insufficient. In a fourth part, we discuss the interaction between our preference principles.

1. Brief overview 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 $\downarrow$). These trees are of 2 types : **auxiliary** or **initial**¹. An auxiliary tree has exactly one distinguished leaf, called "foot node" and marked *. 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) and sentential complements (e.g. object sentential complements). 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 which a **derived** tree (i.e. constituent tree) can be obtained. Figure 1 shows the elementary trees anchored when parsing "Yesterday John kicked the bucket"², as well as the derivation trees obtained both for the "literal interpretation" and for the "idiomatic interpretations" of the sentence. It also shows that both derivation trees yield the same derived tree³. The derivation tree is close to a dependency structure (cf Candito & Kahane 98).

Moreover, linguistic constraints on the well-formedness of elementary trees have been formulated (Abeillé 91) (Frank 92) :

- Predicate Argument Cooccurrence Principle : there must be a leaf node for each realized argument of the head of an elementary tree.

¹ Traditionally initial trees are called α , and auxiliary trees β

² All our examples follow linguistic analyses presented in [Abeillé 91]. Thus we use no VP node and no Wh nor NP traces. But this has no incidence on the application of our preference principles.

³ Dotted lines in derivation trees indicate a substitution, plain lines an adjunction. The number at each node represents the address at which the operation took place, following Gorn convention.

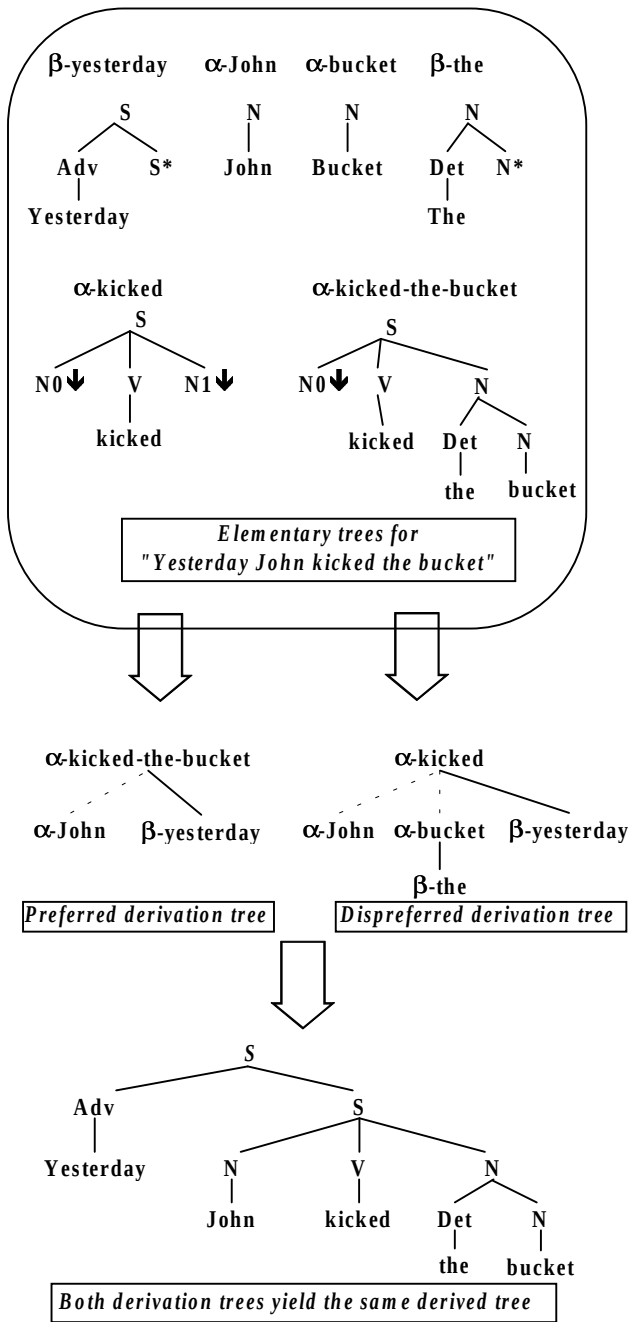


FIGURE 1 : Illustration of LTAG and of Principle 1

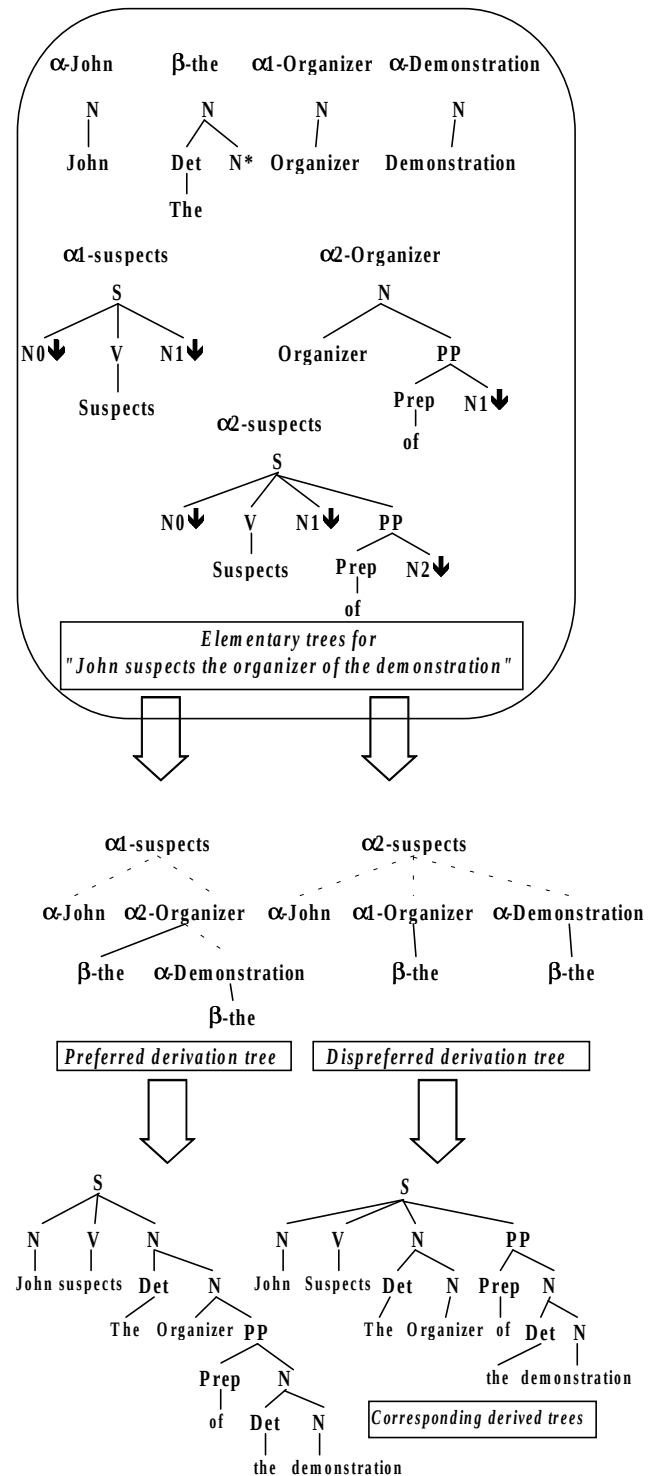


FIGURE 2 : Illustration of Principle 2

- Semantic consistency : No elementary tree is semantically void
- Semantic minimality : an elementary tree corresponds at most to one semantic unit

2. Three preference principles expressed on derivation trees

A vast literature, going back as early as (Kimball 73), addresses structural parsing preferences. Older principles, such as right association (RA) and minimal attachment (MA) have been criticized : Among other things, the interaction between these principles is unclear. These principles lack provision for integration with semantics and/or pragmatics (Schubert (84)), do not clearly establish the distinction between arguments and modifiers⁴ (Ferreira & Clifton (86)) and are English-biased : evidence against RA has been found for Spanish (Cuetos & Mitchell (88)) and Dutch (Brysbaert & Mitchell (96)). Newer structural principles, on the other hand, such as "Attach anyway" (Fodor & Inoue 98), are not integrated nor implemented into wide-coverage grammars.

So, to account for widely accepted preference principles, which are difficult to formulate in terms of constituents trees (idiomatic interpretation of a sentence favored over its literal interpretation (Abeille 95) (Gibbs 85) (Gibbs & Nayak 89)., arguments favored over adjuncts (Abney 89), (Britt & al. 92) and attachment to closest potential governor), (Kinyon 99a) has formulated the three following principle on dependency-like structures within the LTAG framework :

- 1-Prefer the derivation tree with the fewer number of nodes
- 2-Prefer to attach an α -tree low in a derivation tree
- 3-Prefer the derivation tree with the fewer number of β -tree nodes⁵

A discussion on the linguistic adequacy of these principles, as well as on why LTAGs are better than other lexicalized formalisms such as LFG to formulate these principles can be found in Kinyon (99b).

Principle 1 accounts for the preference we have for the idiomatic interpretation of a sentence. In LTAGs, all the set elements of the expression are present in a single elementary tree. We have shown in Figure 1 the derivation trees obtained when parsing "Yesterday John kicked the bucket". The derivation tree for the idiomatic interpretation, which is preferred, has fewer nodes than the derivation tree for the literal interpretation..

⁴ It is argued that MA and RA do distinguish arguments from modifiers, since arguments will yield a constituent tree with fewer nodes, but this relies very heavily on the underlying syntactic framework : it may be true for an X-bar theory, not necessarily for a more "surfastic " theory of syntax.

⁵ This principle was initially presented in (Srinivas & al 95), formulated as "prefer to substitute rather than to adjoin".

Principle 2 captures the preference for an argument to attach to its closest potential governor. So in (a1), "of the demonstration" is preferably attached to "organizer" rather than to "suspect". Similarly, in (a2), "To whom" attaches to "say" rather than to "gives". Figure 2 shows how principle 2 yields the preferred derivation tree for sentence.(a1).

- (a1) John suspects the organizer of the demonstration
- (a2) To whom does Mary say that John gives flowers

Finally, principle 3 accounts for the preference of arguments over adjuncts. So it will allow to retrieve the right attachment in (b1), where "le matin" (the morning) is argument of "regarde" (watches) rather than modifier. It also allows to retrieve the correct attachment in (b2) where "to be honest" is argument of prefer, rather than sentence modifier.

- (b1) Jean regarde le matin (John observes the morning / John watches in the morning)
- (b2) John prefers his daughter to be honest

These principles are easy to implement, so they have yielded practical results⁶ : A parse-ranker has been implemented for French within the FTAG project (cf Abeille & al 99,00), using a semi-automatically generated wide coverage grammar of 5000 elementary trees (Candito 96). This parse ranker, tested on 1000 TSNLP sentences, allows to go down from 2.85 derivations trees / sentence to 1.4 derivation trees / sentence without degrading the quality of parsing (i.e. without discarding "correct" parse trees). These results hint that the three principles are well-motivated from a cognitive point of view. This parse ranker is currently being ported to English and tested on the WSJ.

It is important to note that the distinction between arguments and modifiers can be easily expressed within LTAGs, because in derivation trees elementary trees for arguments are essentially initial (α), while elementary trees for modifiers are auxiliary (β).

It is also important to note that (contrary to RA) these structural preferences are language independent, again because they are formulated on dependancy-like structures and not on constituent structures : we have just seen that they work well for French, although French is argued to be an "early closure" language (Zagar & al 97).

3. Antagonism with lexicalist approaches ?

It has been shown that humans do exhibit frequency effects in language comprehension (Truwell 96), but this does not mean that structural principles are unsound and especially it does not demonstrate that disambiguation systems should resort only to "pure" lexicalist approaches :

One argument against the structural principles presented in 3 would be to say that these structural principles do not

⁶ We do not claim, however, that these principles have yielded better results in automatic disambiguation than statistical parsers which integrate lexical information (e.g. Collins 96). Clearly though, our technique is easier to put to use, esp. for languages for which no training data is available.

exist (i.e. are not observable once frequency effects are taken into account). We disagree for the following reasons :

If the use of such principles was just a mere approximation, it would make it hard to explain that the empirical results are so good. Pure lexicalist approaches have not yielded such results to our knowledge on large real-world data (very little data about lexical preferences are available on a large scale esp. for languages other than English).

Also, lexicalist approaches do not allow to explain how two preferred subcategorization frames interact. For example, if "suspect N of N" and "organizer of N" are two preferred realization frames for respectively "suspect" and "organizer", one still needs to account for the fact that "demonstration" will be attached to "organizer" rather than to "suspect" in "John suspects the organizer of the demonstration"⁷. With the same type of reasoning, although "put N in N" is a common realization frame for arguments of "put", the sentence (c) nonetheless seems incomplete. This can also not be accounted for with a pure lexicalist approach

(c) I've put the book that you were reading in the library

Moreover, lexicalist approaches also do not necessarily account for unknown words⁸ : in a sentence like (d) "at the man" will most likely be preferred as argument of "plups" rather than modifier, although we know nothing about the preferred subcategorization of "plups". So when considering language acquisition, unknown words are still processed, although no data is available regarding the preference of realization of their arguments. Structural preferences thus appear as a much more economical way to acquire new words : here "at the man" is argument of "plups" so "plups" subcategorizes a PP introduced by "at", whereas if one had to rely on frequency effect, it would take much longer to encounter "plups" many times before formulating a hypothesis about its subcategorization, and verifying it. Also, resorting to very few structural preferences for disambiguation seems much more economical and practical than storing huge quantities of frequency information about the lexicon, especially since contrary to structural preferences, information on the lexicon has to change from language to language.

(d) He plups at the man

Finally, to oppose lexicalist approaches and support the structural principles presented in 3, (Kinyon 00) formulated the following hypothesis :

Regardless of which realization of arguments a verb favors, if it can subcategorize a PP introduced by a given Preposition P, then in practice when the verb and a PP introduced by P appear in the same sentence, the PP is either an argument of the verb, or in a position where it can not be argument (i.e. argument of a closer potential governor, or located

in another clause such as inside a relative, or modifier only if the verb is already saturated).

This hypothesis was validated on LeMonde, a one million words annotated and shallow-parsed corpus for French (Clément & Kinyon 00, Abeillé & al 00b). The 100 most frequent verbs were extracted. 56 of these verbs could subcategorize PPs introduced by one or several prepositions, for a total of 71 subcategorization frames. Then, for each of these subcategorization frames, all the sentences where Verb and Prep co-occur were extracted and examined manually. The main findings are the following :

1- Cases of possible ambiguous attachment remain (13.86 % of the sentences examined)

2- 39% of these ambiguous cases are solved when attaching the PP to the closest potential governor. Moreover, the attachment is deemed correct in all cases.

3- The probability for a verb to realize as an argument a PP introduced by a given Preposition P does not help disambiguation and does not predict the proportion of ambiguous attachments encountered when examining sentences where Verb and P co-occur.

4- Rather, the preposition itself is important : "à" yields much more ambiguity than other prepositions such as "avec" or "pour" because it often introduces a temporal or locational expression (e.g. "à l'assemblée nationale" (*in parliament*) / "à 3 heures" (*at 3 o'clock*)). In fact, 46% of the ambiguous cases remaining after applying structural principles 2 and 3 are solved by resorting to very simple semantic information : à + location nouns , à + time nouns are overwhelmingly adjuncts and not arguments.

Therefore, only 4.6 % of ambiguous attachments remain (mainly set phrases such as "lancer un appel au calme"), which could be disambiguated by refining semantic disambiguation. Thus the hypothesis is validated, which indicates that the use of structural principles + basic semantic information allows very efficient disambiguation and again, in a more economical manner than lexical approaches.

As discussed in (Kinyon 99b) though, some lexical preferences seem useful, but formulated not at the level of lexical items, but rather at the level of parts of speech. So for instance, grammatical categories are preferred over lexical categories. So in (e1) clitic will be preferred over noun for "elle", in (e2) "être" (be) will be an auxiliary rather than a lexical verb, and in (e3) "deux" will be a determiner rather than a noun. General lexical preferences of this type have been incorporated in the parse-ranker discussed in 3. Expressing lexical preferences in such general terms is also economical and allows to eliminate some cases of spurious ambiguity.

(e1) Elle court (She runs / It is her who runs)

(e2) Elle est venue (She has arrived / She is an arrival)

(e3) Je vois deux hommes (I see two men)

⁷ Whereas claiming that arguments prefer to attach to their closest potential governors solves this problem.

⁸ One may say that more general information is used when encountering unknown lexical items, but this general idea is not implementable as such.

4. Conflicts between structural principles

One of the main argument against "traditional" structural principles is that the interaction between them is unclear. It has been said for example that in case of conflict, minimal attachment prevails over right association in a sentence such as "He repaints the wall with cracks" thus allowing to account for the garden path effect. Of course, this suffers numerous counter-examples.

With the structural principles presented in 3 and expressed on dependency like structures, it is striking that zero conflicts were encountered, both on the 1000 sentences for French, and on 3000 sentences from the wall street journal for English.

This strongly suggests that these principles are relevant from a cognitive point of view.

Conclusion

We have presented three parsing preference principles expressed on dependency like structures, and shown that these language-independent principles are both psycholinguistically relevant and useful to disambiguate real-word data on a large scale (which has led to the development of a parser-ranker). We also came to the conclusion that an efficient disambiguation scheme involving these structural preferences as well as limited semantic information and "simplified" lexical principles (i.e. expressed in terms of parts of speech) was more economical than acquiring large amounts of lexical data, thus being more appealing both from a practical and from a cognitive point of view. In fact, these structural preferences are a first step towards a psycholinguistically relevant processing model for LTAGs, which allows among other things to predict garden-path phenomena (cf Kinyon 99c, Kinyon 00b).

REFERENCES

- Abeillé A. (1991) Une grammaire lexicalisée d'arbres adjoints pour le français : application à l'analyse automatique. Thèse de doctorat. Université Paris 7.
- Abeillé A. (1995) The flexibility of French idioms. In Idioms LEA. Schenk & al. (eds).
- Abeillé A. Candito M.H. Kinyon A. (1999) FTAG : current status and parsing scheme. Proceedings Vextal'99. Venice.
- Abeillé A. Clément L. Kinyon A. (2000) Building a treebank for French. Proc. LREC'2000. Athens
- Abeillé A. Candito M.H. Kinyon A. (2000) The current status of FTAG. Proc. TAG+5. Paris.
- Abney S. (1989) A computational model of human parsing. Journal of psycholinguistic Research, 18, pp. 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.
- Candito M.H. (1996) A principle based hierarchical representation of LTAG. Proc. 15th COLING. Copenhagen.
- Candito M.H., Kahane S. (1998). Can the TAG derivation tree represent a semantic graph ? an answer in the light of MTT. Proc. TAG+5. Philadelphia.
- Clément L. Kinyon A (2000). Chunking, marking and searching a morpho-syntactically annotated corpus for French. Proc. ACIDCA'2000. Monastir.
- Collins M. (1996) Three generative, Lexicalised Models for statistical parsing. Proc. ACL'97. Madrid.
- 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.
- Ferreira F. Clifton C. (1986) *The independence of syntactic processing*. Journal of Memory and Language, 25,348-368.
- Fodor J.D. Inoue A. (1998). Attach Anyway. In Reanalysis in Sentence Processing. Fodor & Ferreira (eds). Kluwer academic publishers.
- Frank R. (1992) Syntactic Locality and Tree Adjoining Grammar : Grammatical Acquisition and Processing Perspectives. PhD dissertation. Univ. of Pennsylvania.
- Gibbs R. (1985) On the process of understanding idioms. Journal of psycholinguistic research, 14, pp. 465-477.
- Gibbs R., Nayak (1989) Psycholinguistic studies on the syntactic behaviour of idioms. Cognitive Psychology, 21, pp. 100-138.
- Kimball J. (1973) Seven principles of surface structure parsing in natural language. Cognition 2.
- Kinyon A. (1999a) : Parsing preferences with Lexicalized Tree Adjoining Grammars : exploiting the derivation tree. Proc. ACL'99
- Kinyon A. (1999b) : Hiérarchisation d'analyses basée sur des informations dépendanciennes dans le cadre des LTAGs. Proceedings TALN'99.
- Kinyon A. (1999c). Some remarks about the psycholinguistic relevance of LTAGs. CLIN'99. Utrecht
- Kinyon A. (2000a). Structural preferences vs Lexical preferences : some data on French verbs subcategorizing a PP. Poster presented at Cuny'2000. La Jolla. California.
- Kinyon A. (2000b). Hypertags. Proc. COLING'2000. Sarrebrücken.
- Schubert L. (1984). *On parsing preferences*. COLING'84, Stanford. 247-250.
- Srinivas B., Doran C., Kulick S. (1995) : Heuristics and Parse Ranking. Proc IWPT'95. Prag. Czech Republic.

Trueswell, J.C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language*, 35, 566-585.

Zagar D., Pynte J., Rativeau S. (1997) Evidence for early closure attachment on first pass reading times in French. *Quarterly journal of experimental psychology*, 50(A), 421-438.