

ON PARSING PREFERENCES

Lenhart K. Schubert

Department of Computing Science
University of Alberta, Edmonton

Abstract. It is argued that syntactic preference principles such as Right Association and Minimal Attachment are unsatisfactory as usually formulated. Among the difficulties are: (1) dependence on ill-specified or implausible principles of parser operation; (2) dependence on questionable assumptions about syntax; (3) lack of provision, even in principle, for integration with semantic and pragmatic preference principles; and (4) apparent counterexamples, even when discounting (1)-(3). A possible approach to a solution is sketched.

1. Some preference principles

The following are some standard kinds of sentences illustrating the role of syntactic preferences.

- (1) John bought the book which I had selected for Mary
- (2) John promised to visit frequently
- (3) The girl in the chair with the spindly legs looks bored
- (4) John carried the groceries for Mary
- (5) She wanted the dress on that rack
- (6) The horse raced past the barn fell
- (7) The boy got fat melted

(1) - (3) illustrate Right Association of PP's and adverbs, i.e., the preferred association of these modifiers with the rightmost verb (phrase) or noun (phrase) they can modify (Kimball 1973). Some variants of Right Association (also characterized as Late Closure or Low Attachment) which have been proposed are Final Arguments (Ford et al. 1982) and Shifting Preference (Shieber 1983); the former is roughly Late Closure restricted to the last obligatory constituent and any following optional constituents of verb phrases, while the latter is Late Closure within the context of an LR(1) shift-reduce parser.

Regarding (4), it would seem that according to Right Association the PP for Mary should be preferred as postmodifier of groceries rather than carried; yet the opposite is the case. Frazier & Fodor's (1979) explanation is based on the assumed phrase structure rules VP → V NP PP, and NP → NP PP: attachment of the PP into the VP minimizes the resultant number of nodes. This principle of Minimal Attachment is assumed to take precedence over Right Association. Ford et al's (1982) variant is Invoked Attachment, and Shieber's (1983) variant is Maximal Reduction; roughly speaking, the former amounts to early closure of non-final constituents,

while the latter chooses the longest reduction among those possible reductions whose initial constituent is "strongest" (e.g., reducing V NP PP to VP is preferred to reducing NP PP to PP).

In (5), Minimal Attachment would predict association of the PP on that rack with wanted, while the actual preference is for association with dress. Both Ford et al. and Shieber account for this fact by appeal to lexical preferences: for Ford et al., the strongest form of want takes an NP complement only, so that Final Arguments prevails; for Shieber, the NP the dress is stronger than wanted, viewed as a V requiring NP and PP complements, so that the shorter reduction prevails.

Sentence (6) leads most people "down the garden path", a fact explainable in terms of Minimal Attachment or its variants. The explanation also works for (7) (in the case of Ford et al. with appeal to the additional principle that re-analysis of complete phrases requiring re-categorization of lexical constituents is not possible). Purportedly, this is an advantage over Marcus' (1980) parsing model, whose three-phrase buffer should allow trouble-free parsing of (7).

2. Problems with the preference principles

2.1 Dependence on ill-specified or implausible principles of parser operation.

Frazier & Fodor's (1979) model does not completely specify what structures are built as each new word is accommodated. Consequently it is hard to tell exactly what the effects of their preference principles are.

Shieber's (1983) shift-reduce parser is well-defined. However, it postulates complete phrases only, whereas human parsing appears to involve integration of completely analyzed phrases into larger, incomplete phrases. Consider for example the following sentence beginnings:

- (8) So I says to the ...
 - (9) The man reconciled herself to the ...
 - (10) The news announced on the ...
 - (11) The reporter announced on the ...
 - (12) John beat a rather hasty and undignified ...
- People presented with complete, spoken sentences beginning like (8) and (9) are able to signal detection of the errors about two or three syllables after their occurrence. Thus agreement

features appear to propagate upward from incomplete constituents. (10) and (11) suggest that even semantic features (logical translations?) are propagated before phrase completion. The "premature" recognition of the idiom in (12) provides further evidence for early integration of partial structures.

These considerations appear to favour a "full-paths" parser which integrates each successive word (in possibly more ways than one) into a comprehensive parse tree (with overlaid alternatives) spanning all of the text processed.

Ford et al.'s (1982) parser does develop complete top-down paths, but the nodes on these paths dominate no text. Nodes postulated bottom-up extend only one level above complete nodes.

2.2 Dependence on questionable assumptions about syntax

The successful prediction of observed preferences in (4) depended on an assumption that PP postmodifiers are added to carried via the rule VP → V NP PP and to groceries via the rule NP → NP PP. However, these rules fail to do justice to certain systematic similarities between verb phrases and noun phrases, evident in such pairs as

(13) John loudly quarreled with Mary in the kitchen

(14) John's loud quarrel with Mary in the kitchen
When the analyses are aligned by postulating two levels of postmodification for both verbs and nouns, the accounts of many examples that supposedly involve Minimal Attachment (or Maximal Reduction) are spoiled. These include (4) as well as standard examples involving non-preferred relative clauses, such as

(15) John told the girl that he loved the story

(16) Is the block sitting in the box?

2.3 Lack of provision for integration with semantic/pragmatic preference principles

Right Association and Minimal Attachment (and their variants) are typically presented as principles which prescribe particular parser choices. As such, they are simply wrong, since the choices often do not coincide with human choices for text which is semantically or pragmatically biased.

For example, there are conceivable contexts in which the PP in (4) associates with the verb, or in which (7) is trouble-free. (For the latter, imagine a story in which a young worker in a shortening factory toils long hours melting down hog fat in clarifying vats.) Indeed, even isolated sentences demonstrate the effect of semantics:

(17) John met the girl that he married at a dance

(18) John saw the bird with the yellow wings

(19) She wanted the gun on her night table

(20) This lens gets light focused

These sentences should be contrasted with (1), (4), (5), and (7) respectively.

While the reversal of choices by semantic and pragmatic factors is regularly acknowledged, these factors are rarely assigned any explicit role in

the theory; (however, see Crain & Steedman 1981). Two views that seem to underlie some discussions of this issue are (a) that syntactic preferences are "defaults" that come into effect only in the absence of semantic/pragmatic preferences; or (b) that alternatives are tried in order of syntactic preference, with semantic tests serving to reject incoherent combinations. Evidence against both positions is found in sentences in which syntactic preferences prevail over much more coherent alternatives:

(21) Mary saw the man who had lived with her while on maternity leave.

(22) John met the tall, slim, auburn-haired girl from Montreal that he married at a dance

(23) John was named after his twin sister

What we apparently need is not hard and fast decision rules, but some way of trading off syntactic and non-syntactic preferences of various strengths against each other.

2.4 Apparent counterexamples.

There appear to be straightforward counterexamples to the syntactic preference principles which have been proposed, even if we discount evidence for integration of incomplete structures, accept the syntactic assumptions made, and restrict ourselves to cases where none of the alternatives show any semantic anomaly.

The following are apparent counterexamples to Right Association (and Shifting Preference, etc.):

(24) John stopped speaking frequently

(25) John discussed the girl that he met with his mother

(26) John was alarmed by the disappearance of the administrator from head office

(27) The deranged inventor announced that he had perfected his design of a clip car shoe (shoe car clip, clip shoe car, shoe clip car, etc.)

(28) Lee and Kim or Sandy departed

(29) a. John removed all of the fat and some of the bones from the roast
b. John removed all of the fat and sinewy pieces of meat

The point of (24)-(26) should be clear. (27) and (28) show the lack of right-associative tendencies in compound nouns and coordinated phrases. (29a) illustrates the non-occurrence of a garden path predicted by Right Association (at least by Shieber's version); note the possible adjectival reading of fat and ..., as illustrated in (29b).

The following are apparent counterexamples to Minimal Attachment (or Maximal Reduction):

(30) John abandoned the attempt to please Mary

(31) Kim overheard John and Mary's quarrel with Sue

(32) John carried the umbrella, the transistor radio, the bundle of old magazines, and the groceries for Mary

(33) The boy got fat spattered on his arm

While the account of (30) and (31) can be rescued by distinguishing subcategorized and non-subcategorized noun postmodifiers, such a move would lead to the failures already mentioned in section 2.2. Ford et al. (1982) would have no

trouble with (30) or (31), but they, too, pay a price: they would erroneously predict association of the PP with the object NP in

- (34) Sue had difficulties with the teachers
- (35) Sue wanted the dress for Mary
- (36) Sue returned the dress for Mary

(32) is the sort of example which motivated Frazier & Fodor's (1979) Local Attachment principle, but their parsing model remains too sketchy for the implications of the principle to be clear. Concerning (33), a small-scale experiment indicates that this is not a garden path. This result appears to invalidate the accounts of (7) based on irreversible closure at fat. Moreover, the difference between (7) and (33) cannot be explained in terms of one-word lookahead, since a further experiment has indicated that

(37) The boy got fat spattered.
is quite as difficult to understand as (7).

3. Towards an account of preference trade-offs

My main objective has been to point out deficiencies in current theories of parsing preferences, and hence to spur their revision. I conclude with my own rather speculative proposals, which represent work in progress.

In summary, the proposed model involves (1) a full-paths parser that schedules tree pruning decisions so as to limit the number of ambiguous constituents to three; and (2) a system of numerical "potentials" as a way of implementing preference trade-offs. These potentials (or "levels of activation") are assigned to nodes as a function of their syntactic/semantic/pragmatic structure, and the preferred structures are those which lead to a globally high potential. The total potential of a node consists of (a) a negative rule potential, (b) a positive semantic potential, (c) positive expectation potentials contributed by all daughters following the head (where these decay with distance from the head lexeme), and (d) transmitted potentials passed on from the daughters to the mother.

I have already argued for a full-paths approach in which not only complete phrases but also all incomplete phrases are fully integrated into (overlaid) parse trees dominating all of the text seen so far. Thus features and partial logical translations can be propagated and checked for consistency as early as possible, and alternatives chosen or discarded on the basis of all of the available information.

The rule potential is a negative increment contributed by a phrase structure rule to any node which instantiates that rule. Rule potentials lead to a minimal-attachment tendency: they "inhibit" the use of rules, so that a parse tree using few rules will generally be preferred to one using many. Lexical preferences can be captured by making the rule potential more negative for the more unusual rules (e.g., for N → fat, and for V → time).

Each "expected" daughter of a node which follows the node's head lexeme contributes a non-negative

expectation potential to the total potential of the node. The expectation potential contributed by a daughter is maximal if the daughter immediately follows the mother's head lexeme, and decreases as the distance (in words) of the daughter from the head lexeme increases. The decay of expectation potentials with distance evidently results in a right-associative tendency. The maximal expectation potentials of the daughters of a node are fixed parameters of the rule instantiated by the node. They can be thought of as encoding the "affinity" of the head daughter for the remaining constituents, with "strongly expected" constituents having relatively large expectation potentials. For example, I would assume that verbs have a generally stronger affinity for (certain kinds of) PP adjuncts than do nouns. This assumption can explain PP-association with the verb in examples like (4), even if the rules governing verb and noun postmodification are taken to be structurally analogous. Similarly the scheme allows for counterexamples to Right Association like (24), where the affinity of the first verb (stop) for the frequency adverbial may be assumed to be sufficiently great compared to that of the second (speak) to overpower a weak right-associative effect resulting from the decay of expectation potentials with distance.

I suggest that the effect of semantics and pragmatics can in principle be captured through a semantic potential contributed to each node potential by semantic/pragmatic processing of the node. The semantic potential of a terminal node (i.e., a lexical node with a particular choice of word sense for the word it dominates) is high to the extent that the associated word sense refers to a familiar (highly consolidated) and contextually salient concept (entity, predicate, or function). For example, a noun node dominating star, with a translation expressing the astronomical sense of the word, presumably has a higher semantic potential than a similar node for the show-business sense of the word, when an astronomical context (but no show-business context) has been established; and vice versa. Possibly a spreading activation mechanism could account for the context-dependent part of the semantic potential (cf., Quillian 1968, Collins & Loftus 1975, Charniak 1983).

The semantic potential of a nonterminal node is high to the extent that its logical translation (obtained by suitably combining the logical translations of the daughters) is easily transformed and elaborated into a description of a familiar and contextually relevant kind of object or situation. (My assumption is that an unambiguous meaning representation of a phrase is computed on the basis of its initial logical form by context-dependent pragmatic processes; see Schubert & Pelletier 1982.) For example, the sentences Time flies, The years pass swiftly, The minutes creep by, etc., are instances of the familiar pattern of predication

<predicate of locomotion> (<time term>),
and as such are easily transformable into certain commonplace (and unambiguous) assertions about one's personal sense of progression through time. Thus they are likely to be assigned high semantic

potentials, and so will not easily admit any alternative analysis. Similarly the phrases met [someone] at a dance (versus married [someone] at a dance) in sentence (17), and bird with the yellow wings (versus saw [something] with the yellow wings) in (18) are easily interpreted as descriptions of familiar kinds of objects and situations, and as such contribute semantic potentials that help to edge out competing analyses.

Crain & Steedman's (1981) very interesting suggestion that readings with few new presuppositions are preferred has a possible place in the proposed scheme: the mapping from logical form to unambiguous meaning representation may often be relatively simple when few presuppositions need to be added to the context. However, their more general plausibility principle appears to fail for examples like (21)-(23).

Note that the above pattern of temporal predication may well be considered to violate a selectional restriction, in that predicates of locomotion cannot literally apply to times. Thus the nodes with the highest semantic potential are not necessarily those conforming most fully with selectional restrictions. This leads to some departures from Wilks' theory of semantic preferences (e.g., 1976), although I suppose that normally the most easily interpretable nodes, and hence those with the highest semantic potential, are indeed the ones that conform with selectional restrictions.

The difference between such pairs of sentences as (17) and (22) can now be explained in terms of semantic/syntactic potential trade-offs. In both sentences the semantic potential of the reading which associates the PP with the first verb is relatively high. However, only in (17) is the PP close enough to the first verb for this effect to overpower the right-associative tendency inherent in the decay of expectation potentials.

The final contribution to the potential of a node is the transmitted potential, i.e., the sum of potentials of the daughters. Thus the total potential at a node reflects the syntactic/semantic/pragmatic properties of the entire tree it dominates.

A crucial question that remains concerns the scheduling of decisions to discard globally weak hypotheses. Examples like (33) have convinced me that Marcus (1980) was essentially correct in positing a three-phase limit on successive ambiguous constituents. (In the context of a full-paths parser, ambiguous constituents can be defined in terms of "upward or-forks" in phrase structure trees.) Thus I propose to discard the globally weakest alternative at the latest when it is not possible to proceed rightward without creating a fourth ambiguous constituent. Very weak alternatives (relative to the others) may be discarded earlier, and this assumption can account for early disambiguation in cases like (10) and (11).

Although these proposals are not fully worked out (especially with regard to the definition of

semantic potential), preliminary investigation suggests that they can do justice to examples like (1)-(37). Schubert & Pelletier 1982 briefly described a full-paths parser which chains upward from the current word to current "expectations" by "left-corner stack-ups" of rules. However, this parser searched alternatives by backtracking only and did not handle gaps or coordination. A new version designed to handle most aspects of Generalized Phrase Structure Grammar (see Gazdar et al., to appear) is currently being implemented.

Acknowledgements

I thank my unpaid informants who patiently answered strange questions about strange sentences. I have also benefited from discussions with members of the Logical Grammar Study Group at the University of Alberta, especially Matthew Dryer, who suggested some relevant references. The research was supported by the Natural Sciences and Engineering Research Council of Canada under Operating Grant A8818.

References

- Charniak, E. (1983). Passing markers: a theory of contextual influence in language comprehension. Cognitive Science 7, pp. 171-190.
- Collins, A. M. & Loftus, E. F. (1975). A spreading activation theory of semantic processing. Psychological Review 82, pp. 407-428.
- Crain, S. & Steedman, M. (1981). The use of context by the Psychological Parser. Paper presented at the Symposium on Modelling Human Parsing Strategies, Center for Cognitive Science, Univ. of Texas, Austin.
- Ford, M., Bresnan, J. & Kaplan, R. (1981). A competence-based theory of syntactic closure. In Bresnan, J. (ed.), The Mental Representation of Grammatical Relations, MIT Press, Cambridge, MA.
- Frazier, L. & Fodor, J. (1979). The Sausage Machine: a new two-stage parsing model. Cognition 6, pp. 191-325.
- Gazdar, G., Klein, E., Pullum, G. K. & Sag, I. A. (to appear). Generalized Phrase Structure Grammar: A Study in English Syntax.
- Kimball, J. (1973). Seven principles of surface structure parsing in natural language. Cognition 2, pp. 15-47.
- Marcus, M. (1980). A Theory of Syntactic Recognition for Natural Language, MIT Press, Cambridge, MA.
- Quillian, M. R. (1968). Semantic memory. In Minsky, M. (ed.), Semantic Information Processing, MIT Press, Cambridge, MA, pp. 227-270.
- Schubert, L. K. & Pelletier, F. J. (1982). From English to logic: context-free computation of 'conventional' logical translations. Am. J. of Computational Linguistics 8, pp. 26-44.
- Shieber, S. M. (1983). Sentence disambiguation by a shift-reduce parsing technique. Proc. 8th Int. Conf. on Artificial Intelligence, Aug. 8-12, Karlsruhe, W. Germany, pp. 699-703. Also in Proc. of the 21st Ann. Meet. of the Assoc. for Computational Linguistics, June 15-17, MIT, Cambridge, MA., pp. 113-118.
- Wilks, Y. (1976). Parsing English II. In Charniak, E. & Wilks, Y. (eds.), Computational Semantics, North-Holland, Amsterdam, pp. 155-184.