Integrating Statistical and Structural Information in a Distributed Architecture for Syntactic Disambiguation

Paola Merlo
Department of Linguistics
University of Geneva
2 rue de Candolle
1204 Geneva Switzerland
merlo@fapse.unige.ch

Suzanne Stevenson
Dept. of Computer Science,
and Center for Cognitive Science
Rutgers University
New Brunswick, NJ 08903
suzanne@cs.rutgers.edu

Abstract

This paper shows that corpus-based statistical patterns do not entirely concur with human interpretation preferences. Human judges rank reduced relative clauses with verbs from three syntactic classes bi-modally, as relatively easy or extremely difficult, while a large corpus shows a tri-modal distribution of the three classes. We present a parser that accounts for the preference data correctly, guided by structural and frequency factors.

1 Introduction

In order to achieve natural and effective communication, a natural language interface must consistently and efficiently determine the preferred interpretation of an ambiguous input. After many years of NLU research, it has remained difficult to specify a simple set of rules or restrictions for matching human expectations in interpretation. The intractability of heuristic methods, and the recent explosion of availability of on-line text, have inspired a number of researchers to derive interpretation preferences from corpus analysis (e.g., Bod, 1995; Brill and Resnik, 1994; Hindle and Rooth, 1993; Weischedel, Meteer, Schwartz, Ramshaw, and Palmucci, 1993). The extraction of relevant statistics from large corpora has been claimed to enable us to determine human preferences in a more general and accurate manner than previously achieved.

However, a number of studies have discovered discrepancies between preferences in interpretation of ambiguities on the one hand, and corpus counts of the various resolutions of those same ambiguities on the other hand. For example, in an on-line study of conjoined NP attachment preferences to three possible NP sites, Gibson, Schutze, and Salomon (1995) found a significant three-way distinction in preference of attachment to each of the three sites. By contrast, in a corpus study they found only a binary distinction, because two of the attachment possibilities were not significantly different in frequency. One also finds discrepancies for verb subcategorization ambiguities. For example, given verbs that subcategorize for an NP or an S, Frazier and Rayner (1982) found a strong NP attachment preference in interpretation, using stimuli that present a clear bias for the S continuation according to corpus frequencies (Merlo, 1994). Apparently, interpretation preferences for an ambiguity cannot always be determined by the frequency of different resolutions of the ambiguity revealed through corpus analysis.

In this paper, we explore another type of ambiguity for which a simple preference metric derived from corpus frequencies appears insufficient to model human preferences. The ambiguity in question is the main verb/reduced relative (MV/RR) ambiguity exhibited in the classic garden-path sentence, *The horse raced past the barn fell* (Bever, 1970). The difficulty of this sentence is due to the ambiguity of the verb *raced*, which can be either a simple past (the preferred interpretation) or a past participle (the interpretation necessary for a coherent analysis of the sentence). Over the last several years, the contrast in acceptability of the following types of sentences has been noted:

(1) #The horse raced past the barn fell.

(2) The butter melted in the microwave was lumpy.

(3) The patients heard here sound unusually sane.

Since the three sentences are structurally identical, the complete unacceptability of (1) contrasted with the ease of (2) and (3) has defied an explanatory account. Recently, however, it has been claimed that the difference in acceptability arises from verb-specific frequency information (MacDonald, 1994; MacDonald, Pearlmutter, and Seidenberg, 1994; Trueswell, 1995; Trueswell, Tanenhaus, and Garnsey, 1994). In this paper, we demonstrate that a pure frequency approach cannot predict the pattern of human preference data, and propose an alternative model for capturing those preferences. After briefly describing a lexical semantic analysis of the types of verbs exhibited in (1)–(3), we present judgment data that indicates that the first class is hard in a RR construction, while the second and third are relatively easy. Next, we present corpus data that reveals a three-way distinction among the verb classes, and is unable to account for the sharp drop-off in acceptability of the first class. Finally, we present a processing model that can account for the appropriate preference behavior. We conclude with some remarks concerning the import of our findings for pure statistical approaches to disambiguation in an NLU system.

2 Verb Classes and Judgment Data

We make the crucial observation that the prototypically difficult reduced relatives contain a manner of motion verb. In English, these verbs form a subclass of unergative verbs (Levin and Rappaport Hovav, 1995), which may also appear in a transitive/intransitive alternation:

(4a) The horse raced past the barn.

(4b) The rider raced the horse past the barn.

The transitive form of an unergative (4b) is the causative counterpart of the intransitive form (4a), in which the subject of the intransitive becomes the object of the transitive (Brousseau and Ritter, 1991; Hale and Keyser, 1993; Levin and Rappaport Hovav, 1995). A first hypothesis then might be that the causative nature of a verb like *raced* is what leads to difficulty in the reduced relative construction. However, unaccusative verbs, such as *melt*,

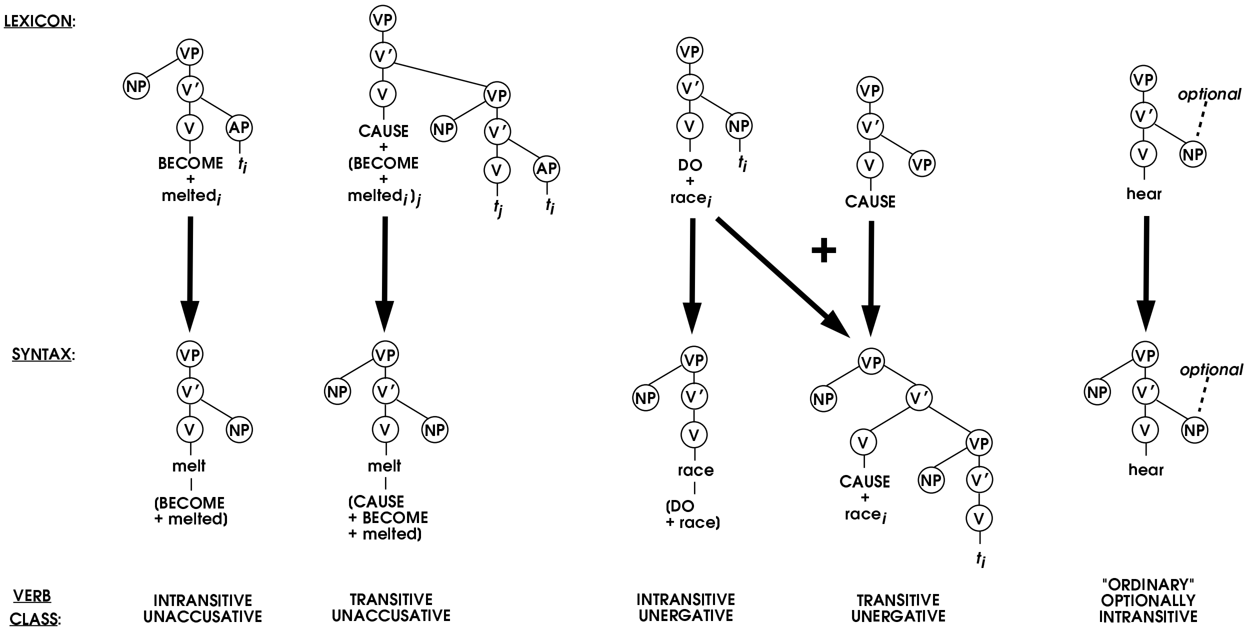

Figure 1: Lexical entries, showing verb incorporation, for the three verb classes, and the corresponding syntactic structures.

do not give rise to processing difficulty in the reduced relative (compare sentence (2) above), yet they can also occur in a transitive/intransitive alternation related by causativization:

(5a) The butter melted in the pan.

(5b) The cook melted the butter in the pan.

We observe, however, that there is a difference in theta role assignments between unergative and unaccusative verbs. In an intransitive unergative, such as (4a), the subject NP is an *agent*, and in an intransitive unaccusative, such as (5a), the subject is a *theme* or *patient*. The theta role assignments to the corresponding arguments of the transitive forms are the same, with the addition of a *causal agent* as subject in both cases. This leads to an unusual situation for the transitive unergatives, in which the subject is a causal agent, but the object is also an agent, of the action expressed by the verb. For example, in (4b), although *the rider* causes *the horse* to race, *the horse* is the *agent* or *actor* of the actual racing.

It has been argued elsewhere that this difference in thematic role assignment is reflected in the syntactic structure of the transitive variant of these verbs (Stevenson and Merlo, 1995). Specifically, an unaccusative can undergo lexical causativization, yielding two distinct intransitive (non-causative) and transitive (causative) forms in the lexicon. However, an unergative cannot undergo lexical causativization, as that entails the lexical subordination of an agent theta role (for details, see Stevenson and Merlo, 1995; Hale and Keyser, 1993). An unergative verb must undergo causativization explicitly in the syntax, yielding a more complex syntactic structure for its transitive form; see Figure 1.

Finally, an optionally intransitive verb such as *hear* has a non-causative transitive/intransitive alternation, in which the object NP is simply not required. Thus, the verb has a single lexical entry with its internal argument marked as optional; see Figure 1. We

	Ambiguous		Unambiguous	
	Verb	Score	Verb	Score
Unaccusative	melt	2	begin	2
	mutate	1.66	break	1
	pour	1.66	freeze	1.5
	reach	1	grow	1
			sink	3.25
Unergative	advance	5	fly	4.25
	glide	5	ring	3.75
	march	5	run	5
	rotate	5	withdraw	3.40
	sail	5		
	walk	3.75		

Table 1: Approximate quantification of the difficulty of unergative and unaccusative verbs in the reduced relative construction. Judgments were ranked on a scale of 1 to 5, with 1 indicating completely acceptable and 5 completely unacceptable.

refer to this third type of verb as an "ordinary" verb, to contrast with the unergative and unaccusative classes defined above.

Experimental results show that such "ordinary" optionally intransitive verbs are easily understood in a reduced relative, given a context that is not strongly biased toward a main verb reading (MacDonald, 1994). We tested the relative difficulty of unergatives and unaccusatives by asking naive informants for acceptability judgments on sentences with RRs containing these verbs. The judgments revealed a clear pattern, shown by the average relative difficulty scores reported in Table 1. All the difficult verbs (score ≥ 3), with the exception of *sink*, belong to the unergative class, while all the easy verbs belong to the unaccusative class. Moreover, even unambiguous past participles of an unergative verb cause difficulty in the RR (average difficulty rating for all unergatives > 4.0).

The fact that differential difficulty clearly patterns with grammatical classes, and the fact that hard reduced relatives occur also when the past participle is unambiguous, leads us to believe that a purely probabilistic model is insufficient to explain the pattern of difficulty. In the next section, we demonstrate that the indication of a grammatical phenomenon which emerges from the acceptability judgments is supported by corpus analysis.

3 Corpus Analysis

Here we present corpus data that disconfirms the hypothesis that difficulty in processing reduced relatives is related either directly to the frequency of the RR construction, or to several properties of lexical items that are highly correlated to the RR construction. Specifically, we investigate the following hypothesis, in which α ranges over a number of relevant properties:

	RR	MV	TOTALS
UNERGATIVES[a]	1	327	328
UNACCUSATIVES[b]	6	358	364
ORDINARY[c]	16	361	377

[a] *race*, *march*, *sail*, *glide*, *fly*
[b] *melt*, *break*, *grow*, *freeze*, *sink*
[c] *fight*, *lecture*, *study*, *surrender*, *watch*

Table 2: Frequency counts of reduced relative and main verb usages.

$\mathbf{H}_0$**:** the difficulty of the RR interpretation is inversely correlated with the frequency of α.

We have tested this hypothesis for several values of α—the RR construction itself, transitivity, passive voice, and past participle frequency—across the three classes of verbs. The counts for the RR construction and transitivity were done manually on 1.5 million words of the Wall Street Journal. The counts for the last two tests were done automatically on the Tree Bank II bracketed and tagged portion of the Wall Street Journal. When these counts were insufficient, they were supplemented by additional manual counts from the first corpus.

Recall that unergative verbs are very difficult in the RR construction, and unaccusative and ordinary verbs are relatively easy. Thus, for the hypothesis to be confirmed, unergatives must exhibit a significantly lower frequency of α compared to unaccusative and ordinary verbs. While we do not have a direct experimental comparison between unaccusative and ordinary verbs, we note that there is no sharp contrast in acceptability between them, as occurs between unergatives and other verbs. Support for the hypothesis therefore also requires there to be little to no difference in α across these two verb classes.

3.1 The Data

$\alpha =$ **RR construction** We compared the frequency of the RR construction and main verb usage for the three classes of verbs; see Table 2. For the unambiguous past participles, the RR occurrences were compared to the appropriate MV usages, even if not homophonous (e.g., *flew*, *flown*), in order to cover the same set of forms as the ambiguous past participles. Applying Fisher's exact test (because some of the expected frequencies are less than 5) revealed that the difference between unergatives and unaccusatives is not significant, with $p > 0.05$. This finding disconfirms the hypothesis that difficulty of RRs is in inverse proportion to its frequency. The corollary hypothesis that ordinary verbs and unaccusatives are statistically identical is also disconfirmed, as the two distributions are significantly different ($\chi^2(1) = 4.3, p < 0.05$).

$\alpha =$ **transitivity** The same sets of verbs were classified according to their transitivity; see Table 3. The unergative and unaccusative distributions are clearly different; $\chi^2(1) = 23.68$,

	TRANS	INTR	TOTALS
UNERGATIVES[a]	86	242	328
UNACCUSATIVES[b]	176	228	404
ORDINARY[c]	268	114	382

[a] *race, march, sail, glide, fly*
[b] *melt, break, grow, freeze, sink*
[c] *fight, lecture, study, surrender, watch*

Table 3: Frequency counts of transitive and intransitive usages.

	ACTIVE	PASSIVE	TOTALS
UNERGATIVES[a]	355	7	362
UNACCUSATIVES[b]	439	24	463
ORDINARY[c]	241	87	328

[a] *jump, march, move, race*
[b] *break, freeze, grow, sink*
[c] *fight, lecture, study, watch*

Table 4: Frequency counts of active and passive usages.

$p < 0.001$. Here the hypothesis—that difficulty of the RR is inversely related to frequency of transitivity—seems to be confirmed. However, comparing the unaccusative and ordinary verbs, the hypothesis is not supported. These two classes of verbs do not appear to be markedly different in acceptability in the RR, but the two distributions are clearly different ($\chi^2(1) = 56.48$, $p < 0.001$).

α = passive voice The MV and RR interpretations also differ in voice, with the MV active and the RR passive. The frequency of active and passive forms across verb classes was tested; the active count includes the simple past and compound perfect tenses, while the passive count includes the full passives, reduced relatives and predicative uses. Comparing unergative and unaccusative verbs (Table 4), we find that the variation in voice is not independent of the verb class ($\chi^2(1) = 5.94$, $p < 0.025$). This result could explain the difference in difficulty, but, again, if we compare the unaccusative and ordinary verbs, we find that a significantly different distribution in the active and passive forms also obtains, and does *not* correspond to a perceived differential difficulty ($\chi^2(1) = 72.48$, p< 0.001).

α = past participle Trueswell (1995) found results supporting the hypothesis that past participle frequency is a predictor of RR difficulty. We therefore also investigated the past

	PP	MV/PRED	TOTALS
UNERGATIVES[a]	76	576	652
UNACCUSATIVES[b]	87	393	480
ORDINARY[c]	147	220	367

[a]*jump*, *march*, *race*, *run*, *sail*
[b]*break*, *freeze*, *grow*, *melt*, *sink*
[c]*fight*, *lecture*, *study*, *surrender*, *watch*

Table 5: Frequency counts of past participle and main verb/predicative usages.

	TRANS	PASSIVE	PAST PRT
UNERGATIVES	.26	.02	.12
UNACCUSATIVES	.44	.05	.18
ORDINARY	.70	.26	.40

Table 6: Relative frequencies for several values of α in the three verb classes.

participle frequencies of our verb classes; the distributions are given in Table 5, where the occurrence of past participles is opposed to the occurrence of the verb in its main or predicative form. A χ^2 test for the unergative and unaccusative verb classes was highly significant ($\chi^2(1)$= 49.98, $p < 0.001$), supporting past participle frequency as a predictor of difficulty. In this case too, though, if the ordinary verb class is compared to unaccusatives, then frequency counts of the past participle make the wrong prediction—a χ^2 test shows that the two distributions are significantly different ($\chi^2(1) = 57.84$, $p < 0.001$).

3.2 Discussion

We have found that the hypotheses of inverse correlation between frequency and difficulty of interpretation is disconfirmed in all cases, including those that link the difficulty to a lexical item. In several cases, the corpus data indicate a three-way distinction between the three verb classes, while preference data indicates a binary distinction, with only the unergative class causing severe processing difficulty. When the corpus data indicates a two-way distinction (α = RR construction), the partitioning of verbs is incorrect, as ordinary verbs contrast with unaccusatives and unergatives. Since, for the other values of α, the relevant frequency of α is significantly less for unergatives than for the other verbs, one might propose that the preference data results from a threshold effect. However, if we look at the global comparison of relative frequencies for these values of α in the three verb classes (Table 6), the data reveal that the unaccusatives pattern more closely to the unergatives than to the ordinary verbs, making the determination of such a threshold a difficult task.

Figure 2: (a) A sample parsing network and (b) the tree structure it represents.

Although the space of frequency-based explanations to be explored is still very large, we fear that a more complex probabilistic model would not achieve a good estimation of the parameters, as it would encounter too sparse data. In the following section, we propose an account of the differential difficulty of reduced relatives based on independently justified architectural restrictions.

4 The Processing Account

Instead of a pure statistical model, we suggest that two influences affect the interpretation of reduced relative clauses: structural complexity and frequency factors. We present a processing model whose architectural restrictions, coupled with the lexical syntactic analysis presented in Section 2, derive the extreme difficulty of reduced relatives with unergative verbs. In fact, the very hard sentences are simply *unprocessable* in this approach, independent of frequency information. However, frequency *can* influence the resolution of the ambiguity in the case of easier reduced relatives.

4.1 A Brief Overview of the Parser

The competitive attachment model is a hybrid connectionist parser that dynamically builds a parse tree by activating processing nodes that represent syntactic phrases (Stevenson, 1994b; Stevenson, 1994a). Attachment nodes are established between two phrasal nodes that are potential *sisters* in the parse tree; see Figure 2. Attachments among individual phrases are determined through distributed processing, forming a network that directly represents syntactic structures. Competition for numeric activation is the crucial mechanism for resolving ambiguities. For an ambiguous input token, multiple phrases are activated, each in proportion to the strength of the corresponding lexical entry. These phrases make the best possible attachments to the current parse tree, and the resulting structural alternatives compete with each other.

In order to achieve uniformity of numeric processing, all phrasal nodes posit an attachment to an *empty node*, so that there is an attachment node to activate even when no overt phrase attaches to that site; see Figure 3. An empty node may indicate lack of an attachment at that location, or it may alternatively represent a trace of a movement operation.

Figure 3: Empty nodes are activated in conjunction with every phrase, at each attachment site.

Because an empty node has no evidence from the input, however, it cannot have articulated structure—it is a single, "dummy" phrasal node. This restriction is necessary to curb the potential explosion of nested empty structure, which could quickly exhaust the available processing nodes in the parser.

4.2 Parsing Reduced Relatives

The restriction on top-down hypothesizing of empty structure plays a crucial role in determining the ease or difficulty of reduced relatives in the competitive attachment parser. Figure 4 shows the syntactic structure of an unergative RR, according to the analysis of Stevenson and Merlo (1995). The causative form of *raced* takes a VP complement (corresponding to the "base" intransitive form of the verb); compare Figure 1. In the RR, this VP requires an empty head (the trace of *raced*) and an empty specifier (the trace of *the horse*). Thus, the VP has empty articulated structure, with no overt elements. This is precisely the type of empty phrase that the parser cannot activate. Figure 5 shows the three network structures that are partially activated in response to the initial input *The horse raced*. The empty node that is sister to *raced* in the RR structure is unable to support a grammatical analysis of the input; compare the VPs in the reduced relatives of Figures 4 and 5.

It is crucial to note that the parser is simply unable to activate the structure needed for a grammatical analysis of this sentence. The parser can only activate a single empty node that stands in the position of the VP complement. The parser *cannot* activate a full projection for the empty embedded VP, which would be necessary to establish the correct empty nodes for the traces of the NP and V movements. The parser thus captures the overwhelming preference for the main verb interpretation of the unergative participle, and the severe difficulty people experience in interpreting it as a reduced relative.

In contrast, the easy reduced relatives, with unaccusative and ordinary verbs, do not face this difficulty. For example, Figure 6 shows the syntactic structure of a RR with the verb *melted*. Because the causativization of *melt* occurs in the lexicon, there is no embedded VP in the syntactic structure of the VP for *melted*. The parser establishes the alternative structures shown in Figure 7 at the point of processing *The butter melted*. Note that here all of the single empty nodes in the RR structure are sufficient for a grammatical analysis of the construction. Competition for the RR interpretation arises from the two main verb structures, but there is no garden path effect. Instead, the preferences of the parser derive from activation differences that arise from contextual influences such as lexical frequency information. Ordinary (optionally intransitive) verbs have the same structure for the RR as the unaccusative verbs, and are therefore also relatively easy to process in the competitive

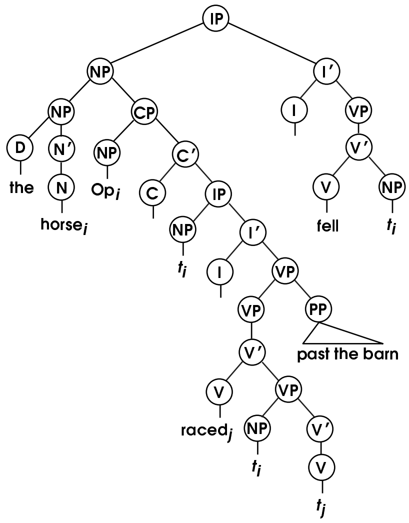

Figure 4: The syntactic structure for *The horse raced past the barn fell.*

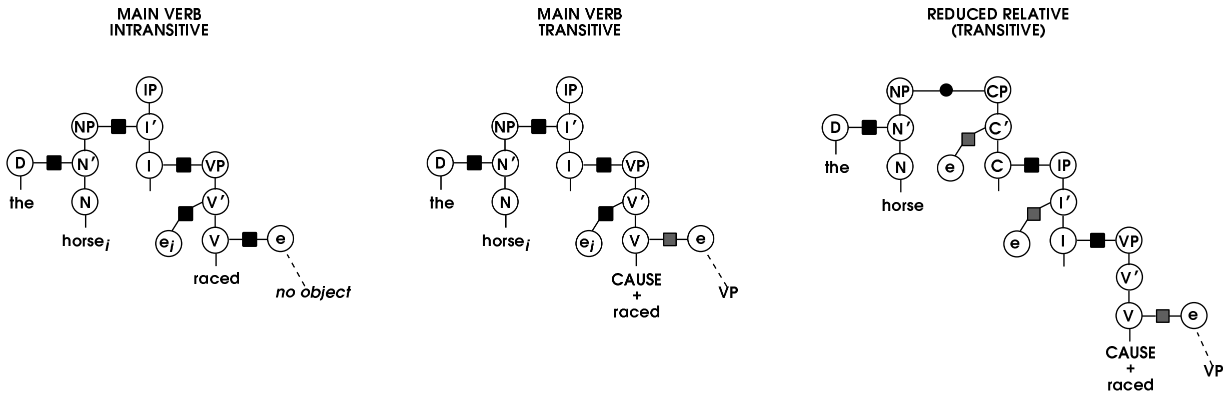

Figure 5: The structures competing for activation at the point of processing *raced* in *The horse raced past the barn fell.*

attachment parser.

5 Conclusions and Future Work

The primary goal in most NLU systems has been to rank preferences for ambiguous input in ways that match the intuitions of human judges. The competitive attachment model can integrate frequency information during the parse, but it is also sensitive to structural representations, allowing it to better match human preferences. We can account for the severe difficulty of reduced relatives with unergative verbs, as well as for the ease of interpretation of reduced relatives with verbs of any other class that support a transitive alternative. A frequency-based account mirroring the corpus results would have ranked both unergatives and unaccusatives as strongly dispreferred, incorrectly. Moreover, such an account would

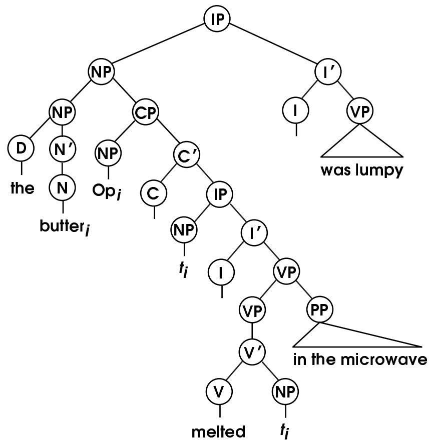

Figure 6: The syntactic structure for *The butter melted in the microwave was lumpy.*

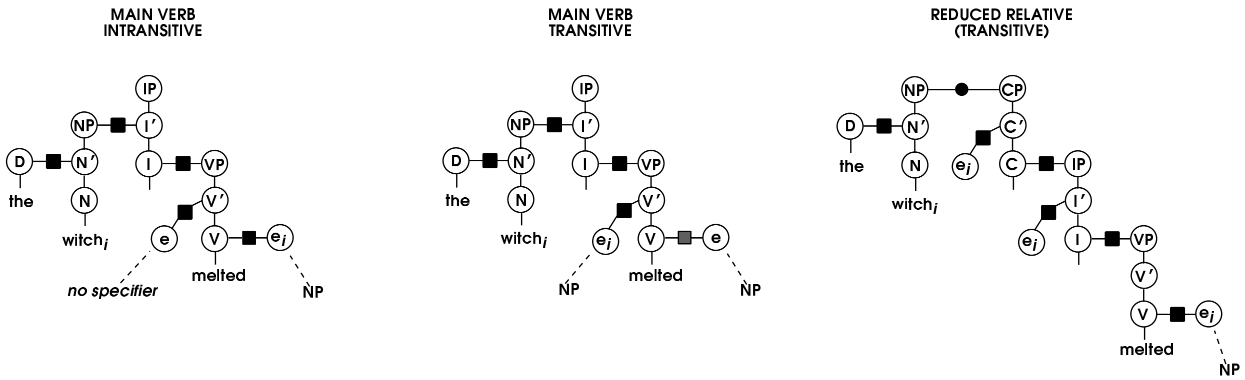

Figure 7: The structures competing for activation at the point of processing *melted* in *The butter melted in the microwave was lumpy.*

have no explanation for the difficulty of unambiguous cases. Therefore, the competititive attachment model reconciles corpus counts and on-line data, because frequency can influence certain cases in which the parsing competition is closer, but it cannot rescue hard cases that are difficult because of structural reasons.

The current work suggests the need for experiments that manipulate a three-valued variable for verb type, to establish the relation between on-line behavioral data on one-hand, and linguistic representation and corpus counts on the other hand. Moreover, an explanation for the observed fact that unaccusatives are rarer than ordinary verbs in the RR, passive constructions and in the transitive voice, is needed, given that their interpretation does not appear to cause difficulty. A difference between the two types of verbs is that an unaccusative has two possible main verb structures, projected from its causative and non-causative lexical entries, while the lexical entry for an ordinary verb projects a single main verb structure, with an optional complement. Future work will focus on the precise characterization of preference differences between these two types of verb which may arise from their difference in lexical semantic analysis.

References

Bever, T. G. (1970). The cognitive basis for linguistic structure. In J. R. Hayes (Ed.), *Cognition and the Development of Language*. New York: John Wiley.

Bod, R. (1995). *Enriching Linguistics with Statistics*. Ph. D. thesis, University of Amsterdam.

Brill, E. and P. Resnik (1994). A rule-based approach to prepositional phrase attachment disambiguation. In *Proceedings of the 15th International Conference on Computational Linguistics (COLING-94)*.

Brousseau, A.-M. and E. Ritter (1991). A non-unified analysis of agentive verbs. In *West Coast Conference on Formal Linguistics*, Number 20, pp. 53–64.

Frazier, L. and K. Rayner (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology 14*, 178–210.

Gibson, E., C. Schutze, and A. Salomon (1995). The relationship between the frequency and the complexity of linguistic structure. In *8th Annual CUNY Conference on Human Sentence Processing*. To appear in *J. of Psycholinguistic Res.*.

Hale, K. and J. Keyser (1993). On argument structure and the lexical representation of syntactic relations. In K. Hale and J. Keyser (Eds.), *The View from Building 20*, pp. 53–110. MIT Press.

Hindle, D. and M. Rooth (1993). Structural ambiguity and lexical relations. *Computational Linguistics 19*(1), 103–120.

Levin, B. and M. Rappaport Hovav (1995). *Unaccusativity*. Cambridge, MA: MIT Press.

MacDonald, M., N. Pearlmutter, and M. Seidenberg (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review 101*(4), 676–703.

MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes 9*(2), 157–201.

Merlo, P. (1994). A corpus-based analysis of verb continuation frequencies for syntactic processing. *Journal of Psycholinguistic Research 23*(6), 435–457.

Stevenson, S. (1994a). Competition and recency in a hybrid network model of syntactic disambiguation. *Journal of Psycholinguistic Research 23*(4), 295–322.

Stevenson, S. (1994b). *A Competitive Attachment Model for Resolving Syntactic Ambiguities in Natural Language Parsing*. Ph. D. thesis, Computer Science, University of Maryland.

Stevenson, S. and P. Merlo (1995). Lexical structure and processing complexity. Submitted for publication.

Trueswell, J. C. (1995). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language*. In press.

Trueswell, J. C., M. K. Tanenhaus, and S. M. Garnsey (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language 33*, 285–318.

Weischedel, R., M. Meteer, R. Schwartz, L. Ramshaw, and J. Palmucci (1993). Coping with ambiguity and unknown words through probabilistic models. *Computational Linguistics 19*(2), 359–382.