Surprisal Derives the Recent Filler Heuristic in Mildly Context Sensitive Grammars

Kyle Grove

Linguistics

Cornell University

211 Morrill Hall

Ithaca, NY 14853, USA

`kwg33@cornell.edu`

Abstract

This paper provides a new account for why online processing of filler-gap relative clause dependencies is more difficult in cases where filler-gap interacts with object control than in cases involving subject control, as reported by Frazier et al. (1983). Frazier et al. (1983) argued for a Recent Filler heuristic in which the parser expects to discharge the most recent filler at every gap site. We observe that statistical subcategorization preferences on the control verb and the embedded verb 'sing' interact, favoring subject control disambiguation.

We employ surprisal (Hale, 2001) as a complexity metric on filler-gap structures by construing control as a Movement operation in Minimalist Grammars (Stabler, 1997). We obtain greater surprisals for the Distant Filler condition, deriving the prediction that the Recent Filler heuristic falls out from statistical subcategorization preferences.

1 Introduction

In filler-gap constructions, [1] the parser must map a non-local dependency between an extraposed element (the 'filler' nominal element) and a position in the subcategorization frame (the 'gap') of a verb to-be-determined. When filler-gap sentences contain complex verb structure such as control verbs, ambiguity can result because there are multiple potential gap sites the filler could conceivably have been extraposed from. The complexity of this task

is demonstrated in 1, where a single prefix (1a. and 1b.) can be disambiguated in two quite different ways.

(1) a. Everyone liked the woman the little child begged to sing those stupid French songs.

   b. Everyone liked the woman the little child begged to sing those stupid French songs for.

(Frazier et al., 1983, 203)

(2) a. The child$_1$ begged the woman$_2$ to t$_2$ sing those songs.

   b. The child$_1$ begged to t$_1$ sing those songs for the woman.

With object control (OC) verbs (1a and 2a), the subject of the infinitival 'sing' is the object of the transitive control verb 'begged', but with subject control (SC) verbs (1b and 2b), the subject of the infinitival is the subject of an intransitive control verb. While these conditions can generally be disambiguated at the control verb in non-relativized contexts (2), relativization of the control verb's object removes the disambiguating cue in object control cases. Thus, the prefix 'Everyone loved the woman the child begged to sing those ...' is locally ambiguous, and is disambiguated by either the continuation (... *for.*), in the subject control case, or the sentence abruptly ending (.), in the object control case.

Frazier et al. (1983) argued that the lesser difficulty of SC filler-gap supports an account where a strictly serial parsing strategy is guided by a Recent Filler heuristic. The Recent Filler heuristic holds that in

cases where the parser is presented with multiple fillers for a single gap, it expects to map the most recent filler to that gap. When this expectation is defeated by new sentential material, the parser must backtrack until it can recover to an analysis congruent with this material.

The Recent Filler heuristic was disconfirmed in Boland et al. (1990). Boland et al. (1990) employed a semantic mismatch paradigm for object control verbs in which the raiser-to-object is potentially an implausible subject of the embedded infinitival. This effect is seen in 3, where 3b is semantically anomalous due only to the infinitival verb; horses and outlaws both make excellent receivers of signals, but horses lack the ability to surrender weapons.

(3) a. The cowboy signaled the outlaw to surrender his weapons quietly.

b. The cowboy signaled the horse to surrender his weapons quietly.

(Boland et al., 1990, 416)

By applying semantic mismatch to filler-gap dependencies, an extraposed element could be manipulated for semantic plausibility as a potential filler for a gap site. Boland et al. (1990) developed Distant Filler sentences which featured such plausability mismatches, using wh-questions. In these sentences, the recent filler matched the gap-site verb for plausability, but the distant filler did not.

(4) a. Which outlaw did the cowboy signal to surrender his weapons quietly?

b. Which horse did the cowboy signal to surrender his weapons quietly?

(Boland et al., 1990, 417)

Boland et al. (1990) used an online plausability monitoring task in which participants where asked to incrementally indicate whether the sentence was plausible. If participants employed the Recent Filler heuristic, then they should be unaware of plausability mismatches which obtain only on the Distant Filler structural analysis. Participants detected implausibility immediately, suggesting that they do not rely on a Recent Filler heuristic.

2 Hypothesis

We argue contra Frazier et al. (1983) that the Recent Filler effect arises from the parser's statistical knowledge of verb subcategorization. We observe (in Fig. 14, Appendix) that the Recent Filler analysis requires a prepositional attachment for the verb 'sing' and a subject control frame for the control verb, while the Distant Filler analysis requires there to be no PP attachment on 'sing', and an object control frame for the control verb. We anticipate the verb 'sing', used throughout the materials in Frazier et al. (1983), to exhibit frequent prepositional phrase attachment. We hypothesize that corpus-derived probabilistic weightings on PP-attachment and control verb subcategorization frames are incorporated in a probabilistic grammar that biases towards the Recent Filler analysis. As sentences are parsed, the probabilities in this grammar represent degrees of belief which are prone to revision as evidence is collected; dramatic revisions of belief suggest that the parse has been particularly difficult. We employ surprisal (Hale, 2001) as a psycholinguistic linking theory which characterizes the severity of this belief update. We predict that object control Distant Filler continuations will have greater surprisal at the disambiguating continuation than subject control Recent Filler continuations.

3 Background

3.1 Grammar

How best to represent the argument sharing that control and raising verbs exhibit has generated debate among syntacticians and semanticists (Jacobson, 1992; Chomsky and Lasnik, 1993; Steedman, 1994; Hornstein, 2000). Chomsky and Lasnik (1993) represents the external argument of the infinitival verb as being satisfied by a phonetically unrealized pronominal, PRO, which is to be mapped thematically to the corresponding argument in the control verb by a structural construal operation external to the core grammar. Hornstein (2000) argues instead that the Movement operation available to the core grammar is responsible for this mapping. Jacobson (1992) focuses more on the subcategorization frames of the verbs than the arguments them-

selves. On an approach such as in Jacobson (1992), a control or raising verb acts as a functor on the infinitival verb, mapping the arguments of the infinitival onto the matrix verb.

We constructed a mildly context sensitive Minimalist Grammar (MG) (Stabler, 1997) that implements a non-pronominal theory of control (Jacobson, 1992; Pollard and Sag, 1994; Steedman, 1994; Hornstein, 2000) as a type of Movement operation between the subject position of the embedded verb and an object or subject position in the control verb. We leave as an empirical question whether other approaches to control would yield the same human sentence processing prediction. Rendering control as Move in MG is advantageous in several ways. First, properties of the Minimalist Grammars formalism have been studied in depth: sentences in an MG language can be parsed efficiently and parses in MG possess a context-free backbone which can be leveraged with PCFG methods. Second, since relative clause extraposition is also treated as Move in our grammar, we can model the complexity of the interaction between filler-gap and control more easily than we could if control were handled in some other component of the grammar.

In MGs, a lexical item has a distinct set of syntactic (SYN) features, which uniquely determine a movement chain that the lexical item can participate in (Hale and Stabler, 2005). Thus, we model lexical subcategorization frames in MGs by a one-to-many encoding between phonetic (PHON) features and SYN features: an ambiguous control verb such as 'wanted' is modeled in an MG lexicon with redundant lexical entries, as shown in Fig. 1.

::=>V =D v   to :: =v inf
wanted :: =inf +k v   wanted :: =inf +k V
the :: =N D -k   the :: =N D
woman :: N   child :: N

Figure 1: MG Fragment for Control

Derived Trees for Subject and Object Control are depicted in Figs. 2 and 3.

In both subject and object control verbs, the control dependency is triggered by the Merge of a nominal category whose SYN is D -k in the specifier of
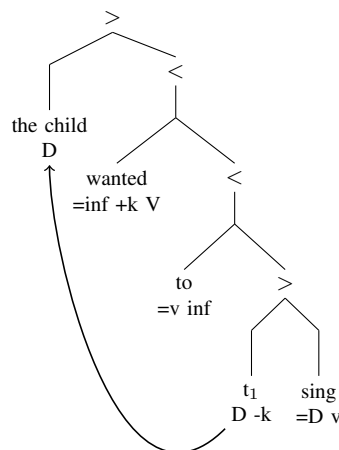

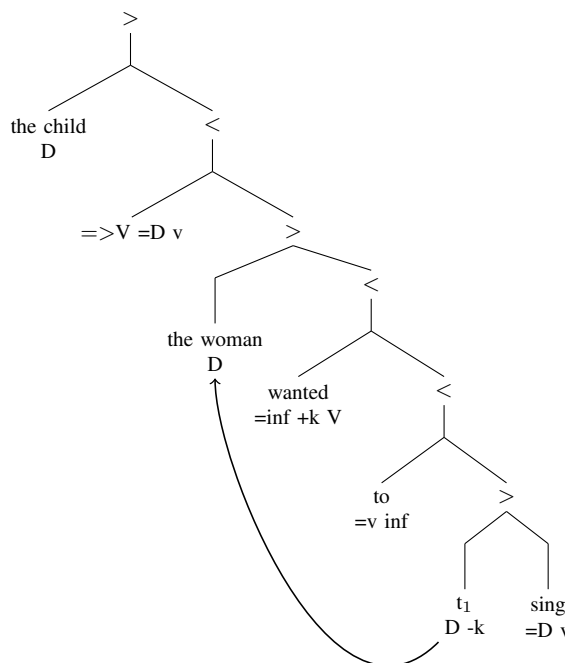*Figure 2:* Derived Tree for Subject Control

Figure 3: Derived Tree for Object Control

the infinitival verb. The landing site of this nominal category is determined by the location of the corresponding +k attractor feature. When the control verb has SYN =inf +k v as in Fig. 2, the nominal with -k feature will move to the specifier of 'little v', the subject position of the control verb. Subject-to-subject movement derives the semantic intuition

that in 'The child asked to sing', the child has two semantic roles: it is both an asker and a (potential) singer. However, when the control verb has SYN =inf +k V, such in Fig. 2, the subject of the embedded verb raises to the object position of the matrix verb, specifier of V. Raising-to-object derives the semantic intuition that in 'The child asked the woman to sing', the woman is a potential singer but the recipient of the request; the child is the asker.

The object of an object control verb could undergo further extraposition via relativization. The interaction of an ambiguous control verb and relativization renders thematic assignment difficult because the extraposed object could either have moved from the object of the control verb or from another position in the structure. In Figs. 4 and 5, we demonstrate a promotion analysis of relative clauses following Kayne (1994) and Bianchi (2004). On this analysis, the relative pronoun and relativized nominal are a constituent underlyingly. As seen in Fig. 5, the =N feature on relative pronoun 'who' triggers Merge with the N category feature on woman, yielding a category whose PHON is 'who woman' and syntactic feature inventory is D -arel. This D is Merged into an argument position as the embedded verb's argument structure is composed with Merge operations up to the head '=T +arel Crel'. The Merge of '=T +arel Crel' with the derived category 'T -arel' establishes an embedded clause with syntactic features '+arel Crel -arel'. The '+arel/-arel' feature pair cancels by triggering Movement of the -arel feature, extracting 'who woman' out of the embedded clause as indicated by the lower arrow in Fig. 5. Merge of '=Crel +nom agrD' ultimately results in the remnant movement of 'woman', as indicated by the higher arrow in Fig. 5. This series of movement correctly derives the ordering of subject and object relative clauses in English. This strategy treats reduced relative clauses by substituting the null relative pronoun '::=N D -arel' in Fig. 4 for its non-null counterpart 'who :: = N D -arel'.

the :: =>agrD D	:: =Crel +nom agrD
:: =T +arel Crel	=v T
who :: =N D -arel	:: =N D -arel
woman :: N	sing :: =D v

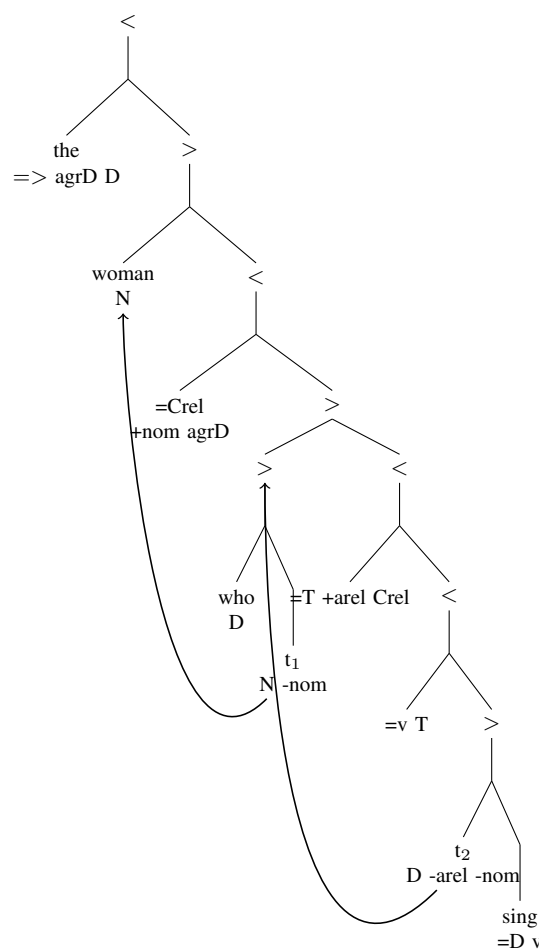*Figure 4:* MG Fragment for Relative Clause

Figure 5: Derived Tree for Relative Clause

3.2 Surprisal

Following Hale (2001), we employ information theory to model the strangeness of parser actions. This study pursues the hypothesis that information about verbal subcategorization frequencies biases the human sentence processor against object control (Distant Filler) filler-gap resolutions of the ambiguous prefix. We predict that a parser with this stochastic grammatical knowledge will model this expectancy by exhibiting greater surprisal on object control/Distant Filler continuations than on subject control/Recent Filler continuations.

Surprisal (Hale, 2001) hypothesizes that perceived difficulty of human sentence processing at a token

of interest is associated with the unexpectedness of the new token. On a given string, the surprisal of a token situated between positions *n-1* and *n* is the logarithm of the ratio of the probabilities of prefixes starting at 0 and ending at *n-1* and *n*.

$$\text{surprisal} = \log_2 \frac{\alpha_{n-1}}{\alpha_n}$$

Figure 6: Surprisal of a word given a PCFG

Surprisal formalizes the intuition that some words are syntactically costly to incorporate, by measuring the rate at which those words reduce the total probability allocated to all incrementally viable analyses. Surprisal predicts garden pathing when new tokens rule out much probability mass.

4 Methodology

We used Tregex (Levy and Andrew, 2006) on Penn Treebank (Marcus et al., 1994) to obtain counts for object control and subject control for each control verb used in Experiment 2 of Frazier et al. (1983), using the queries indicated in Fig. 7. We considered a verb frame as an instance of subject control when the verb node was sister to an S node which dominated a null pronominal subject (NP-SBJ) whose annotation was either PRO or -NONE-. We considered a verb frame as an instance of object control when the verb frame was sister to an S-node whose NP-SBJ had non-null string yield. We excluded instances of passivized object control, which structurally resemble cases of subject control.

Subject Control
VP<(/VB.?/<(/expect.[s|ed|ing]/)
\$+NP \$(S<(NP-SBJ<-NONE-\$(VP<TO))))
VP < (/VB.?/ < (/expect.[s|ed|ing]/)
\$+NP \$(S<(NP-SBJ<PRO\$(VP<TO))))
Object Control
VP < (/VB.?/ < (/expect.[s|ed|ing/])
\$(S<(NP-SBJ<!-NONE-\$(VP<TO))))

Figure 7: Tregex Queries for Control Verbs

We obtained the following counts in Fig. 8 for the four verbs in Experiment 2 of Frazier et al. (1983), verifying that subject control is prevalent in the Penn Treebank.

Verb	Subject Control	Object Control
want	344	47
expect	509	200
choose	23	1
ask	16	35

Figure 8: Corpus Counts for Control Verbs

We also obtained counts for prepositional phrase attachment preferences for the verb 'sing', as shown in Fig. 9, also verifying that PP-attachment is particularly frequent for the verb 'sing'.

Verb	PP	¬ PP
sing	7	6

Figure 9: Corpus Counts for PP Attachment

We also obtained counts for reduced relative constructions and main clause constructions, as well as counts for the transitivity of 'sing'. These factors were common across all conditions. For each of the factors, we constructed a parameter by converting the count into a ratio, through dividing the individual outcome's count by the summed count of all possible outcomes. We built a representative minitreebank of $4 * 2^4 = 64$ sentences, where each sentence contained: either a subject or object control verb; in either a main clause or reduced relative clause usage; with or without a prepositional phrase attachment to the verb; and with either a transitive or intransitive use of the verb 'sing', for each of the four control verbs. Each sentence was weighted with the product of the parameters particular to that condition.

An MG statistical prefix parsing system was used to obtain surprisals for the ambiguous prefix common to 1a and 1b and the subject and object control continuations. For each prefix, a parse forest is built; this parse forest is equivalent to a context-free grammar (Billot and Lang, 1989), which can be augmented with probabilities to obtain a probabilistic context free grammar. At training time, the parser uses Weighted Relative Frequency Estimation (Chi, 1999) to estimate a PCFG model of the minitreebank. At testing time, the parser constructs a probabilistic model at each prefix; these prefixes are represented as straightline finite state automata whose suffixes are self loops. The parser estimates for each

$$_0\text{v}_6 \rightarrow {}_0\text{D}_2 \; {}_2 = \text{D} \; \text{v}_6$$

Figure 10: A situated MG rule.

prefix automaton a weighted intersection PCFG using the renormalization technique in Nederhof and Satta (2006). In this intersection PCFG, each category is 'situated' with indices, i.e. the situated category is the product of grammatical categories and automaton transitions used in its derivation.

5 Results

We found greater surprisals on distant filler continuations for three of the four verbs.

Verb	Recent Filler	Distant Filler
wanted	1.448 bits	2.978 bits
expected	1.819 bits	2.176 bits
asked	2.392 bits	1.608 bits
chose	1.213 bits	4.241 bits

Figure 11: Surprisal Results

These compare to the Frazier et al. (1983) results, presented below.

Verb	Recent Filler	Distant Filler
wanted	980 msec.	1168 msec.
expected	997 msec.	1082 msec.
asked	969 msec.	1132 msec.
chose	915 msec.	1050 msec.

Figure 12: Frazier Mean Reaction Times

The surprisal data suggest that no special heuristic is required to explain the Recent Filler expectation in Frazier et al. (1983); the expectation arises naturally on a probabilistic grammar which has knowledge of subcategorization frequencies for verbs. The Recent Filler analysis of the ambiguous prefix requires that 1) the verb 'sing' exhibit PP-attachment; 2) the control verb be a subject control verb. The Distant Filler analysis requires the conjunction of two lower probability events: 1) that the verb 'sing' exhibit no PP-attachment; 2) that the control verb be an object control verb. The Distant Filler continuations yield greater surprisals because the parser

must segue rapidly from a highly probable parse forest which is uncommitted to verb information to a much less probable parse forest which is committed to generally unlikely beliefs about the matrix and embedded verb.

6 Discussion

Our claim that the surprisal results reflect rapid shifts in parser beliefs about control and PP-attachment is borne out by examining the parse forest conditioned on each prefix. We examined the parse forest conditioned on the ambiguous prefix (of the 'want' condition) and found an MG category which was parent of two different MG rules; one which reflected the Recent Filler/Subject Control strategy and one which reflected the Distant Filler/Object Control strategy. We depict these rules in Fig. 13.

0.766(: v -arel -nom)→(: +k v -arel -nom -k)
0.234(: v -arel -nom)→(: =D v -arel -nom) (:D)

Figure 13: Probabilistic MG Rules for Recent and Distant Filler

The derived MG category (v -arel -nom) is simply a verbal category with features that license relative clause extraction. In the Recent Filler rewrite, this category is formed from a unary Move application which moves 'the child' to derive Subject Control via the (+k,-k) case feature pair. In the Distant Filler rewrite, (v -arel -nom) is formed from the binary Merge application where 'the child' is merged as the subject of the control verb; 'woman' has already moved for Object Control. The probabilities [2] attached to these rules show that the Recent Filler bias falls out from statistical verb subcategorization information; the parser predicts that the Recent Filler continuation is three times as likely as the Distant Filler continuation.

We built a Minimalist Grammar which treats Control as Move so that we could easily model the interactions of control and filler-gap. Future work would look to explore whether other approaches to control verbs would yield similar surprisal results. We could

---

[2]Importantly, all the categories below this production in the 'branch' have deterministic rewrites with probability 1.

for instance operationalize Jacobson (1992) syntactically with MG adjunction; we would allow the control verb to incorporate the infinitival verb directly, rather than using Move to share nominal arguments between them. Alternatively, we could develop a system that treats control using a null pronominal utilizing Conjunctive Grammars (Okhotin, 2001) to simulate the seperate Base and Control modules of a Government and Binding style grammar.

We derived Frazier et al. (1983)'s Recent Filler preference as an epiphenomenon of a statistical parser's knowledge of verb subcategorizations. The embedded verb 'sing' exhibits an affinity for PP-attachment which together with knowledge of sub-categorization rates of control verbs directly gives rise to the effect in Frazier et al. (1983). Our model showed that the Recent Filler effect is more likely due to the rapid integration of verbal subcategorization frames which provide rich information about the structural environment.

References

Valentina Bianchi. 2004. Resumptive relatives and lf chains. In Luigi Rizzi, editor, The Structure of CP and IP, pages 76–114. Oxford University Press, Oxford.

Sylvie Billot and Bernard Lang. 1989. The structure of shared forests in ambiguous parsing. In Proceedings of the 27th annual meeting on Association for Computational Linguistics, pages 143–151. Association for Computational Linguistics Morristown, NJ, USA.

Julie E. Boland, Michael K. Tanenhaus, and Susan M. Garnsey. 1990. Evidence for the immediate use of verb control information in sentence processing. Journal of Memory and Language, 29(4):413–432.

Zhiyi Chi. 1999. Statistical properties of probabilistic context-free grammars. Computational Linguistics, 25(1):131–160.

Noam Chomsky and Howard Lasnik. 1993. The theory of principles and parameters. Syntax: An international handbook of contemporary research, 1:506–569.

Lyn Frazier, Charles Clifton, and Janet Randall. 1983. Filling gaps: Decision principles and structure in sentence comprehension. Cognition, 13(2):187–222.

John T. Hale and Edward P. Stabler. 2005. Strict deterministic aspects of minimalist grammars. Logical aspects of computational linguistics, pages 162–176.

John T. Hale. 2001. A probabilistic Earley parser as a psycholinguistic model. In Proceedings of NAACL, volume 2, pages 159–166.

Norbert Hornstein. 2000. Move! A Minimalist Theory of Construal. Blackwell, Oxford.

Pauline Jacobson. 1992. Raising without movement. 48:149–194.

Richard S. Kayne. 1994. The Antisymmetry of Syntax. MIT Press.

Roger Levy and Galen Andrew. 2006. Tregex and Tsurgeon: tools for querying and manipulating tree data structures. In Proceedings of the Fifth International Conference on Language Resources and Evaluation. Citeseer.

Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1994. Building a large annotated corpus of English: The Penn Treebank. Computational linguistics, 19(2):313–330.

Mark-Jan Nederhof and Giorgio Satta. 2006. Estimation of consistent probabilistic context-free grammars. In Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, pages 343–350. Association for Computational Linguistics Morristown, NJ, USA.

A. Okhotin. 2001. Conjunctive grammars. Journal of Automata, Languages and Combinatorics, 6(4):519–535.

Carl J. Pollard and Ivan A. Sag. 1994. Head-driven phrase structure grammar. University of Chicago Press.

Edward P. Stabler. 1997. Derivational minimalism. In Logical Aspects of Computational Linguistics., pages 68–95. Springer.

Mark Steedman. 1994. Binding and Control in CCG and its Relatives. IRCS Technical Reports Series, page 161.

7 Appendix

Object Control/Distant Filler

Everyone loved   the   woman   the girl begged GAP to GAP   sing those ... songs.

=Crel +nom agrD   =inf +k V

surprisal = 2.75 bits

Subject Control/Recent Filler

Everyone loved   the   woman   the girl begged   to GAP   sing those ... songs   for GAP.

=Crel +nom agrD

=inf +k v

surprisal = 1.72 bits

Relativization (A') Movement
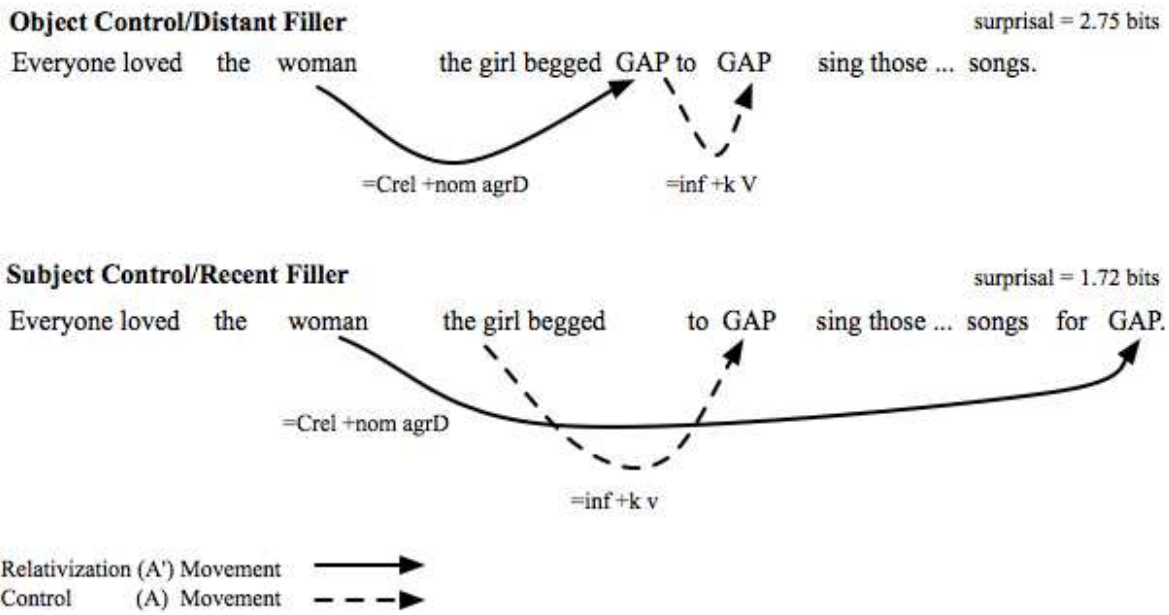Control   (A) Movement

Figure 14: Online Processing of Control and Relativization with Mean Surprisals