Complete feature list for phrasal predictability models

		FEATURE TYPE				
		NGRAM	HEAD	HIST	PRENOM	EXT
1.	Immediately preceding word trigram, bigram, and unigram	•				
2.	Constituent history: list of syntactic categories of currently expanded tree node			•		
3.	Node distance: number of sisters appearing to right of head daughter of currently expanded node			•		
4.	Head word of currently expanded node		•			
5.	Head (part-of-speech) tag of currently expanded node		•			
6.	Conjunction of node distance and head word		•	•		
7.	Conjunction of node distance and head tag		•	•		
8.	Conjunction of constituent history and head word		•	•		
9.	Conjunction of constituent history and head tag		•	•		
10.	Word distance: number of words appearing to right of head daughter			•		
11.	Conjunction of word distance and head word		•	•		
12.	Conjunction of word distance and head tag		•	•		
13.	Animacy of currently expanded NP		•			
14.	Conjunction of animacy and constituent history		•	•		
15.	Final determiner preceding head word of NP				•	
16.	Conjunction of constituent history and final determiner			•	•	
17.	Conjunction of head word and final determiner		•		•	
18.	Whether the head noun has a prenominal possessor				•	
19.	Prenominal adjective modifying head noun (recorded for every such prenominal adjective)				•	
20.	Conjunction of prenominal adjective and head word		•		•	
21.	Conjunction of prenominal adjective, final determiner, and head word		•		•	
22.	Conjunction of prenominal adjective and constituent history			•	•	
23.	External grammatical function of currently expanded NP, defined as the conjunction of the syntactic category of the NP's mother and whether the NP's head sister has appeared yet					•
24.	Matrix-clause verb (if any), defined as the word heading lowest S or VP node ancestor of the currently expanded NP, so long as no SBAR or SBARQ node intervenes					•
25.	Part-of-speech tag of the matrix-clause verb					•

Features 1–9 are used for all nodes; feature 10-25 are used only in the expansion of NP nodes. All features except type 1 were conjoined with the syntactic category of the currently expanded node.

Control factors for syntactic reduction models

There are twenty control factors and one control interaction in the model. They are summarized in . Together the controls constitute 27 parameters. A priori, this should be a reasonable number of factors since the database contains approximately 1,400 cases of the less frequent outcome (RC*s with a relativizer), which should be enough for up to 140 parameters in the model.

Factor name	**Description**	**Type**
L(PRECEDING)	Log length of material preceding modified NP	continuous
L(ADJACENCY)	Log length of material between head and RC*	continuous
L(RC*)	Log length of RC* in words	continuous
GAP-EMBEDDING	Gap part of a clausal complement in RC*?	categorical (2)
GAP-EMBEDDING x L(RC*)		continuous (2)
SUBJECT FORM	Accessibility coding of RC* subject	categorical (3)
SUBJECT ANIMACY	Animacy of RC* subject	categorical (3)
P(PRECEDING WORD)	Log. frequency of word preceding RC*	continuous
P(INITIAL WORD)	Log. frequency of word after relativizer site	continuous
MATRIX NEGATION	Is the matrix clause negated?	categorical (2)
MATRIX VERB	Type of matrix verb	categorical (4)
PRECEDING DISFLUENCY	Disfluencies between head and RC*	count
PRECEDING PAUSE	Pause immediately preceding RC*?	categorical (2)
INITIAL DISFLUENCY	Disfluencies in RC* before/in subject	count
LATER DISFLUENCY	Disfluencies in remainder of RC*	count
LOG SPEECH RATE	Log. speech rate around RC* start	continuous
SQ LOG SPEECH RATE	Squared log. speech rate around RC* start	continuous
WITHIN-SPEAKER PERSIST.	Preceding within-speaker prime (if any)	categorical (3)
ACROSS-SPEAKER PERSIST.	Preceding across-speaker prime (if any)	categorical (3)
OCP PRECEDING PHON	Preceding phon [+fricative] or [+dental]?	categorical (2)
SPEAKER GENDER	Gender of speaker	categorical (2)
	Total control parameters in full model	27