

Structural and non-structural expectations in Chinese sentence comprehension

A growing body of psycholinguistic research suggests that human sentence comprehension is predictive. When we read incrementally, each incoming word introduces information that helps us to shape expectations about the rest of the sentence (Hale, 2001; Levy, 2008). The information that a new word carries can be structural, e.g. grammatical category, hierarchical phrase structure, or syntactic movement. It often revises existing parses of the sentence. In addition, a new word renders non-structural information, such as thematic relation, lexical semantics and information structure. The present work examines the inclusion of additional non-structural features in a word-by-word model of reading difficulty. In particular, it presents a computational modeling of Chinese relative clause (RC) comprehension that involves both formal grammar fragments and the non-structural feature animacy. The predictions faithfully reflect the reported animacy effect at the head noun (Wu et al. 2012).

The animacy feature has been found to play a facilitating role in sentence comprehension (Traxler et al., 2002, 2005). In a Chinese study, Wu et al. (2012) investigated how the frequency distribution of animacy would impact the processing of Chinese RCs with different animacy manipulations in examples (1) and (2) below. The results indicate that animacy serves as an important cue for thematic role assignment and affects how ambiguities are resolved. In particular, Wu and colleagues found a facilitation animacy effect at the post-RC head noun position: subject RCs with animate heads (1a and 1b) and Object RCs with inanimate heads (2c and 2d) were easier than their counterparts respectively. This result mirrors the finding in corpus analyses and the observation in cognitive psychology.

- (1) a. SR; Head: +anim; NP within RC: +anim
raokai baoan de jizhe ...
bypass guard DE reporter ...
'the reporter who bypassed the guard ...'
- b. SR; Head: +anim; NP within RC: –anim
raokai damen de jizhe ...
bypass gate DE reporter ...
'the reporter who bypassed the gate ...'
- c. SR; Head: –anim; NP within RC: +anim
zazhong baoan de jidan ...
smash guard DE egg ...
'the egg which smashed into the guard ...'
- d. SR; Head: –anim; NP within RC: –anim
zazhong damen de jidan ...
smash gate DE egg ...
'the egg which smashed into the gate ...'
- (2) a. OR, Head: +anim, NP within RC: +anim
jizhe raokai de baoan ...
reporter bypass DE guard
'the guard who the reporter bypassed ...'
- b. OR, Head: +anim, NP within RC: –anim
jidan zazhong de baoan ...

egg smash DE guard
'the guard who the egg smashed into ...'

c. OR, Head: –anim, NP within RC: +anim
 jizhe raokai de damen ...
 reporter bypass DE gate
'the gate which the reporter bypassed ...'

d. OR, Head: –anim, NP within RC: –anim
 jidan zazhong de damen...
 egg smash DE gate
'the gate which the egg smashed into ...'

In this work, to model the animacy effect in Chinese RC comprehension, we adopted the information-theoretic notion Entropy Reduction (ER, Hale, 2006). ER formalizes the amount of information contributed by a word in reducing structural uncertainties. A higher ER degree often relates to a longer reading time delay during sentence comprehension (Frank, 2013), including head-final RCs in Chinese, Japanese, and Korean (Yun et al, 2015). It yields a quantitative viewpoint on ambiguities that sentence comprehenders face when dealing with prenominal relativized constructions. It also uncovers a range of language-specific factors that all pertain to the distribution on “unchosen” alternatives. The present work extends this modeling framework to include non-structural features, such as animacy.

A Minimalist Grammar (Stabler, 1997) was prepared to cover Chinese RCs and other relevant structures, e.g. DE possessives (“NP de NP”). The grammar fragment is unlexicalized to avoid any bias on word choice. It is also subcategorized in a way that diacritics such as *–Animate* and *–Inanimate* are added to the noun phrase category and subsequently to parent nodes like NPs and VPs. Weighting relevant construction types by treebank attestation counts (CTB 7), including the frequency distribution of noun phrase animacy, allows us to estimate probabilistic “intersection” grammars conditioned on prefixes (Nederhof & Satta, 2008). This modeling work samples syntactic alternatives from intersection grammars to get an intuitive picture of how uncertainties (both structural and non-structural) are reduced during parsing. The calculated reading difficulty at the RC head is shown in the table below. ER predicts that Subject RCs with animate head and ORs with inanimate head (in bold) are easier at the head noun than their counterparts. These predictions are consistent with the reading data reported by Wu et al (2012). Further investigation also explores linguistically plausible interpretations for the reported animacy effect, including the NP animacy distribution within possessives as well as the interaction between head noun animacy and main verb transitivity.

	Head	NP within RC	ER (bits) at Head		Head	NP within RC	ER (bits) at Head
Subject RCs	+anim	+anim	1.45	Object RCs	+anim	+anim	1.46
	+anim	–anim	1.19		+anim	–anim	1.44
	–anim	+anim	1.47		–anim	+anim	1.28
	–anim	–anim	1.57		–anim	–anim	1.26

In sum, examining contextualized syntactic alternatives along with non-structural feature animacy shows how processing difficulty in Chinese sentence comprehension reflects the uncertainty associated with alternative expectations. By using probabilistic grammars based on corpus counts, this methodology leverages a strong grammar-parser relationship.