

ANIMACY IN SENTENCE PROCESSING ACROSS  
LANGUAGES: AN INFORMATION-THEORETIC  
PROSPECTIVE

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Zhong Chen

August 2014

© 2014 Zhong Chen  
ALL RIGHTS RESERVED

ANIMACY IN SENTENCE PROCESSING ACROSS LANGUAGES: AN  
INFORMATION-THEORETIC PROSPECTIVE

Zhong Chen, Ph.D.

Cornell University August 2014

This dissertation is concerned with different sources of information that affect human sentence comprehension. It focuses on the way that syntactic rules interact with non-syntactic cues in real-time processing. It develops the idea first introduced in the Competition Model of MacWhinney in the late 1980s such that the weight of a linguistic cue varies among languages.

The dissertation addresses this problem from an information-theoretic perspective. The proposed Entropy Reduction metric (Hale, 2003) combines corpus-retrieved attestation frequencies with linguistically-motivated grammars. It derives a processing asymmetry called the Subject Advantage that has been observed across languages (Keenan & Comrie, 1977). The modeling results are consistent with the intuitive structural expectation idea, namely that subject relative clauses, as a frequent structure, are easier to comprehend. However, the present research takes this proposal one step further by illustrating how the comprehension difficulty profile reflects uncertainty over different initial substrings. It highlights particular disambiguation decisions that contribute to processing difficulties found in object relative clauses in English, Italian and Chinese.

More importantly, this dissertation examines the role that the universally applicable animacy cue plays in understanding relativized structures. Based on the frequency distribution of animacy in treebanks, this line of modeling work

not only provides finer-grained explanations for the animacy effect at the head noun reported in previous experimental literature, but also contributes to integrate an important functional notion into the formal linguistic framework in a unique way. Incorporating functional features like animacy allows us to explore rich, cognitively-plausible grammars in human sentence comprehension.

## BIOGRAPHICAL SKETCH

Zhong Chen (陈钟) was born in November, 1984 in Nanjing, Jiangsu, China. He received his Bachelor of Arts degree in Chinese Linguistics in 2006 from Nanjing Normal University, China, with a specialization in Chinese Information Processing. Before entering Cornell University in 2008, he studied at Michigan State University in the Dual Ph.D. program in Linguistics and Computer Science & Engineering. In 2012, he received a Master of Arts degree in Linguistics from Cornell University. Following completion of his Ph.D. in Linguistics with a minor in Cognitive Science, he will work as an Assistant Professor of Chinese at the Rochester Institute of Technology.

To my parents, Weimin Chen and Hong Yuan, who made all of this possible,  
for their unselfish love, encouragement and support.

## ACKNOWLEDGEMENTS

My deepest thanks go to the chair of my guidance committee, Professor John Hale, who has advised me for eight years. John gave me a chance to continue my study on computational linguistics in this great country. When we first met in 2006, I was immediately impressed by the curiosity, passion, and knowledge he brought to his research. As a junior professor, he has been under a lot of pressure. But he creates a comfortable research environment for me, helping me gradually acquire new skills and to shape my own research approach. He would never push me to do any work that I am not interested in. Instead, he always provides me fascinating ideas and encourages me to develop topics that I truly love. He teaches me to be professional, honest and to think critically. He respects our cultural differences but also tells me the truth of US academic system without reservations so that I can start my own career here. I am grateful to have such a wonderful advisor and friend and I can never express enough of my gratitude to him.

I also wish to thank my collaborator and mentor Professor Shravan Vasishth from Potsdam University, Germany. Through working with him, I have learned how to develop a research plan, how to run psycholinguistic experiments, and how to analyze experimental data from scratch. Shravan often sacrifices his rest time at home to have video-meetings with me because of our time difference. He patiently explains everything that I need for the project. He shares his experience in the field with me and encourages me to think differently and thoroughly. I respect that he always sets a strict and high standard for research and makes sure that all details are perfectly taken care of. I wish that I will have more chances to work with him in future years.

The research in this dissertation could never be finished without assistances

and suggestions from my professors, collaborators and colleagues. Guidance committee members Professors Barbara Lust and John Whitman have spent a lot of time helping me develop every stage of this work. The idea of comparing a linguistic cue across languages was motivated by a conversation with Yifei He. I teamed up with Tim Hunter and Jiwon Yun in building the modeling tool CCPC. T. Florian Jaeger helped me get access to the `tgrep2` version of the Switchboard corpus which was then used in the English part. Cristiano Chesi provided me the Siena University Treebank he worked on so that I could obtain corpus counts for Italian relative clauses. I also would like to thank Simone De Lemos and Margaret Renwick for their judgement on Italian examples. The Chinese part benefits from previous collaborations and discussions with Lena Jäger, Charles Chien-Jer Lin, and Fuyun Wu. I appreciate the generous help from Yaling Hsiao and Maryellen MacDonald for allowing me to use their extremely valuable corpus data on Chinese relative clauses. I also would like thank the East Asian Program of Cornell and the Chiang Ching-kuo Foundation for all the financial support I received during the completion of this dissertation.

I was determined to choose computational linguistics as my career when I was still a college student because it is a field that perfectly combines two of my interests, computers and languages. I graduated from one of the first undergraduate computational linguistics programs in China. Trained as a “non-traditional” linguist, I took mathematics and computer science courses that had never been considered in a Chinese linguistics program. I was also fortunate enough to join Professor Xiaohe Chen’s research team as a sophomore, working on an automatic word segmentation system. This undergraduate research experience was invaluable to me. Although my job at that time was neither innovative nor technical, it gave me a chance to see how a research project was

motivated, designed, and implemented in a timely fashion. I always see Professor Chen's strict and orderly research work as my example. After I continued my study in the US, I still kept in touch with many of my former professors, colleagues, and classmates. Xiaopeng Bai, Professor Baojia Li, Bin Li, Hui Li, Xiaolei Song, and Runhua Xu helped me one way or another on many of my research projects. I am truly grateful for all they have done.

Before transferring to Cornell, Michigan State University was my first home in this country. During my two years there, I started to get into the world of formal linguistics and had the opportunity to study and work with many great professors, including Joyce Chai, Rong Jin, Yen-Hwei Lin, Alan Munn, and Chunhong Teng. My sincere thanks also go to fellow Spartans, Vineet Bansal, Jing Fu, Ching-Ni Hsieh, Ho-Hsin Huang, Chiung-Yao Wang, Chia-Hsin Yeh, Xiaofei Zhang, and Yisu Zhou, for making my life colorful.

I spent six wonderful years and met a lot of wonderful people at Cornell. I benefited from working with my labmates Neil Ashton, Marisa Ferrara Boston, Anca Chereches, Effi Georgala, Kyle Grove, and David Lutz. Fellow Chinese students, Hongyuan Dong, Jixing Li, Nan Li, Zhiguo Xie, and Hao Yi, have all given me precious suggestions on my Chinese linguistic studies. Christina Bjorndahl, Alison Fisher, and Ika Nurhayani shared their comments with me in the research workshop on one of my papers which this work was built on. I also thank those who helped me practice my job talks, especially Amui Chong and Todd Snider. I learned the essential Cornell methodology of language teaching from my TA supervisor Qiuyun Teng and my colleague Dongming Shen De Angelis. I am sure that my future job as the Coordinator of the Chinese program at RIT will greatly benefit from it.

In the past two years, I made many good friends in Binghamton where the

majority of this dissertation was written. I am indebted to Jessica Chang, Cheng-Yu Chao, Chi-Chi Lei, Wilson Lin, Matt Peng, Joey Tsai, Kelly Tseng, Chieh-Ru Wu, and Danfeng Zhang. Thank you for playing basketball with me, providing me food when I had no time to cook, bringing me laughs when I was depressed, taking care of our little girl when Yijen and I were both sick, driving me to Ithaca for my weekly meeting or even allowing me to writing up this dissertation in your room. I wish to express my gratitude to all of you.

Most importantly of all, I am incredibly lucky to have support of my beloved family on this journey, my parents, in-laws, grand-parents and most importantly my wife and daughter. I met my wife Yijen Huang when we were both in the Linguistics program at Michigan State. As a linguist, Yijen understands how abstract and sophisticated that a linguistic problem could be. She patiently listens to my complains, comforts me when I have hard times, and provides me with insightful suggestions. She always encourages me to pursue the academic career that I am passionate about, even though it is not the most profitable one. Despite all her own busy work as a graduate student, she takes a lot of family responsibilities, especially after our daughter Olivia Chen was born. It is amazing to see how fast baby Olivia grows up everyday and I often wish that the progress of this dissertation could also keep up that pace. But finally I made it to this destination. Thank you, Olivia, for bringing happiness into my life!

## TABLE OF CONTENTS

Biographical Sketch . . . . .	iii
Dedication . . . . .	iv
Acknowledgements . . . . .	v
Table of Contents . . . . .	ix
List of Tables . . . . .	xi
List of Figures . . . . .	xiii
<b>1 Introduction</b>	<b>1</b>
1.1 Introduction . . . . .	1
1.2 Subject Advantage in relative clauses: an universal effect . . . . .	2
1.3 Animacy: a cross-linguistic sentence processing cue . . . . .	4
<b>2 Modeling incremental processing difficulties</b>	<b>9</b>
2.1 Entropy Reduction as a processing complexity metric . . . . .	10
2.2 Cornell Conditional Probability Calculator . . . . .	12
<b>3 English Relative Clauses</b>	<b>19</b>
3.1 The processing of English relative clauses . . . . .	20
3.2 Experimental studies of processing English relatives . . . . .	25
3.2.1 The subject-object asymmetry in English relatives . . . . .	25
3.2.2 The Role of Animacy in English RC Processing . . . . .	28
3.3 Modeling English RC processing . . . . .	33
3.3.1 English grammars . . . . .	33
3.3.2 English prediction I: the Subject Advantage . . . . .	41
3.3.3 English prediction II: the Animacy Effect . . . . .	46
<b>4 Italian Relative Clauses</b>	<b>53</b>
4.1 Italian relative clauses . . . . .	54
4.2 Experimental studies of Italian RC processing . . . . .	55
4.2.1 The Subject-Object Asymmetry . . . . .	55
4.2.2 The effect of animacy . . . . .	59
4.3 Modeling Italian RC processing . . . . .	66
4.3.1 Italian grammars . . . . .	66
4.3.2 Counting structures in Italian treebanks . . . . .	70
4.3.3 Italian prediction I: the Subject Advantage . . . . .	73
4.3.4 Italian prediction II: the head animacy effect . . . . .	77
<b>5 Chinese Relative Clauses</b>	<b>82</b>
5.1 Ambiguities in Chinese RCs . . . . .	83
5.2 Experimental studies of Chinese RC processing . . . . .	88
5.3 Modeling Chinese RC processing . . . . .	97
5.3.1 The Chinese Grammars . . . . .	97
5.3.2 Counting constructions in the Chinese treebank . . . . .	103

5.3.3	Chinese prediction I: the Subject Advantage . . . . .	108
5.3.4	Chinese prediction II: the head noun animacy . . . . .	113
<b>6</b>	<b>Conclusions</b>	<b>118</b>
6.1	Compare the animacy cue weight in three languages . . . . .	118
6.2	Future Work . . . . .	123
6.3	Conclusion . . . . .	127
<b>A</b>	<b>LIST OF ABBREVIATIONS</b>	<b>129</b>
	<b>References</b>	<b>130</b>

## LIST OF TABLES

2.1	A “mini-treebank” example used to weight an English MCFG. . .	14
2.2	The top ten derivations generated by an English grammar . . . . .	15
2.3	The top ten derivations generated by an intersection grammar conditioned on the prefix “David” . . . . .	16
3.1	Processing principles proposed for relative clauses (adopted from Yun, Chen, Hunter, Whitman, and Hale (2015)) . . . . .	22
3.2	Self-paced reading evidence for the subject advantage at the em- bedded verb in English RCs (Grodner & Gibson, 2005) . . . . .	26
3.3	Eye-tracking evidence for the subject advantage at both the embedded noun phrase and the embedded verb in English RCs (Staub, 2010) . . . . .	28
3.4	English ORs with inanimate head nouns are processed easier than those with animate heads (Traxler, Morris, & Seely, 2002). . .	30
3.5	Sample of MG lexical entries under the promotion analysis for English relative clauses . . . . .	35
3.6	Attestation counts from the Penn Treebank (Brown and WSJ cor- pora) are used to weight the grammar for English RCs . . . . .	39
3.7	Attestation counts from the Switchboard corpus are used to weight the animacy grammar for English RCs . . . . .	40
3.8	ER predicts additional processing difficulties at both the noun phrase and the embedded verb in English object relatives, con- sistent to the eye-tracking data of Staub (2010). Observed or pre- dicted subject advantages are highlighted in bold. . . . .	43
3.9	ER predicts that the subject advantage (OR minus SR) disappears in the RC region when an English OR has an inanimate head noun.	47
3.10	ER predicts that the reduced subject advantage in RCs with inani- mate heads is driven by the additional reading difficulty at the embedded SR verb. . . . .	48
4.1	Approximate reading times (ms) of the RC region in Italian SRs and ORs (Conditions (4.3)a and d), adopted from Di Domenico and Di Matteo (2009) . . . . .	59
4.2	Aggregated counts of active OR with different animacy patterns in two elicitation tasks of Belletti and Chesi (2011) . . . . .	65
4.3	Sample of MG lexical entries under the adjunction analysis for Italian relative clauses . . . . .	67
4.4	Sample of MG lexical entries under the promotion analysis for Italian relative clauses . . . . .	69
4.5	Animacy-encoded MG fragment derives an Italian OR with ani- mate subject and inanimate object . . . . .	71
4.6	Attestation counts from the Turin University Treebank are used to weight the grammar for Italian RCs . . . . .	72

4.7	Corpus counts of different matrix clause types in Italian . . . . .	73
4.8	Corpus counts of Italian RCs with various animacy patterns . . .	73
4.9	Corpus counts of Italian <i>cui</i> relatives with various animacy patterns . . . . .	74
4.10	ER predictions of Italian ORs with different animacy patterns compared against the production study results of Belletti and Chesi (2011) . . . . .	79
4.11	ER predictions of Italian ORs with different animacy patterns compared against the self-paced listening results of Arosio, Guasti, and Stucchi (2011) . . . . .	79
4.12	The subject advantage shrinks when OR has an inanimate head noun (Traxler et al., 2002) . . . . .	80
4.13	Predicted processing difficulty for the RC region . . . . .	81
5.1	Sample of MG lexical entries used to construct a Chinese relative clause . . . . .	99
5.2	Animacy-encoded MG fragment derives a Chinese OR with animate subject and inanimate object . . . . .	101
5.3	Attestation counts from Chinese Treebank 7 are used to weight the grammar for “disambiguated” Chinese RCs . . . . .	104
5.4	Tokens of Chinese RCs with various animacy patterns (Y. Hsiao & MacDonald, 2013, Table 3, Page 6) . . . . .	106
5.5	Animacy pattern of Chinese possessive structures . . . . .	107
5.6	Animacy pattern of Chinese RC heads with different types of matrix verb . . . . .	107
5.7	The frequency distributions of syntactic alternatives after reading the RC region is different between SRs and ORs (second row of Figure 5.7). . . . .	111
5.8	ER predictions of Chinese RCs with different animacy patterns . .	114
6.1	Predicted ERs at the “RC+head” region and the relative frequency of various RC conditions derive the animacy cue weights for English, Italian, and Chinese. . . . .	122

## LIST OF FIGURES

2.1	A modeling procedure illustration of CCPC system . . . . .	18
3.1	The SR advantage in the RC region by working memory group, quantified by the total time difference spent between SRs and ORs in an eye-tracking study (Traxler, Williams, Blozis, & Morris, 2005). . . . .	31
3.2	Mean regression-path durations for the four conditions in Example (3.4) (Lowder & Gordon, 2012) . . . . .	32
3.3	An English object relative clause derived by the promotion MG fragment . . . . .	36
3.4	The MG fragment allows the relativization of indirect object with P-stranding. . . . .	37
3.5	An English subject relative clause derived by the promotion MG fragment specifying the animacy of noun phrases . . . . .	38
3.6	Predicted reading difficulties of English subject and object relative clauses . . . . .	42
3.7	The transition at the embedded verb in English ORs . . . . .	44
3.8	The transition at the embedded verb in English SRs . . . . .	45
3.9	The transition at the NP determiner in English ORs . . . . .	46
3.10	An inanimate RC prefix is more ambiguous than an animate RC prefix because of more high-ranking OR alternatives. . . . .	49
3.11	ER predicts that the processing difficulty of having a inanimate subject is reduced by the clausal boundary (Lowder & Gordon, 2012). . . . .	50
3.12	The inanimate subject prefix allows a variety of alternatives. The main verb greatly disambiguates the prefix. . . . .	51
4.1	Self-paced reading times of Italian RC conditions (4.3)a-b in experiment 1 of Di Domenico and Di Matteo (2009) . . . . .	58
4.2	Self-paced reading times of Italian RC conditions (4.3)c-d in experiment 2 of Di Domenico and Di Matteo (2009) . . . . .	59
4.3	The subject advantage increases when ORs have animate heads (Arosio et al., 2011). . . . .	61
4.4	Italian object relative clause derived by an “adjunction” grammar . . . . .	67
4.5	Italian object relative clause derived by a “promotion” grammar . . . . .	68
4.6	Italian <i>cui</i> -relative clause derived by a “promotion” grammar . . . . .	70
4.7	Predicted word-by-word reading difficulties of Italian RCs using an “adjunction” grammar . . . . .	75
4.8	Predicted word-by-word reading difficulties of Italian RCs using a “promotion” grammar . . . . .	76
4.9	Transitions at the first set of words within RC region in Italian . . . . .	77
4.10	Transitions at the third set of words within RC region in Italian . . . . .	78

5.1	Reading times of each region in Chinese RCs (C.-J. C. Lin & Bever, 2006) . . . . .	91
5.2	Reading times of each region in Subject-modifying Chinese RCs (Jäger, Chen, Li, Lin, & Vasishth, Submitted, Experiment 1) . . . . .	95
5.3	Derivation tree of Chinese ORs using MG fragment in Table 5.1 . . . . .	100
5.4	X-bar tree of “disambiguated” Chinese ORs in (5.8b) built through syntactic derivations in Figure 5.3 . . . . .	101
5.5	X-bar tree of inanimate Chinese ORs with animate RC subject in (5.7) built through syntactic derivations in Figure 5.2 . . . . .	103
5.6	ER predictions of “disambiguated” Chinese RCs . . . . .	109
5.7	Derivations and their conditional probabilities in “disambiguated” Chinese RCs . . . . .	111
5.8	ER at the head noun in Chinese RCs; detail of last two rows in Figure 5.7. Here, expectations that the RC will be headless are highlighted in red. . . . .	112
5.9	SR transitions at heads with different animacy . . . . .	115
5.10	OR transitions at heads with different animacy . . . . .	116

# CHAPTER 1

## INTRODUCTION

### 1.1 Introduction

The competence and performance debate in linguistics has lasted for half a century since Chomsky (1965). In the field of human sentence processing, there exist a large number of studies discussing how readers' linguistic knowledge affects the understanding of utterances in real time. A consensus is that the sentence comprehension is guided by syntactic structures. Indeed, Miller (1962, p.752) has argued that "the proper functioning of our syntactic skill is an essential ingredient in the process of understanding a sentence." Yet it is hard to conclude that our ability of reading and speaking is solely driven by a grammar with syntactic constraints. In other words, the human parser may well use pieces of non-syntactic information when needed during sentence processing. In recent years, more and more research efforts have been made to explore how syntactic and non-syntactic factors interact in language comprehension. The research reported in this dissertation contributes to this ongoing investigation.

The present work employs a human sentence processing model to evaluate the effect of noun phrase animacy, a non-syntactic cue, in relative clause processing. This evaluation is done computationally by a complexity metric, Entropy Reduction (Hale, 2003). Predicted reading difficulties are used to compare against observed experimental data. The cross-linguistic nature of the modeling allows for an examination of the animacy cue weight in understanding relativized structures in three different languages, essentially re-conceptualizing existing ideas like the Competition Model of MacWhinney and Bates (1989),

from an information-theoretic prospective.

The results demonstrate that the animacy effect at the head noun seen in relative clauses of different languages can be predicted by a model incorporating fine-grained syntactic and non-syntactic information. The results also support the expectation-based sentence processing theory such that the comprehension difficulties correlate with readers' expectations about structures as well as, such as semantics and pragmatics. More importantly, the modeling system used here can visualize the probability distribution on syntactic alternatives at any parser state. Therefore, the modeling work often furnishes specific and linguistic interpretations for the comprehension difficulty inconsistencies between different types of relative clause. Many of these explanations have not yet been discovered in the psycholinguistic literature.

The rest of this chapter provides more details about the motivation of this dissertation topic, especially the choice of target sentence type: relative clauses and the choice of a non-syntactic cue: animacy.

## **1.2 Subject Advantage in relative clauses: an universal effect**

Sentence comprehension research has always been keen on identifying cross-linguistic constraints that lead to different online parsing difficulty patterns. The underlying idea assumes that the cognitive processes of sentence comprehension are subject to certain UNIVERSAL PRINCIPLES which should apply regardless of the language in question. In other words, although to some extent languages differ, it is possible to summarize a core set of universal parsing constraints. This task is a very attractive and psycholinguists have pursued it with

remarkable success.

In contrast to the universal processing principles, there also exists UNIVERSAL PROCESSING EFFECTS that are consistent across different languages. Perhaps the most widely-studied example is the processing of relative clauses (RCs) that modify an NP. The RC defines a particular proposition to restrict the potential referents (Comrie & Kuteva, 2005). When a noun phrase element is relativized from its original position in the underlying structure, it usually leaves a gap inside the relative clause. In this case, there is no overt element that is co-referential with the head NP. From the earliest studies of relative clauses, researchers have been interested in how different kinds of extraction behave. Keenan and Comrie (1977) surveyed a wide range of relative clauses across the world and formulated an important claim: the Accessibility Hierarchy (AH) of relativization, as shown in (1.1).

(1.1) The Accessibility Hierarchy (Keenan & Comrie, 1977)

subject > direct object > indirect object > oblique > possessor

The “larger than” symbol here indicates that the position on its left is easier to be relativized than the one on its right. For example, according to the AH, subjects are more likely to be relativized than objects and it is very hard to relativize possessives. The fact that subject relatives (SRs) are more common than object relatives (ORs) across languages is also attested in the processing literature involving a variety of methods, a list of which is listed on Page 21. It would be ideal if there were a theory that could explain this observation across languages.

Indeed, a number of processing principles have been developed to account

for the Subject Advantage in RC processing, as summarized in Table 3.1 on Page 22. The following chapters review some of those theories using examples in English, Italian and Chinese. The choice of the three languages above are not unintentional. For example, Chinese relative clauses are unique because they precede the head NP which they modify. Of the 825 languages in a database of Dryer (2005), 141 (or 17.1%) languages are prenominal, including many languages in Asia (Velupillai, 2012). They are a good test for sentence processing theories because the beginning words of prenominal RCs are the same with those in simple sentences, i.e. they share the same initial substrings. Without the indicator of an upcoming RC, such as a relative pronoun or a complementizer, several sentence-medial ambiguities will compete with the RC analysis in incremental processing. Another reason for selecting those three target languages in this study is to evaluate the animacy cue in sentence processing. Experimental studies have identified that comprehenders in those three languages more or less use their knowledge of noun phrase animacy in understanding relativized structures. Section 1.3 explains it in details.

### **1.3 Animacy: a cross-linguistic sentence processing cue**

As discussed earlier, the interaction between syntactic and non-syntactic information has been central to sentence processing research. Major classes of sentence comprehension models have assumed that syntactic structures determines how a sentence is interpreted, in this case, identifying the role of participants in an event. In contrast, the evaluation of role prototypicality involves non-syntactic information such as animacy and definiteness. However, these models differ in terms of how two classes of information interact during pro-

cessing. Frazier and colleagues (Frazier, 1979; Frazier & Rayner, 1982; Frazier & Clifton, 1996) examine this issue by adopting a MODULAR viewpoint and argue for a “two-stage” approach. In their model, the initial stage of analyzing a sentence only employs syntactic categorial information and a small set of structural preference principles. Non-syntactic cues, such as animacy, semantic plausibility and frequency of occurrence only influence processing choices at a later stage. Therefore, syntactically determined role identification precedes the non-syntactically determined role prototypicality. Contrastively, INTERACTIVE models assume that all information types are jointly taken into account from the very first stages of processing when they become available (MacDonald, Pearlmuter, & Seidenberg, 1994; Trueswell & Tanenhaus, 1994). On this architecture, the role prototypicality of arguments in the sentence may guide the choice of syntactic structure, hence influencing role identification.

Another classic study on non-syntactic cues is the COMPETITION MODEL proposed by MacWhinney and colleagues (MacWhinney, Bates, & Kliegl, 1984; MacWhinney & Bates, 1989). Their work takes an avowedly cross-linguistic perspective. It emphasizes the unique properties of each language and how these affect the interpretation of a sentence. It also motivated a line of research that sought to determine how speakers integrate cues such as word order, case, and animacy in order to arrive at an interpretation.

One of the crucial arguments that makes the competition model distinct from others is that languages differ not simply with respect to their unique properties but the relative weight of properties assigned in each language. The same type of linguistic cues may vary across languages in their “cue validities”, that is, as characterized by MacWhinney et al. (1984), “the information value of a

given linguistic device as a cue to an underlying meaning or intention". Cue validity is determined by a combination of how often a cue is available ("cue applicability") and how reliable the cue leads to the correct role identification ("cue reliability").

According to the competition model, the interpretation of a transitive sentence, e.g., which argument is agent and which argument is patient, is seen as a competition among different cues. A strong cue in one language could be among the weakest ones in another language. For example, they found that in Italian, the strongest cue of role identification is verb morphology, whereas in English, it is extremely weak as a cue (Bates, McNew, MacWhinney, Devescovi, & Smith, 1982). Word order, often seen as a syntactic information type, is a dominating cue in English with a high cue validity but is considered to be relatively weak in Chinese (Li, Bates, & MacWhinney, 1993).

One of the cross-linguistically applicable cues in interpreting a sentence is animacy. Contrastively, many other cues, such as agreement (or the verb morphology) could only be applied to languages like English, Italian and German but not to Chinese, which does not have clear inflectional markings. MacWhinney et al. (1984) and Li et al. (1993) conduct several offline sentence-interpretation tasks in which participants were asked to choose one of the two nouns in a sentence as the agent or the subject. Results suggest that the degree to which the animacy information determines the interpretation of arguments varied from one language to another. The effects are relatively weak in English, somewhat stronger in Italian and relatively strong in Chinese and German.

This animacy cue result is surprising because syntactic information is traditionally believed to be the primary determinant of role identification. Non-

syntactic cues such as animacy and definiteness only influence role prototypicality, i.e., whether a less prototypical Agent or Patient leads to a harder transitive construction. Moreover, MacWhinney and colleagues' experiments do not reveal the influence of animacy information in online sentence processing. This latter question calls for a more precise analysis of exactly how animacy interacts with syntactic constructions. In fact, MacWhinney et al.'s conclusions were not upheld in later studies such as Traxler et al. (2002, 2005). In these later experiments, Traxler found animacy to be highly effective in English as cue disambiguating agents from patients.

This dissertation addresses the animacy cue weight issue from the perspective of computational modeling. A complexity metric for incremental sentence comprehension is employed to model the observed animacy-driven reading pattern in three languages, in particular the RC head noun animacy effect. This inclusion or exclusion of animacy factors in the computational model leads to two different sets of predicted reading difficulties. The cue weight can therefore be quantified as the size of change between the two. The study integrates fine-grained linguistic information, both structural and non-structural, with formal grammars, which replaces MacWhinney et al.'s coarse-grained interpretation of animacy cue across languages.

As the first step of modeling the role of animacy in incremental sentence comprehension, this dissertation adopts a simple classification of noun phrase animacy, namely animate versus inanimate, in a way similar to previous works like MacWhinney et al. (1984) and Traxler et al. (2002). However, it does not mean that the animacy only has a binary categorization. Earlier linguistic typology works like Silverstein (1976) and Dixon (1979) have noticed that the gram-

matical processes in different languages seems to be sensitive to the relative degree of animacy of the noun phrases involved. In particular, they propose that continuous categories ranging from most animate to least animate can be ranked under an ANIMACY HIERARCHY.

(1.2) The Animacy Hierarchy (Dixon, 1979, 85)<sup>1</sup>

1st Person Pronoun > 2nd Person Pronoun > 3rd Person Pronoun >  
Proper Noun > Human Common Noun > Animate Common Noun >  
Inanimate Common Noun

The animacy modeling framework described in this dissertation can be naturally extended if this kind of continuous ranking of animacy category is being considered.

The rest of this dissertation is organized as follows. Chapter 2 introduces the complexity metric, Entropy Reduction, and the modeling methodology. Chapters 3 to 5 investigate the RC processing of English, Italian and Chinese respectively with a focus on both the Subject Advantage and the animacy effect at the head noun. Chapter 6 compares the predicted animacy cue weights across three target languages and discusses future research directions.

---

<sup>1</sup>Croft (2003) calls (1.2) “Extended Animacy Hierarchy” and argues that it involves three distinct but related functional dimensions: person hierarchy, referentiality hierarchy and semantic animacy hierarchy.

## CHAPTER 2

### MODELING INCREMENTAL PROCESSING DIFFICULTIES

This Chapter presents the modeling procedure adopted in the present dissertation research. The information-theoretic notion, Entropy Reduction (ER), is used to predict word-by-word reading difficulties of relative clauses in three languages. ER has been advanced as an alternative to Surprisal (Hale, 2001), also a complexity metric for human sentence comprehension. However, Hale (2006) shows that ER accurately derives the subject-object asymmetry in English relative clauses while surprisal cannot (Levy, 2008). Section 2.2 introduces a computer program, Cornell Conditional Probability Calculator (CCPC), to calculate the probability of alternative completions of initial substrings from user-supplied weighted grammars. Those probabilities determine how uncertain a parser state is at a given point in a sentence, as quantified by the Entropy of the set of viable derivations.

The CCPC not only computes ER values more precisely than the original work in Hale (2006) by properly renormalizing a weighted grammar, but also enumerates the syntactic alternatives that are in play at a given point, essentially visualizing the contents of a ranked parallel parser state. This new feature allows psycholinguists to conveniently investigate how structural disambiguations are related to comprehension difficulties.

## 2.1 Entropy Reduction as a processing complexity metric

A growing body of work in the field of sentence processing suggests that it is possible to model incremental comprehension difficulties using information-theoretic notions (Hale, 2001, 2003, 2006; Levy, 2008). For instance, Entropy Reduction is a complexity metric that quantifies the cognitive effort expended on a word. First introduced by Wilson and Carroll (1954), ER was applied to an artificial language which they created, by way of demonstrating its potential utility for morphosyntactic analysis. Hale (2003) revived the Entropy Reduction hypothesis and applied it to the analysis of processing asymmetries in English. The main idea is that words reduce uncertainty about the structure of the sentence. To some extent, ER is a generalization of the “flipping the preferred interpretation” idea in Narayanan and Jurafsky (2002). Here the flip only counts if the reader moves towards a less-confused state of mind.

Entropy Reduction allows for the possibility of parallel parsing. The uncertainty that ER deals with reflects the ambiguity between multiple parses, including expectations about as-yet-unheard words. As new words come in, given what one has already read, the probability of grammatical alternatives fluctuates. The idea of ER is that decreased uncertainty about the whole sentence, including probabilistic expectations, correlates with observed processing difficulty. Such processing difficulty reflects the amount of information that a word supplies about the overall disambiguation task in which the reader is engaged.

The average uncertainty of specified alternatives can be quantified using the fundamental information-theoretic notion, entropy, as formulated below in def-

inition (2.1). In a language-processing scenario, the random variable  $X$  in (2.1) might take values that are derivations on a probabilistic grammar  $G$ . We could further specialize  $X$  to reflect derivations proceeding from various categories, e.g., “NP”, “VP”, “S” etc. Since rewriting grammars always have a start symbol, e.g. “S”, the expression  $H_G(S)$  reflects the average uncertainty of guessing any derivation that  $G$  generates.

$$H(X) = - \sum_{x \in X} p(x) \log_2 p(x) \quad (2.1)$$

This entropy notation extends naturally to express conditioning events. If  $w_1 w_2 \dots w_i$  is an initial substring of a sentence generated by  $G$ , the conditional entropy  $H_G(S|w_1 w_2 \dots w_i)$  will be the uncertainty about just those derivations that have  $w_1 w_2 \dots w_i$  as a prefix. Conditional entropies can be calculated using standard techniques from computational linguistics such as chart parsing (Bar-Hillel, Perles, & Shamir, 1964; Nederhof & Satta, 2008). Textbooks like Jurafsky and Martin (2008, Chapter 13) offer a tutorial presentation of these ideas. By abbreviating  $H_G(S|w_1 w_2 \dots w_i)$  with  $H_i$ , the cognitive load  $ER_i$  reflects the difference between conditional entropies before and after  $w_i$ , a particular word in a particular position in a sentence. Formula (2.2) defines the ER complexity metric. It says that cognitive work is predicted whenever the uncertainty about the sentence’s structure, as generated by the grammar  $G$ , goes down after reading in a new word.

$$ER_i = \begin{cases} H_{i-1} - H_i & \text{if this difference is positive} \\ 0 & \text{otherwise} \end{cases} \quad (2.2)$$

Intuitively, disambiguation occurs when the uncertainty about the rest of the sentence decreases. In such a situation, readers’ “beliefs” in various syntactic alternatives take on a more concentrated probability distribution (Jurafsky,

1996). The disambiguation work spent on this change is exactly the Entropy Reduction. By contrast when beliefs about syntactic alternatives become more disorganized, e.g. there exist many equiprobable syntactic expectations, then disambiguation work has not been done and the parser has gotten more confused. The background assumption of ER is that human sentence comprehension is making progress towards a peaked, disambiguated parser state and that the disambiguation efforts made during this process can be quantified by the reductions of structural uncertainty conveyed by words.

ER proposal is not to be confused with another widely applied complexity metric, Surprisal (Hale, 2001; Levy, 2008), which is the conditional expectation of the log-ratio between forward probabilities of string prefixes before and after a word. Surprisal and ER have different properties, for instance entropy reductions are additive whereas surprisals are not. Blachman (1968) clarifies of the difference between surprisal and ER on a mathematical level<sup>1</sup>. ER has had rather better success in modeling sentence-medial ambiguities, such as those found in English object relatives. Surprisal has not led, as yet, to much insight into these effects. Chapter 3 will discuss the application of Surprisal and Entropy Reduction to English RC processing in details.

## 2.2 Cornell Conditional Probability Calculator

This section introduces the Cornell Conditional Probability Calculator (CCPC).<sup>2</sup> By combining a formal grammar and structural frequency information, CCPC

---

<sup>1</sup>See Yun et al. (2015, Appendix C) for a short review.

<sup>2</sup>CCPC is designed by a team of computational linguists at Cornell University, led by John Hale. Other team members include Zhong Chen, Tim Hunter, and Jiwon Yun. The software system is now freely available at <http://conf.ling.cornell.edu/compling/software.htm>

can enumerate all legitimate structural alternatives that shares a common string prefix and at the same time computes the conditional probability associated with each alternative. The calculated ER values reflect a theoretical prediction about the pattern of comprehension difficulty on a word-by-word level.

Since structural alternatives are syntactic constructions, the CCPC modeling procedure starts from a user-prepared grammar which covers the target sentences, e.g. relative clauses. CCPC can use context-free phrase structure grammars, but it also supports Minimalist Grammars (MGs) involving movement in the style of Stabler (1997). MGs are a transformational formalism that adopts ideas from the Minimalist Program (Chomsky, 1995). Stabler’s formalization involves two generalized transformations: MERGE and MOVE. MERGE is a binary rule, analogous to ordinary context-free grammar rules or function application in categorial grammar. MOVE is a unary rule that is non-concatenative. Using a translator written by Guillaumin (2005), CCPC converts a MG to an equivalent Multiple Context Free Grammar (MCFG) (Seki et al., 1991). Michaelis (2001) and Harkema (2001) have shown that MCFGs are mildly context-sensitive (Joshi, 1985). This means that while the derivation trees have a context-free, tree-like structure, and thus can be viewed as a weighted grammar, the derived languages exhibit the rich non-local dependencies, including crossing dependencies, that we find in natural language. The following chapters demonstrates how CCPC uses MGs to model the processing of relative clause structures.

Weighting the MCFG rules by attestation counts of relevant structures in the corpus forms a weighted grammar (WMCFG). CCPC incorporates an automatic option to estimate weights for a MCFG by parsing a “mini-Treebank”. Hunter (2013, Page 11) has provided us such an example in Table 2.1.

20	mary sleeps
20	john sleeps
10	mary sleeps on tuesday
10	john sleeps on tuesday
2	mary sleeps on tuesday on tuesday
2	john sleeps on tuesday on tuesday

Table 2.1: A “mini-treebank” example used to weight an English MCFG.

The frequencies associated with each sentence are retrieved from parsed corpora, such as treebanks. CCPC goes through the complete list in this mini-treebank and parses all sentences. At the same time, it estimates the probability of rules used to construct relevant structures based on the structural frequency in the mini-treebank. This procedure is automatic but has one limit since it cannot handle ambiguous strings. There is no way for the CCPC system to specify which derivation is intended.

For the accuracy of ER modeling, this dissertation research does not employ the automatic weighting procedure described above, but rather hand-weight MCFGs with construction frequencies from corpora. This requires a deep understanding on both MG and MCFG formalisms. Examples will be given as we walk through concrete modeling tasks in following chapters. Section 4.2 of Yun et al. (2015) also explains the grammar weighting procedure in a tutorial fashion.

The weighted grammar allows us to calculate the probability of constructing a complete derivation. Table 2.2 is an example of ten derivations generated by an English MG with highest probabilities. Although in this table each probability number seems to be associated with a grammar-produced “string”, e.g.,

0.00379689	<i>he matter -ed</i>
0.00315572	<i>David matter -ed</i>
0.00236679	<i>Sally matter -ed</i>
0.00221223	<i>they matter -ed</i>
0.00221223	<i>I matter -ed</i>
0.00144425	<i>the treat be -s clever</i>
0.00129323	<i>the treat be -s young</i>
0.00129323	<i>the treat be -s right</i>
0.00129323	<i>the treat be -s poor</i>
0.000991171	<i>the treat be -s strange</i>

Table 2.2: The top ten derivations generated by an English grammar

“0.00379689 *he matter -ed*”, they are in fact the probability of a syntactic derivation. Therefore, it is possible that the same “string” shows up more than once in this kind of derivation list. Under that circumstance, such “string” is ambiguous and corresponds to multiple derivations. Hand-weighting the MCFG ensures that the system calculates the entropy values correctly.

As Section 2.1 describes, the quantity that ER advances as a cognitive model is a conditional entropy. These values reflect uncertainty about every analysis and every grammatically possible sequence of words that can follow a given prefix string. To compute these conditional entropies, CCPC uses chart parsing to recover probabilistic “intersection” grammars  $G'$  conditioned on each prefix of the sentences of interest (Nederhof & Satta, 2008). An intersection grammar derives all and only the sentences in the language of  $G$  that are consistent with the initial prefix. It implicitly defines comprehenders’ expectations about how the sentence continues. Given the prefix string, the conditional entropy of the start symbol models a reader’s degree of confusion about which construction he or she is in at that point in the sentence. Comparing the conditional entropies

0.0416261	<i>David matter -ed</i>
0.00815858	<i>David doesnt matter</i>
0.00607872	<i>David matter -s</i>
0.00594986	<i>David matter</i>
0.00263401	<i>David pay -ed for the treat</i>
0.0017826	<i>David sell -ed the treat</i>
0.00177774	<i>David tell -ed the treat</i>
0.00173645	<i>David get -ed the treat</i>
0.00173645	<i>David leave -ed the treat</i>
0.00167712	<i>David have -ed pay -en for the treat</i>

Table 2.3: The top ten derivations generated by an intersection grammar conditioned on the prefix “David”

before and after adding a new word, any decrease quantifies disambiguation work that, ideally, could have been done at that word.

Besides computing entropies, CCPC also enumerates the “remainder set” of syntactic alternatives from intersection grammars to get an intuitive picture of how uncertainties are reduced during parsing. For example, given an intersection grammar conditioned on the prefix “David”, CCPC is able to output all derivations that are still in play. Ten of them with the highest probability are listed in Table 2.3. Derivations in Table 2.3 shares the same prefix, in contrast to those generated by a non-intersection grammar in Table 2.2.

These syntactic alternatives are essential in understanding how ER values are calculated. More importantly, by looking at the remainder set of derivations from one word to the next, psycholinguists will be able to come up with linguistic interpretations for certain processing effect observed at a given word, especially when the effect was caused by disambiguation.

To summarize, the modeling procedure in the present dissertation work can be illustrated in a flowchart in Figure 2.1. By using a handwritten grammar, this work focuses on just a set of linguistically relevant alternatives. By using corpus counts, it attempts to define a realistic, frequency-sensitive notion of expectations in performance.

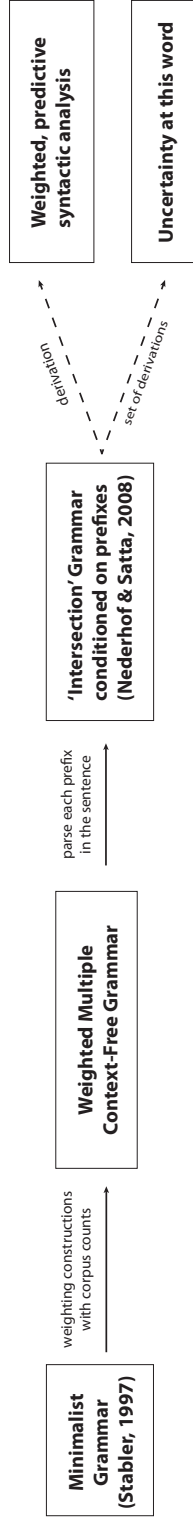


Figure 2.1: A modeling procedure illustration of CCPC system

## CHAPTER 3

### ENGLISH RELATIVE CLAUSES

Section 2 has introduced the idea of Entropy Reduction, as a complexity metric for predicting processing difficulties occurred during incremental comprehension (Hale, 2003, 2006). The CCPC system is designed to allow psycholinguists to investigate the structural uncertainty at each word by enumerating a list of “still in play” syntactic derivations ranked by their conditional probabilities. It therefore makes it easier to find linguistic interpretations for observed processing effects.

As the first of three modeling examples reported in this dissertation, Section 3 first examines one of the most widely-studied findings in the sentence processing literature, the Subject Advantage in English relative clauses. The present research contributes to the discussion on this topic by offering new insights into the parsing strategy and the expectation of structures that derive the additional comprehension difficulty in object-extracted relatives. ER values calculated by CCPC accurately predict the loci of subject advantage observed in recent eye-tracking experiment (Staub, 2010). Importantly, instead of interpreting the subject-object asymmetry with an “two-way approach” of both memory burden and structure frequency (Lewis & Vasishth, 2005; Demberg & Keller, 2009), ER modeling can be seen as an unified explanation of the amplified processing difficulties observed in English ORs.

By adding the frequency distribution of noun phrase animacy to a probabilistic grammar fragment, CCPC also models the animacy effect at the RC head noun (Traxler et al., 2002, 2005). In addition, the comparison of predicted reading difficulties between a simple sentence and an RC with the same animacy

pattern supports the argument that the magnitude of animacy effect is highly related to the structure of the sentence (Lowder & Gordon, 2012).

This chapter is presented as follows. Section 3.1 reviews different theories proposed for English RC processing. Section 3.2 discusses the key experimental evidence on this topic, followed by a detailed report of the modeling procedure and the comparison of ER predictions against experimental observations in Section 3.3.

### 3.1 The processing of English relative clauses

English relative clauses have been one of the major targets of human sentence processing research since 1970s, mostly because of their unique and yet less ambiguous structure. They have been an ideal platform for evaluating a variety of sentence processing theories. A pair of relative clause construction in English is shown in Example (3.1).

(3.1) a. Subject relatives (SRs)

The reporter<sub>*i*</sub> who *e<sub>i</sub>* attacked the senator left the room.

b. Object relatives (ORs)

The reporter<sub>*i*</sub> who the senator attacked *e<sub>i</sub>* left the room.

In (3.1a), the noun phrase “the reporter” modified by the relative clause is not only the subject of the matrix clause, but also the agent of the embedded verb “attacked”. The extraction/relativization of “the reporter” from the underlying position within the relative clause is notated in these examples with a

gap, symbolized by the empty category  $e_i$ . Similarly, in (3.1b), “the reporter” also serves as the head of a relative clause, although it is co-indexed with the empty element at the embedded object position. These indices indirectly suggest which interpretation the construction receives: one in which the logical object of “attacked” is identical with the surface subject, “reporter”. Mastering this referential notation is important for what follows in Chapters 4 and 5, for instance in the description of Italian RCs that begins on Page 53.

A large literature documents the finding that subject relatives (SRs) are easier to process than object relatives (ORs) in English, a processing asymmetry known as the Subject Advantage. The subject advantage has been observed in a variety of different measures, including: reading times (King & Just, 1991), eye-tracking (Traxler et al., 2002), ERP (King & Kutas, 1995), fMRI (Just, Carpenter, Keller, Eddy, & Thulborn, 1996) and PET (Stromswold, Caplan, Alpert, & Rauch, 1996). This processing effect has also been suggested in other languages, including those languages where relative clauses appear before the head noun (C.-J. C. Lin, 2008). The processing of prenominal RCs will be discussed in Section 5.

A variety of more or less universal processing principles, as shown in Table 3.1, have been advanced as candidate explanations for the universal subject advantage pattern. Among them, recent studies have appealed particularly to the working memory idea and to the structural expectation idea. The first one explains the subject advantage in terms of a reduced memory load in SRs compared to ORs whereas the latter one suggests that readers have higher structural expectations for SRs because they are more frequently used.

Broad Categories		General Proposals
WORD ORDER	(Bever, 1970; MacDonald & Christiansen, 2002)	The sequence of words in SRCs is closer to the canonical word order than that in ORCs.
PARALLEL FUNCTION	(Sheldon, 1974)	SRCs are easier to process than ORCs because their head nouns play the same role in both the main clause and the subordinate clauses.
PERSPECTIVE MAINTENANCE	(MacWhinney, 1977, 1982)	SRC structures maintain the human perspective and should be easier to process than those that shift it, e.g. ORCs.
ACCESSIBILITY HIERARCHY	(Keenan & Comrie, 1977)	Universal markedness hierarchy of grammatical relations ranks the relativization from subject higher.
WORKING MEMORY	LINEAR DISTANCE: (Wanner & Maratsos, 1978; Gibson, 2000; Lewis & Vasishth, 2005)	ORCs are harder because they impose a greater memory burden.
	STRUCTURAL DISTANCE: (O'Grady, 1997; Hawkins, 2004)	
STRUCTURAL EXPECTATION	TUNING HYPOTHESIS: (Mitchell, Cuetos, Corley, & Brysbaert, 1995; Jurafsky, 1996)	SRCs occur more frequently than ORCs and therefore are more expected and easier to process.
	SURPRISAL: (Hale, 2001; Levy, 2008)	ORCs are more difficult because they require a low-probability rule.
	ENTROPY REDUCTION: (Hale, 2006)	ORCs are harder because they force the comprehender through more confusing intermediate states.

Table 3.1: Processing principles proposed for relative clauses (adopted from Yun et al. (2015))

Among the MEMORY-BASED ACCOUNTS, Gibson and colleagues (Gibson, 1998, 2000; Warren & Gibson, 2002; Grodner & Gibson, 2005) have tried to explain the processing of English RCs using the Dependency Locality Theory (DLT). In particular, the sentence-initial head noun, such as “the reporter” in Example (3.1), is kept in the working memory until the embedded verb “attacked” is being processed. Due to the longer distance (both linear and structural) between the two phrases in ORs than in SRs, integrating, or building the syntactic dependency between the two, is considered to be harder in ORs. The DLT defines the so-called integration cost in terms of the number of new discourse referent between the head noun and the embedded verb.

Lewis and colleagues (Lewis & Vasishth, 2005; Lewis, Vasishth, & Van Dyke, 2006) have also proposed a memory-based account model which essentially applies Anderson’s memory theory, as realized in the ACT-R cognitive architecture (Anderson et al., 2004), to the special domain of sentence processing. Because of this greater coherence with other cognitive theories, Lewis and Vasishth’s proposal is conceptually superior to earlier memory-based theories such as Wanner and Maratsos (1978)’ HOLD hypothesis, Gibson (1998, 2000)’s DLT etc. Their model acknowledges the locality effect such that constructions that involve long-distance dependencies result in heavier working memory load and longer processing delays. In addition, they emphasize the importance of interference in the memory retrieval process. Readers carry additional memory burden when reading an English OR because the embedded subject, such as “the senator” in Example (3.1b), serves as a potential subject candidate for the memory retrieval at the embedded verb. This kind of interference is similarity-based since both “the reporter” and “the senator” are noun phrases. Evidence for the SIMILARITY-BASED INTERFERENCE has been reported on other process-

ing cues as well, such as number, gender and animacy.

In contrast to the memory-based theories, the difficulty of processing English ORs can also be explained in terms of comprehenders' expectations for alternative syntactic constructions. Perhaps the most direct way to compare expectations on SRs and ORs is their relative frequencies in a language, as evidenced by corpus counts (Roland, Dick, & Elman, 2007). SRs are more common than ORs in almost all languages, including English. The EXPECTATION-BASED THEORIES have different instantiations. One leading idea is SURPRISAL (Hale, 2001; Levy, 2008), which has been proposed to reflect the degree of "surprise" when reading each word in a sentence. For example, the surprisal of ORs is higher than SRs because deriving an OR structure requires a syntactic rule with lower probability. Surprisal is able to link itself to the more general Tuning Hypothesis because the probability associated with each grammatical rule is estimated from attestation counts in corpora, for example, treebanks. To derive a well-attested construction usually involves rules with a higher probability.

As discussed in Chapter 2, the Entropy Reduction idea can also be seen as an instantiation of the expectation-based theories. However, it differs from other ideas in this category, such as surprisal, in terms of how the structural frequency information are reflected in processing difficulties. The key idea of ER is that the reduction of uncertainty level is a result of disambiguation. The processing cost spent at a word correlates with the magnitude of the entropy drop. Although ER might seem less intuitive than surprisal, it has already been used to made a number of successful predictions on various processing effects, including RCs with prenominal heads (Yun et al., 2015).

It is important to know that both classes of processing theories, memory-based and expectation-based accounts, are able to explain a range of processing patterns. And in some cases, the experimental results could not be explained by a single factor. The processing of English relative clauses have been argued as an example for a two-way approach that is co-determined by both the working memory burden and structural expectations. Section 3.2 below addresses some of the most crucial experimental findings on this topic and discusses in details on the predictions made by different processing theories.

## **3.2 Experimental studies of processing English relatives**

### **3.2.1 The subject-object asymmetry in English relatives**

The subject advantage in English RCs is robust. Almost all existing experimental literature have shown that subject-extracted RCs are easier to process than object-extracted ones. However, one of the major differences among previous studies is the exact location of where the processing of ORs is delayed. Investigating the location of subject advantage is important because it can be used to evaluate different processing theories.

Grodner and Gibson (2005) conducted two self-paced reading experiments on English SRs and ORs. Participants were asked to press a button in order to read each word in a experimental sentence in their own pace. There was only one word at a time on the screen while both the preceding and forthcoming

words are dashed. In their Experiment 1, they tested standard RCs like the pair of examples shown in (3.1). Table 3.2 listed the recorded reading times at each word in an RC. It adopts the convention in English RC literature to compare the reading time at the same word in SRs and ORs. The column from left to right follows the word order in ORs.

	Type	Relative pro	Det	Noun	RC verb
Reading Time (ms)	SR	349.8	334.3	384	<b>354.8</b>
	OR	343	348.1	357.6	<b>422</b>

Table 3.2: Self-paced reading evidence for the subject advantage at the embedded verb in English RCs (Grodner & Gibson, 2005)

Grodner and Gibson’s finding in Experiment 1 was that ORs were read much slower than SRs late in the RC, in particular, at the embedded verb with a slowdown of 66.2 ms. At the embedded subject (“Det Noun”), ORs were actually read slightly faster than SRs (12.6 ms). This result can be explained by memory-based approaches such as the DLT because of the additional integration cost at the embedded verb in ORs. Expectation-based surprisal can also explain the subject advantage at the embedded verb, but only by attributing it to the processing spillover from the preceding embedded noun phrase (Levy, 2008). The English RC modeling using surprisal predicts that the processing difficulty in ORs should incur early at the embedded subject, or even at the onset of the subject NP, namely the determiner (Hale, 2001; Levy, 2008). This is because readers have a higher expectation for SRs than for ORs after reading the relative pronoun. Seeing the next word, a determiner, is relatively surprising.

In Experiment 2, Grodner and Gibson (2005) tested whether the observed difficulty at the OR verb was indeed an effect of locality. The idea was to insert more materials between the embedded subject and the verb in an English OR.

One example would be adding an intervening prepositional phrase (PP) which involved a noun phrase, i.e. a new discourse referent:

(3.2) The reporter<sub>i</sub> who the senator *from New York* attacked *e<sub>i</sub>* left the room.

Reading times at the embedded verb such as “attacked” in this manipulated OR increased rather than decreased. On the other hand, there was no significant reading time increase at the inserted phrase. This finding provided convincing evidence for the memory-based approaches because the memory burden at the embedded verb kept increasing when the verb was taken away from the the head noun. Surprisal, however, was unable to predict the increased difficulty at the OR verb in this case since there was no reading time delay at the PP. The increased processing difficulty at the embedded verb was not the spillover from the head noun. In fact, the surprisal model of Levy (2008) predicted that the reading time at the embedded verb in ORs decreased when adding more intervening materials. He acknowledged that:

“surprisal has a major effect on word-by-word processing difficulty, but that truly non-local (i.e., long-distance) syntactic dependencies such as relativization and wh-question formation are handled fundamentally differently from local syntactic dependencies, and the retrieval and integration of a long-distance dependent incurs a substantial processing cost comparable to the cost of a highly surprising word”

Although surprisal was unable to capture the locality effect at the English OR verbs, recent evidence from an eye-tracking study (Staub, 2010) provided

some supports to the expectation-based processing theories. Staub (2010) used the similar type of standard SRs and ORs as Grodner and Gibson (2005) did. However, he found that the total time of eye-fixation (Go-past time) is longer in ORs than in SRs not only at the RC verb but also at the embedded noun phrase. The subject advantage was statistically significant throughout the RC region, as highlighted in bold in Table 3.3. In addition, readers were more likely to look back when they read a noun phrase in ORs, as evidenced by the higher percentage of regression. Taking both results together, Staub (2010) argued that both structural expectation-based accounts and memory retrieval-based accounts may co-determine the processing difficulty in English relative clauses. This “two-way” proposal is also compatible with a computational modeling using reading data for naturally occurring relative clauses (Demberg & Keller, 2009).

	Type	Relative pro	Det	Noun	RC verb
Go-past time (ms)	SR	283	<b>272</b>	<b>375</b>	<b>333</b>
	OR	303	<b>382</b>	<b>459</b>	<b>420</b>
Percentage of regression	SR	0.12	<b>0.12</b>	<b>0.16</b>	0.17
	OR	0.11	<b>0.36</b>	<b>0.40</b>	0.15

Table 3.3: Eye-tracking evidence for the subject advantage at both the embedded noun phrase and the embedded verb in English RCs (Staub, 2010)

### 3.2.2 The Role of Animacy in English RC Processing

In addition to the subject advantage effect, a number of previous work have also shown that the processing load of ORs is modulated by the animacy of noun phrases (English: Traxler et al. (2002, 2005), Dutch: Mak, Vonk, and Schriefers (2002, 2006), Chinese: Wu, Kaiser, and Andersen (2012)). Although RCs with

two animate nouns are widely used as experimental stimuli, including the two English studies discussed above, this type of RC is not the most frequent one in natural language usage (Roland et al., 2007). In fact, there seems to be correlations between RC type and head noun animacy (Gennari & MacDonald, 2008). For example, ORs tend to occur with inanimate heads than with animate ones. The frequency pattern of animacy in RCs coins the animacy pattern in simple sentences, e.g. there are more inanimate objects than animate objects. In English, Traxler et al. (2002, 2005) tested the so-called ANIMACY EFFECT in a series of eye-tracking experiments. Using the four conditions in Example (3.3), the magnitude of subject advantage was found related to the animacy of RC head.

(3.3) Experimental conditions of Traxler et al. (2002)

- a. SR, +animate, –animate

The director that watched the movie received a prize.

- b. OR, +animate, –animate

The director that the movie pleased received a prize.

- c. SR, –animate, +animate

The movie that pleased the director received a prize.

- d. OR, –animate, +animate

The movie that the director watched received a prize.

All four conditions in Example (3.3) have contrastive animacy between two noun phrases. Conditions a and b have the animate head noun “the director” while the head nouns of Conditions c and d, “the movie”, are inanimate. The major finding in Traxler et al. (2002) was that ORs with inanimate head (Condition d) were read much faster. Therefore, the size of subject advantage effect

decreases, as shown in Table 3.4. Here, the subject advantage is calculated as the total time difference at the three-word RC region between SRs and ORs.

Type	Head-NP	Embedded-NP	Total time (ms)	Subj. Adv.
SR	+anim	-anim	881	<b>294</b>
OR	+anim	-anim	1175	
SR	-anim	+anim	847	<b>39</b>
OR	-anim	+anim	886	

Table 3.4: English ORs with inanimate head nouns are processed easier than those with animate heads (Traxler et al., 2002).

Traxler et al. (2005) tested this animacy effect at the head noun in three groups of experimental subjects with different working memory capacity (WMC). As shown below in Figure 3.1, the reduced subject advantage was found in inanimate RCs across all three groups. There was even a reversed processing pattern in the high WMC group such that ORs with inanimate head were read slightly faster. The ER modeling in Section 3.3 predicts a similar animacy effect reported by Traxler and colleagues.

Gennari and MacDonald (2008) argued that the animacy effect in English ORs can be treated as frequency-driven. ORs with inanimate head noun are easier because objects are more likely to be inanimate. On the contrary, if the head of OR is animate, it contradicts with the comprehenders' expectation such that the object or the patient needs to be an inanimate noun.

Table 3.4 shows that the reading time changes dramatically when the animacy of an OR head is manipulated (Lines 2 and 4). The same effect was not found for subject relatives (Lines 1 and 3). Motivated by this pattern, Lowder and Gordon (2012) carried out an eye-tracking study on SRs. They investigated the relationship between the animacy effect and the syntactic structure by com-

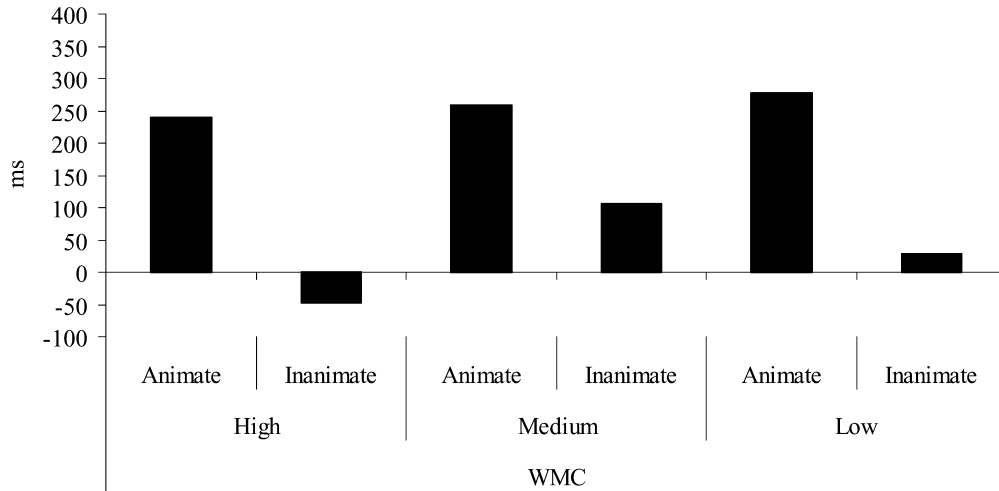


Figure 3.1: The SR advantage in the RC region by working memory group, quantified by the total time difference spent between SRs and ORs in an eye-tracking study (Traxler et al., 2005).

paring two pairs of conditions below.

(3.4) Experimental conditions of Lowder and Gordon (2012)

a. SR, +animate subject

The cowboy **that** concealed the pistol was known to be unreliable.

b. Simple sentence, +animate subject

The cowboy concealed the pistol last night in the saloon.

c. SR, -animate subject

The pistol **that** injured the cowboy was known to be unreliable.

d. Simple sentence, -animate subject

The pistol injured the cowboy last night in the saloon.

Examples (3.4)a and c are subject relatives, one with animate head noun and the other with inanimate head noun. Examples (3.4)b and d are simple sentence conditions created simply by dropping the complementizer “that” of the SRs. In that way, the subject-verb dependency is within the same clause in Conditions b and d but crosses one clausal boundary in Conditions a and c.

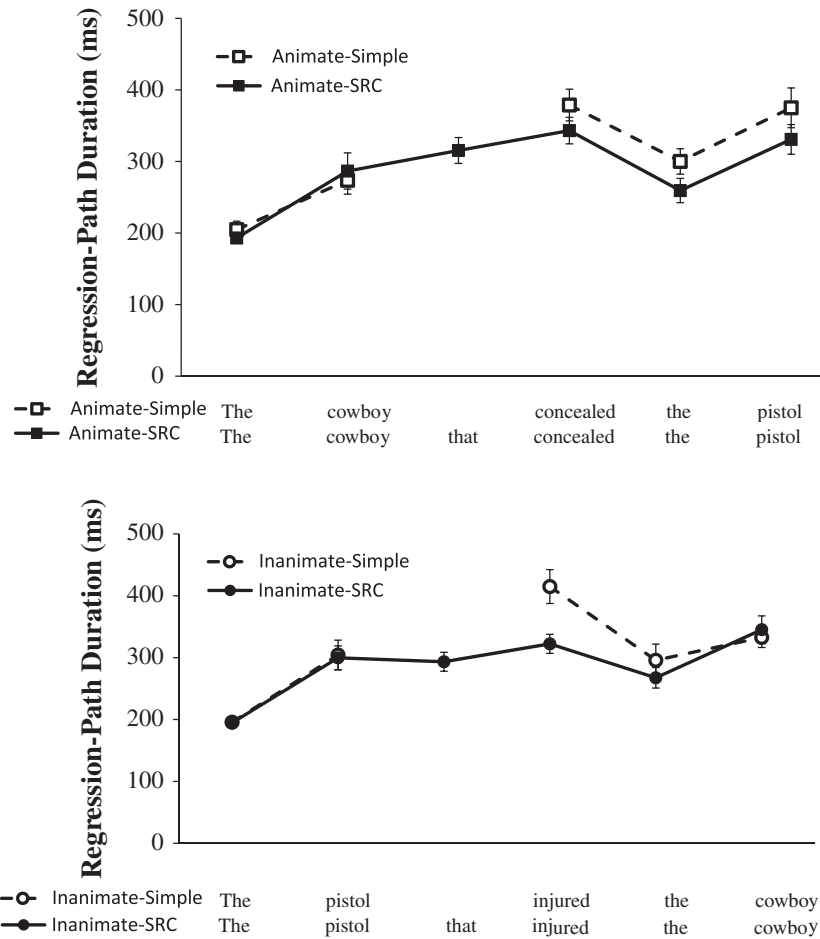


Figure 3.2: Mean regression-path durations for the four conditions in Example (3.4) (Lowder & Gordon, 2012)

Figure 3.2 illustrates the mean regression-path durations across four conditions. The bottom subfigure depicts an interesting processing contrast at the verb “injured”, that is, simple sentences with inanimate subjects are read much

slower than their counterpart SRs with inanimate head. The facilitated processing in inanimate SRs was not found in SRs with animate head noun. This result suggests that the animacy dependency between the subject and the verb is affected by the syntactic structure. The animacy effect is reduced when the subject and the verb belong to different clauses, in this case, the main clause and the RC.

To summarize, previous studies on English relative clauses have supported the universal subject advantage, as discussed in Section 1.2 of Chapter 1. On the other hand, animacy seems to affect RC processing, in a way that mirrors corpus frequency. Using the CCPC system, Section 3.3 reports modeling work which derives both English patterns from formal grammars and corpus counts.

### **3.3 Modeling English RC processing**

This section presents the modeling procedure and results of English relative clause processing. The ER values accurately capture two major effects which have been observed in experimental studies, namely the subject advantage and the animacy effect at the RC head noun. To do so, the system requires language knowledge of both the relative clause structure and the frequency distribution of relevant constructions, including animacy-specified noun phrases.

#### **3.3.1 English grammars**

The parser state's degree of uncertainty at any time hinges on legitimate syntactic alternatives that are still in play. Therefore, it is necessary to identify non-trivial syntactic constructions that could compete with the RC analysis during

incremental sentence comprehension. This is especially crucial when modeling the processing of head-initial RCs. Because the relative pronoun has already signalled a forthcoming relative clause, the targeted relative clause analysis is essentially competing with its own variants. The two English MG fragments (with or without specifying animacy) prepared for this modeling task capture a range of linguistic features, including verb transitivity, adjunct modification, reduced or explicit complementizer. They serve as the necessary complement to the syntactic treatment of relative clause itself.

Most transformational syntax literature on relative clauses focuses on the WH-movement of NP element, for example, how the object “the reporter” in (3.5)a appears at the sentence initial position and attaches to an RC in (3.5)b. Hale (2006) implemented two syntactic treatments of English RCs in his modeling work, namely the adjunction analysis (Chomsky, 1977, 1986, 1995) and the promotion analysis (Vergnaud, 1985; Kayne, 1994; Bianchi, 2002). Using (3.5) as an example, the basic idea of the adjunction analysis involves two steps. First, the WH-movement deposits the relative pronoun, such as “who”, from the underlying object position to the leftmost edge of the constituent “the senator attacked” and forms a complementizer phrase (CP). Second, this CP is adjoined to the head noun phrase “the reporter” as a postmodifier.

- (3.5) a. Simple sentence/Underlying Form of the OR  
The reporter attacked the senator.
- b. Object relatives (ORs)  
The reporter<sub>*i*</sub> who the senator attacked *e<sub>i</sub>* left the room.

Hale (2003) wrote a grammar expressing the adjunction analysis of RC and

tested it in an ER modeling work. However, he found the results did not correlate with the Accessibility Hierarchy of Keenan and Comrie (1977), as measured by repetition accuracy scores in Keenan and Hawkins (1987). The modeling based on a promotion analysis of RC, on the other hand, mirrors the Accessibility Hierarchy. Under the promotion analysis, both the relative pronoun “who” and the head noun “reporter” are transformed to the left edge of the CP. The modeling work reported in this chapter adopts the promotion analysis in both MG fragments. Table 3.5 lists eight MG entries that are used to derive English RCs. Abstract names such as “Noun” or “Vt” are used to avoid potential lexical bias in the modeling.

	Terminal	Syntactic feature	Note
1	Noun	N	Noun phrase
2	Noun	N, $-f$	Noun phrase with promotion feature $-f$
3	Det	=N, $-case$	Determiner Phrase
4	<b>who</b>	=N, $+f$ , <b>D</b> , $-case$ , $-wh$	$+f$ promotes NP to form a DP with $-wh$ feature
5	Vt	=D, $+case$ , V	verb selecting a DP as object
6		=>V, =D, v	verb selecting a DP as subject
7		=v, $+case$ , T	tense
8		=T, $+wh$ , Rel	$+wh$ feature hoisting DP

Table 3.5: Sample of MG lexical entries under the promotion analysis for English relative clauses

Among those eight lexical entries, the relative pronoun “who” has a  $+f$  feature that allows it to promote a noun to its left and to form a relativizable DP “noun who”, marked by the  $-wh$  feature. This  $wh$ -element will eventually be moved to the Spec of RelP, as illustrated in Figure 3.3. This same notation, in which a minus comes before a movement-licensing  $wh$  feature, figures prominently in Chapters 4 and 5 where fragments of Italian and Chinese play a key role in accounting for the processing difficulty profile of relative clauses in each language.

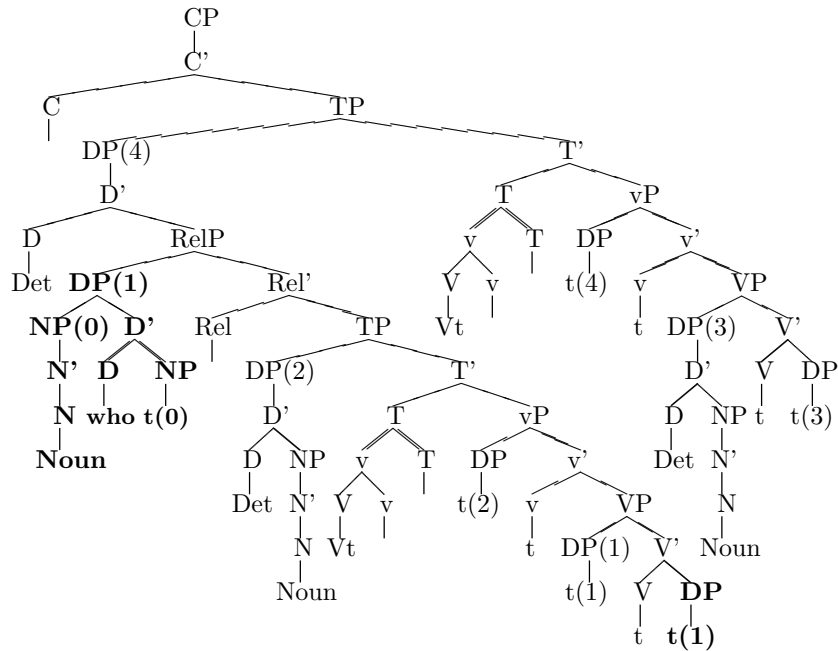


Figure 3.3: An English object relative clause derived by the promotion MG fragment

Wh-elements can be relativized from either subject or object position. Figure 3.4 shows a third alternative where the DP is extracted from the PP to the left edge of RelP. The rule that triggers this wh-movement is the same as the one used in deriving SRs or ORs, as long as the wh-element carries a *-wh* feature. Note that the grammar in its current form only considers NP-extraction. In this way, the preposition “to” remains inside of an RC, i.e. this is a case of P-stranding. The MG framework allows moving the preposition together with the noun, which will be implemented in the treatment of the Italian relative pronoun “cui” in Chapter 4.

To model the animacy effect in English RC processing, a grammar specifying the animacy of noun phrase is needed. This animacy-encoded grammar has dif-

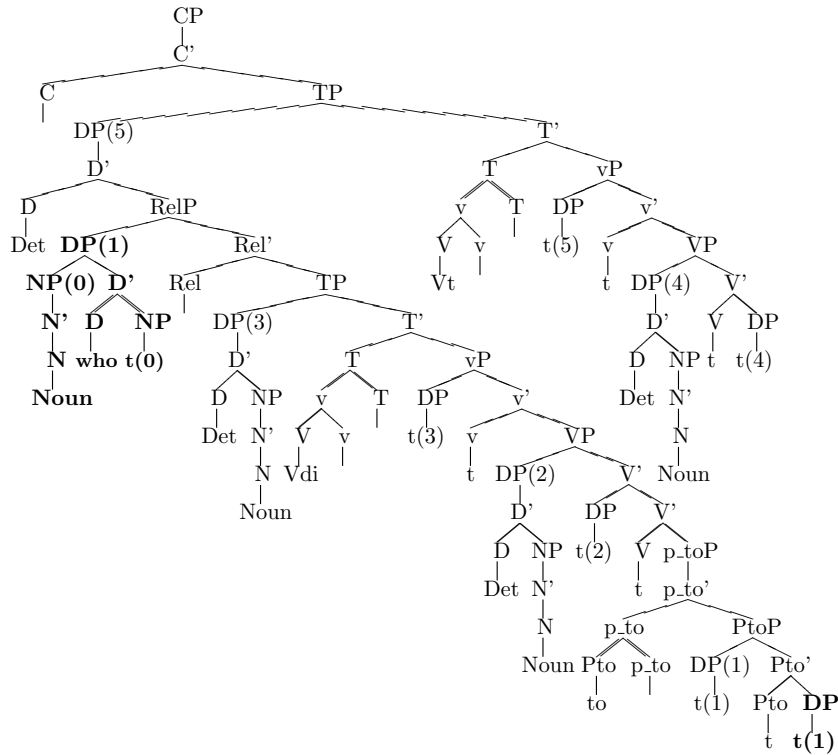


Figure 3.4: The MG fragment allows the relativization of indirect object with P-stranding.

ferent lexical entries for animate nouns versus inanimate nouns, as highlighted by the terminal nodes in Figure 3.5. The grammar also differentiates RC verbs by extraction-site, i.e. “VSR” and “VOR”. This is a case of grandparent annotation in the sense of Johnson (1998) to ensure that fine-grained probabilistic information can be captured in the grammar weighting stage, especially when the animacy of noun phrases interacts with the RC type. Using one category “V” instead of two categories “VSR” and “VOR” obscures this sort of distributional difference.

CCPC transforms MG derivations into multiple context-free rules in the



subject relative	6817
unreduced object relative	2043
reduced object relative	1658
relativization from adjunct	136
noun with determiner	28600
relative clause with determiner	5287
intransitive verb with pronoun subject	4108
transitive verb with pronoun subject	6845
intransitive verb with NP subject	2065
transitive verb with NP subject	4204
transitive verb with NP object	7101
ditransitive verb with NP object	348
transitive verb with pronoun object	4389
ditransitive verb with pronoun object	240
NP as indirect object	3358
pronoun as indirect object	568
transitive verb in OR	2973
ditransitive verb in OR	259
embedded pronoun subject in OR	1787
embedded NP subject in OR	1423

Table 3.6: Attestation counts from the Penn Treebank (Brown and WSJ corpora) are used to weight the grammar for English RCs

grammar. This corpus search is conducted using the animacy-annotated hand-parsed Switchboard corpus of conversational American English (Zaenen et al., 2004). This annotated section consists of about 110,000 sentences with about 300,000 NPs. Zaenen et al. (2004) define an animacy hierarchy with ten classes including HUMAN, ORG (organizations), ANIMAL, MAC (automata), VEH (vehicles), PLACE, TIME, CONCRETE (other physical objects), NONCONC (abstract entities), and MIX (NPs describing heterogeneous groups of entities). For the simplicity of modeling, those ten classes are grouped into a binary classification, i.e., animate or inanimate.<sup>1</sup> Following Bowman and Chopra (2012),

<sup>1</sup>See Section 1.3 for the discussion on the Animacy Hierarchy (Silverstein, 1976; Dixon, 1979).

HUMAN, VEH, MAC, ORG, and ANIMAL are considered animate, and the remaining classes inanimate.

Earlier corpus study by Roland et al. (2007) also considered the frequency of noun phrase animacy in relative clauses. However, they only hand-coded a random sample of 100 relative clauses each from the Brown and Switchboard corpora for animacy. Zaenen et al.'s work makes it possible to conduct finer-grained investigation on how animacy distributes in all kinds of syntactic relationships. Table 3.7 records a part of construction counts from the Switchboard corpus used in this modeling work.

SR with animate head	499
OR with animate head	133
SR with inanimate head	727
OR with inanimate head	1074
embedded animate object	59
embedded inanimate object	124
embedded pronoun object	82
matrix animate object	4660
matrix inanimate object	14914
matrix pronoun object	7130
animate indirect object	13
inanimate indirect object	77
pronoun indirect object	9
animate NP within an adjunct	11219
inanimate NP within an adjunct	13427
inanimate subject with intransitive verb	1945
inanimate subject with transitive verb	2186
animate subject with intransitive verb	11002
animate subject with transitive verb	8931
intransitive verb with an adjunct	13509
transitive verb with an adjunct	8557
transitive verb without an adjunct	35281
reduced OR with animate head	71
reduced OR with inanimate head	474

Table 3.7: Attestation counts from the Switchboard corpus are used to weight the animacy grammar for English RCs

Corpus counts in Table 3.7 confirm several claims about English RCs. Subject relatives tend to have animate head nouns while the head of object relatives are more likely to be inanimate. Pronouns are treated separately from animate or inanimate nouns in the corpus study. Table 3.7 suggests that pronouns are even more probable than animate nouns as the embedded object. In addition, it is more likely for an animate subject to have an intransitive verb, although it is not the case when the subject is inanimate.

### **3.3.2 English prediction I: the Subject Advantage**

The prepared English grammar fragment weighted by frequencies of constructions allows CCPC system to calculate ER predictions for standard English RCs in Example (3.1). The critical region is the three words after the relative pronoun, i.e. verb, determiner and noun in SRs and determiner, noun and verb in ORs.

Figure 3.6 plots ER degree at each word in English RCs. The big bump at the embedded verb in the OR condition indicates a predicted processing delay. Table 3.8 compares ER predictions directly with reported experimental observations. The surprisal modeling result of Hale (2001) is also listed. Recall that Grodner and Gibson (2005) only found the subject advantage at the RC verb position whereas eye-tracking data of Staub (2010) show evidence for the additional processing delays at both the noun phrase and the verb within the OR region. Early modeling work using surprisal and ordinary context-free grammars was able to correctly predict the subject advantage in English RCs but only at the embedded noun phrase. ER results of subject and object asymmetry are

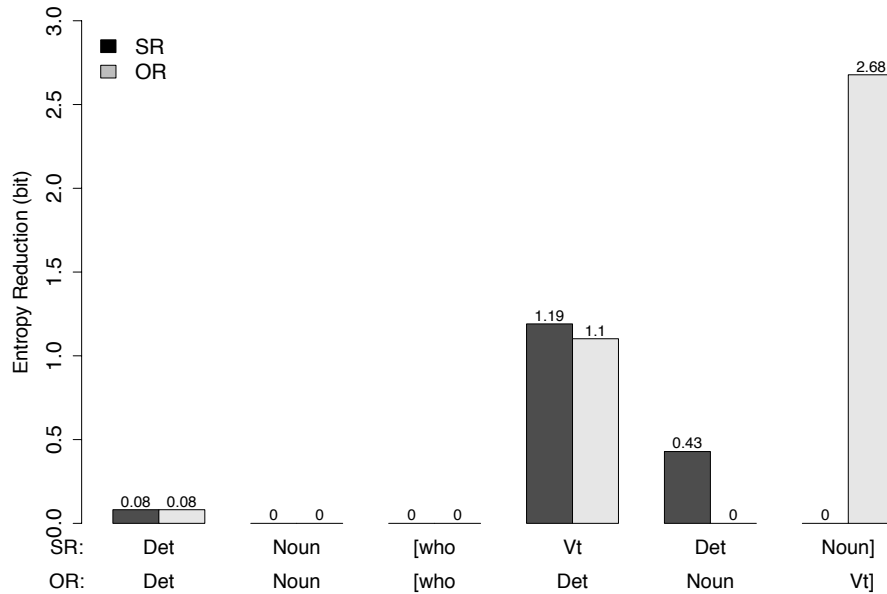


Figure 3.6: Predicted reading difficulties of English subject and object relative clauses

listed in the bottom row. At the determiner of embedded noun phrase, ORs are predicted to be harder than SRs (SR: 0.43 bit vs OR 1.1 bit). The subject advantage is more pronounced later at the RC verb (SR: 1.19 bit vs OR 2.68 bit). The predicted subject advantage<sup>2</sup> throughout the RC region is therefore compatible with the results of Staub (2010).

As discussed in Chapter 2, the advantage of the CCPC system is that it can enumerate all legitimate syntactic alternatives given a prefix string. This is useful in arriving at a linguistic interpretation of this disambiguation work. CCPC is capable of illustrating how structural uncertainties fluctuate during the transitions from one word to the next. The subject and object asymmetry in English

<sup>2</sup>The modeling in this dissertation is not the first attempt of using ER to predict the processing of English RCs. However, the calculated ER values based on a probabilistic context-free grammar only predicts a late effect of subject advantage (c.f. discussion above in Section 3.2). Object relatives are harder at the embedded verb (Hale, 2003).

	Type	Relative pro	Det	Noun	RC verb
Self-paced RT (ms) (Grodner & Gibson, 2005)	SR	349.8	334.3	384	<b>354.8</b>
	OR	343	348.1	357.6	<b>422</b>
Total eye-fixations (ms) (Staub, 2010)	SR	283	<b>272</b>	<b>375</b>	<b>333</b>
	OR	303	<b>382</b>	<b>459</b>	<b>420</b>
Surprisal (Hale, 2001)	SR	1	<b>1.6</b>		0.2
	OR	1	<b>4.53</b>		0
Entropy Reduction (bit) this study	SR	0	<b>0.43</b>	0	<b>1.19</b>
	OR	0	<b>1.1</b>	0	<b>2.68</b>

Table 3.8: ER predicts additional processing difficulties at both the noun phrase and the embedded verb in English object relatives, consistent to the eye-tracking data of Staub (2010). Observed or predicted subject advantages are highlighted in bold.

relatives serves as the first demonstration of this feature.

Evidence from Grodner and Gibson (2005) and Staub (2010) both suggest that the subject advantage exists at the embedded verb. The surprisal modeling fails to predict such an effect at this position (Hale, 2001; Levy, 2008). ER work by Hale (2003) accurately predicted the additional processing burden at the OR verb and attributed it to the syntactic disambiguation. In particular, after seeing the OR verb, the comprehender would know that no recursive modification of the embedded noun phrase will be allowed. The present ER modeling provides a second explanation based on the transitivity status of the embedded verb, as shown in Figure 3.7.

Figure 3.7 illustrates two sets of syntactic derivations given the prefix strings “Det Noun who Det Noun” and “Det Noun who Det Noun Vt”. Those syntactic alternatives are ranked by their own conditional probabilities. Each box represents the parsing state at a given point in the sentence, in this case, before and after the critical word, the embedded verb. The calculated Entropy value

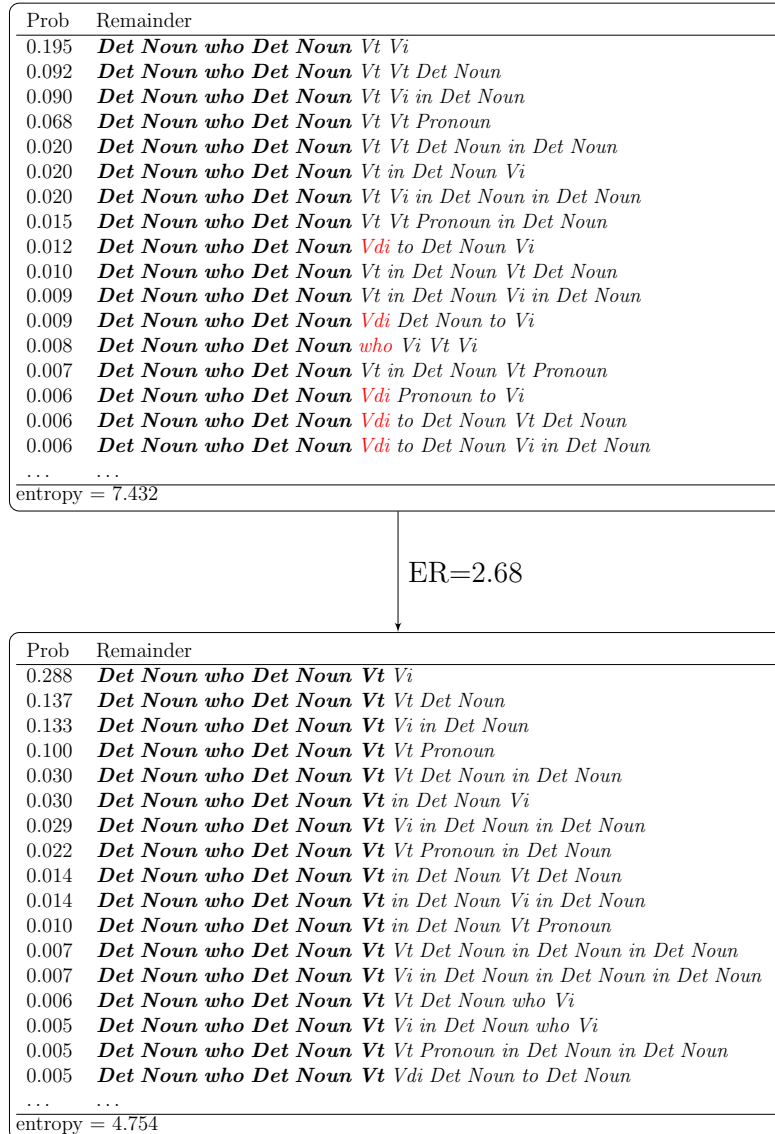


Figure 3.7: The transition at the embedded verb in English ORs

is larger in the top box (7.432 bits). Before the embedded OR verb, the uncertainty degree is high because there exist syntactic derivations other than OR with a transitive verb. As highlighted in red, the prefix in the top box could also continue with a ditransitive verb followed by a “to-phrase”. Reading in the em-

bedded OR verb disconfirms this possibility and the uncertainty degree drops dramatically to 4.754 bits. Therefore, the transition from the embedded subject to the embedded verb in English ORs results in an larger ER value and mirrors the observed experimental findings <sup>3</sup>.

In contrast, Figure 3.8 suggest that the transition from the relative pronoun to the embedded SR verb is less taxing with an ER of 1.19 bits, mostly because the SR prefix “Det Noun who Vt” is still ambiguous. The right box in Figure 3.8 shows that the parser state at that point is quite uncertain (ER 7.343 bits). Comprehenders still need to choose between a determiner phrase or a pronoun as the embedded object. Distinguishing pronouns from other noun phrases is non-trivial as the frequency distribution of pronoun have been found to influence the processing pattern of English RCs (Real & Christiansen, 2007).

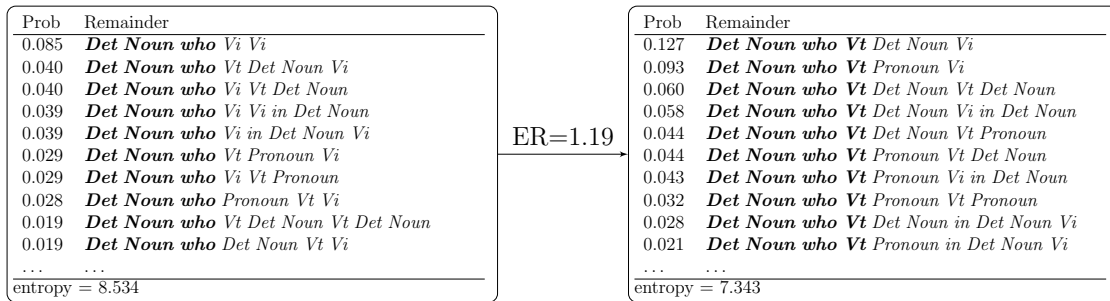


Figure 3.8: The transition at the embedded verb in English SRs

Consistent with what the surprisal has predicted, ER comparison between SRs and ORs in Table 3.8 also shows a subject advantage at the embedded noun phrase, more precisely, at the NP determiner. The underlying reasons for the contrast at this position are similar to the surprisal account such that after en-

<sup>3</sup>It is safe to say that the choice of implementing the verb class using three notations, i.e., “Vt”, “Vi”, or “Vdi”, instead of one “V”, will not influence the ER modeling results. CCPC calculates the conditional probability for each syntactic derivation rather than for each surface string. Changing the surface string notation alone will not alter how structures are derived.

countering the relative pronoun “who”, the OR analyses are less expected than the SR analyses. ER explains this contrast from the disambiguation perspective. The OR prefix is more disambiguated at the determiner. In the left box of Figure 3.9, the highlighted OR derivations are ranked low in the list. Therefore, more disambiguation efforts are made when processing the following determiner since it is part of an OR, a less frequent structure.

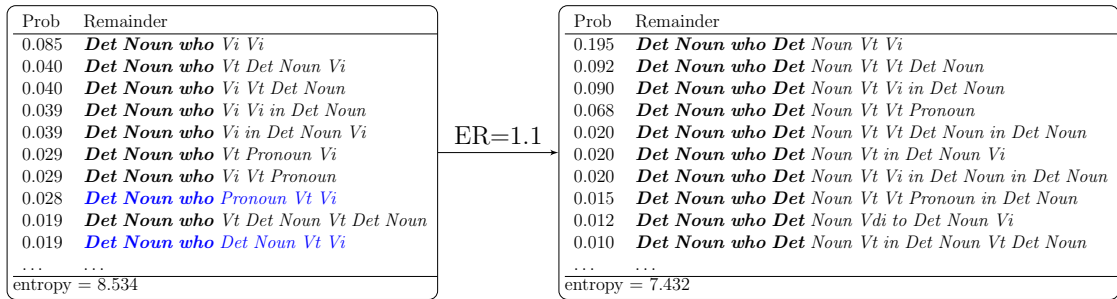


Figure 3.9: The transition at the NP determiner in English ORs

The ER modeling above is suffice to derive the Subject Advantage at both the verb and noun phrase within an RC, a result that precisely mirrors Staub’s recent eye-tracking data. It shows that there is no need to explain the subject-object asymmetry with a “two-way” approach including both DLT and Surprisal. Section 3.3.3 tests the ER metric with another prominence effect in English RC processing: the animacy of the head noun.

### 3.3.3 English prediction II: the Animacy Effect

The subject advantage is not the only processing effect that ER can predict for English RCs. As discussed extensively in Section 3.2, previous literature has identified the animacy of noun phrases as a crucial cue in RC comprehension.

This section reports two modeling works on this issue in order. The results will be compared against observed experimental data in Traxler et al. (2002, 2005) and Lowder and Gordon (2012).

As Section 3.2.2 explains, one observation in Traxler et al. (2002, 2005) was that ORs with inanimate head nouns were read faster than those with animate heads. Therefore, the processing difficulty difference between SRs and ORs reduces when the sentence starts with an inanimate noun phrase. This result is based on a weighted grammar fragment annotated with NP animacy. Table 3.9 directly compares ER predictions with published data in Traxler et al. (2002). Both of them cover the three-word RC region.

	Type	Head-NP	Embedded-NP		Diff.
Eye-tracking Total time (ms) (Traxler et al., 2002)	SR	+anim	-anim	881	<b>294</b>
	OR	+anim	-anim	1175	
	SR	-anim	+anim	847	<b>39</b>
	OR	-anim	+anim	886	
Entropy Reduction (bit)	SR	+anim	-anim	2.04	<b>1.01</b>
	OR	+anim	-anim	3.05	
this study	SR	-anim	+anim	3.72	<b>-0.73</b>
	OR	-anim	+anim	2.99	

Table 3.9: ER predicts that the subject advantage (OR minus SR) disappears in the RC region when an English OR has an inanimate head noun.

When the head noun is animate, a subject advantage is predicted. The ER difference between SRs and ORs is 1.01 bits. However, when the head noun becomes inanimate, the subject advantage is much reduced in terms of total reading times (from 294 ms to 39 ms). ER predicts the fact that SRs are harder than ORs in the RC region (with a difference of -0.73 bits). The reversed pattern predicted here mirrors the eye-tracking results of readers with high working

memory capacity, as shown in Figure 3.1. This is not a surprising prediction since, to some extent, people with high WMC are able to maintain more derivations simultaneously in the memory, which is similar to the parallel processing that CCPC implements.

To understand why SRs with inanimate head are so difficult to process, Table 3.10 examines the exact location where the Object Advantage occurs. Comparing four conditions with different animacy patterns, it seems that the disappeared subject advantage is most likely driven by the higher ER value at the embedded SR verb (1.86 bits). With a frequency-based approach in mind, this effect is intuitive. Since subjects tend to be animate rather than inanimate, integrating an inanimate subject with the embedded verb requires more cognitive efforts.

RC prefix	Type	Det	InaniN	RC verb
Det AniN who	SR		1.0	<b>1.04</b>
	OR		0.93	2.12
		Det	AniN	
Det InaniN who	SR		1.86	<b>1.86</b>
	OR		1.71	1.28

Table 3.10: ER predicts that the reduced subject advantage in RCs with inanimate heads is driven by the additional reading difficulty at the embedded SR verb.

However, when exploring the sorted lists of “in-play” remainders, a more detailed explanation presents itself. Figure 3.10 illustrates two parsing states with different RC prefixes. The Entropy is higher in the bottom one with an inanimate head noun because subject relatives and object relatives are both considered as possible alternatives. On the contrary, when the head noun preceding the relative pronoun “who” is animate, only two OR variants are ranked among the top ten derivations with highest probabilities. Since the inanimate RC pre-

fix is highly ambiguous, integrating it with the embedded verb requires more processing effort.

Prob	Remainder
0.106	<i>Det AniN who Vt Det InaniN Vi</i>
0.078	<i>Det AniN who Vt Pronoun Vi</i>
0.056	<i>Det AniN who Vt Det AniN Vi</i>
0.055	<i>Det AniN who Vi Vi</i>
0.026	<i>Det AniN who Vt Det InaniN Vt Det InaniN</i>
0.020	<i>Det AniN who Vt Pronoun Vt Det InaniN</i>
0.019	<i>Det AniN who Det AniN Vt Vi</i>
0.018	<i>Det AniN who Pronoun Vt Vi</i>
0.017	<i>Det AniN who Vt Det InaniN Vdi Det AniN to Det InaniN</i>
0.015	<i>Det AniN who Vt Det InaniN Vdi Det InaniN to Det InaniN</i>
...	...
entropy = 7.929	

Animate prefix

Prob	Remainder
0.047	<i>Det InaniN who Vt Det InaniN Vi</i>
0.047	<i>Det InaniN who Det AniN Vt Vi</i>
0.046	<i>Det InaniN who Pronoun Vt Vi</i>
0.035	<i>Det InaniN who Vt Pronoun Vi</i>
0.034	<i>Det InaniN who Vi Vi</i>
0.025	<i>Det InaniN who Vt Det AniN Vi</i>
0.016	<i>Det InaniN who Vt Det InaniN Vt Det InaniN</i>
0.016	<i>Det InaniN who Det AniN Vt Vt Det InaniN</i>
0.016	<i>Det InaniN who Pronoun Vt Vt Det InaniN</i>
0.012	<i>Det InaniN who Vt Pronoun Vt Det InaniN</i>
...	...
entropy = 9.092	

Inanimate prefix

Figure 3.10: An inanimate RC prefix is more ambiguous than an animate RC prefix because of more high-ranking OR alternatives.

ER accurately predicts the animacy effect at the English RC head. As shown in Figure 3.10 and discussed in greater detail in Section 2.2, the remainder sets lead us to a properly linguistic interpretation of the disambiguation work at the heart of this animacy effect. On the same track, the work done by Lowder and Gordon (2012) is another perfect test for us to further evaluate the relia-

bility of CCPC. Their experimental design was quite simple. As described in Example (3.4), SRs were tested against simple sentences that share the common prefix in animacy. The idea was that the animacy preference for subject-verb relationship can be reduced when the subject and the verb locates in two different clauses. Indeed, they found that the complementizer “that” in between an inanimate NP and the verb reduces the animacy effect at the verb since “that” creates a clausal boundary between the two, as shown in Figure 3.2 on Page 32.

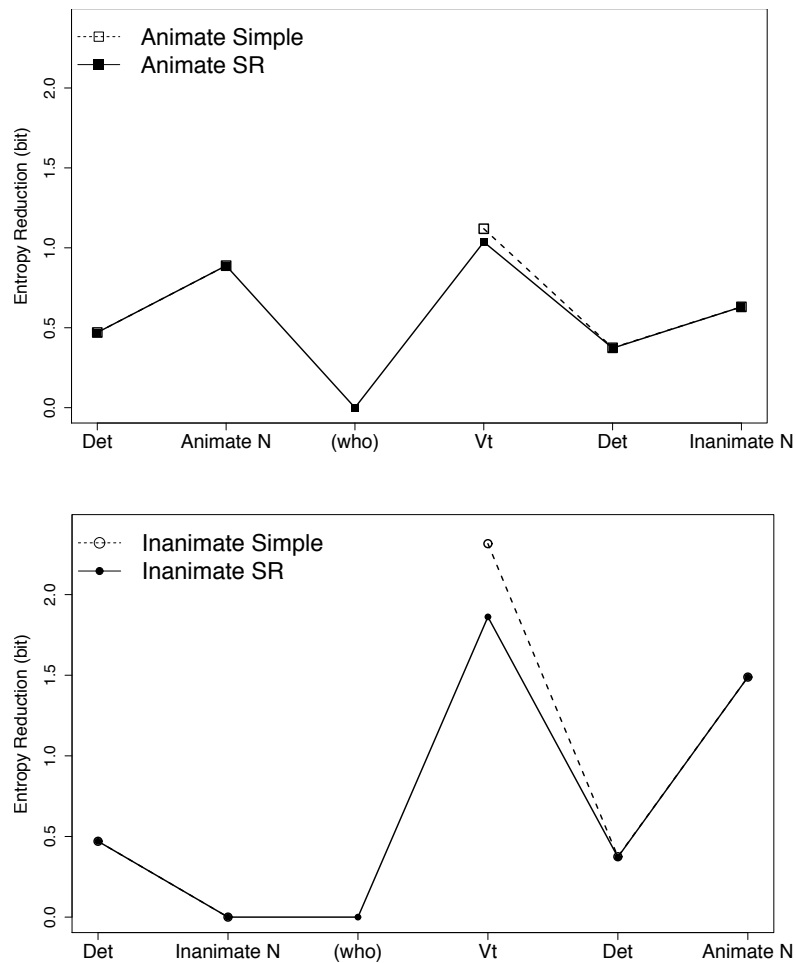


Figure 3.11: ER predicts that the processing difficulty of having a inanimate subject is reduced by the clausal boundary (Lowder & Gordon, 2012).

ER predicts a similar effect at the verb in inanimate conditions. As shown in Figure 3.11, ER value for verbs (“Vt”) are higher in simple sentence conditions. The ER difference between the SR and its counterpart simple sentence is amplified given an inanimate subject.

As part of the modeling results reported earlier in this section, the transition from the inanimate RC prefix “Det InaniN who” to the next “Det InaniN who Vt” has a ER value of 1.86 bits (see Table 3.10). But the reason that the subject-verb integration have an even greater ER in simple sentences is still unknown. CCPC provides a possible answer in Figure 3.12. The prefix with an inanimate subject is very ambiguous because the transitivity of the forthcoming main verb is still undetermined. At the main verb, the frequency distribution of syntactic alternatives in the parser state are more concentrated. Specifically, the top three derivations in the right box have a total probability of 80%. The transition from a high uncertainty state to a much lower one results in a 2.32 bits reduction of entropy.

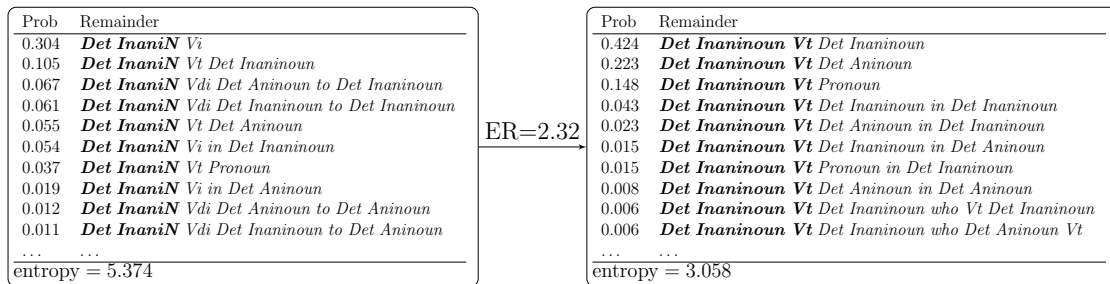


Figure 3.12: The inanimate subject prefix allows a variety of alternatives. The main verb greatly disambiguates the prefix.

To summarize, Chapter 3 examines a range of processing literature in English RCs. The ER model predicts the subject advantage at both the embedded NP and at the RC verb. In addition, two reported animacy effects have been

modeled. The ER hypothesis, along with the CCPC system, therefore takes the expectation-based processing theory one step further by integrating frequencies of non-syntactic cues with formal syntactic analyses, as tested by the successful modeling of English RC processing.

## CHAPTER 4

### ITALIAN RELATIVE CLAUSES

Chapter 2 shows that by enumerating all possible syntactic alternatives at a given word, along with their probabilities, CCPC system is able to capture how uncertain a reader is during sentence processing using the information-theoretic notion Entropy. The numerical pattern of uncertainty reduction indeed mirrors reading time delays that have been observed in published experimental work. Chapter 3 models the Subject Advantage observed in the English RC processing literature. In addition, using an weighted grammar enriched with animacy information, the modeling results also confirm the reading difficulty pattern when the animacy of noun phrases are manipulated (Traxler et al., 2002).

This chapter turns to a different language, Italian. As discussed in Chapter 1, the competition model of MacWhinney et al. (1984) predicts that Italian speakers are more likely to rely on animacy information than English speakers. Since Traxler et al. (2002) have already identified the important role that animacy plays in RC processing, it would be interesting to see whether the animacy effect also applies in Italian, even with a larger scale. However, there exists experimental literature on Italian RCs which suggests differently. In particular, (Belletti & Chesi, 2011) report that Italian speakers do not rely on animacy information to produce Italian ORs. Section 4.2 provides a survey of previous experimental works. Section 4.3 reports the modeling of Italian RC processing taking published results as the ground truth. As in the English study of Chapter 3, this modeling of the Italian RC processing-difficulty profile focuses on two issues: the subject-object asymmetry and the influence of head-noun animacy.

## 4.1 Italian relative clauses

Italian relative clauses, similar to their English counterparts, are post-nominal, i.e., the head noun and the relative pronoun precede the relative clause. In Italian conversations, most relative clauses are introduced by the relative pronoun *che*<sup>1</sup> and its variant *cui*. The relative pronoun *che*, usually immediately following the head noun, functions either as the subject or the direct object of the verb. The head noun phrase which *che* refers to can be animate or inanimate (Maiden & Robustelli, 2007). Example 4.1 below presents a pair of Italian RCs with the relative pronoun *che*.

(4.1) a. Italian Subject relative with *che*

*Il ragazzo<sub>i</sub> [ che e<sub>i</sub> guarda il pagliaccio ] baciò*  
The boy [ CHE e<sub>i</sub> watch-3SING the clown ] kissed  
*la ragazza*  
the girl  
'The boy who is watching the clown kissed the girl.'

b. Italian Object relative with *che*

*Il ragazzo<sub>i</sub> [ che il pagliaccio guarda e<sub>i</sub> ] baciò*  
The boy [ CHE the clown watch-3SING ] kissed  
*la ragazza*  
the girl  
'The boy who the clown is watching kissed the girl.'

In both (4.1a) and (4.1b), the head noun phrase *il ragazzo* "the boy" is relativized from inside the RC. There is an empty category  $e_i$  co-indexed with the head noun. Similar to its English counterpart, more material intervenes between the head noun and the empty category in Italian SRs than in ORs.

---

<sup>1</sup>*Che* and *cui* can be replaced by *il quale* or *la quale* depending on number and gender.

The second relative pronoun *cui*, on the other hand, must be used after a preposition. (4.2) provides two examples of this kind of RC in which *cui* follows the preposition *da* and *di* respectively.

- (4.2) a. *Il luogo da cui venivi era lontano*  
The place from CUI you-were was away  
*'The place from which you came was distant.'*
- b. *Del dialogo di cui oggi si vedono i risultati ...*  
The dialogue of CUI today you see the result-pl ...  
*'The dialogue of which you see the result today ...'*

Both *che* and *cui* relatives are implemented by the MG grammar fragments, along with other relevant construction that could play a role in processing a Italian RC prefix string. Section 4.3 will describe the grammar engineering efforts in details. But before that, Section 4.2 reviews a number of important experimental studies of Italian RCs.

## 4.2 Experimental studies of Italian RC processing

### 4.2.1 The Subject-Object Asymmetry

Relative clauses in Italian, just like in many other languages, have been studied extensively in the experimental literature. One consensus of those works is that Italian object relatives are both difficult to comprehend and to produce. This observation has been confirmed in experiments targeting both children and adults (Arosio, Adani, & Guasti, 2009; Belletti & Contemori, 2009; Di Domenico & Di Matteo, 2009; Adani, van der Lely, Forgiarini, & Guasti, 2010; Contemori &

Belletti, 2011). In addition, previous literature also looks into how different linguistic cues facilitate the processing of Italian RC, including animacy (Arosio et al., 2011; Belletti & Chesi, 2011), gender (Adani et al., 2010; Belletti, Friedmann, Brunato, & Rizzi, 2012) and number (Adani et al., 2010) agreements on the verb.

Among those studies, Di Domenico and Di Matteo (2009) explored the parsing strategies that derive the subject-object asymmetry in Italian RC comprehension. They tested four different types of *che* relatives. In addition to the two RC types exemplified in (4.1), Di Domenico and Di Matteo introduced two other RC conditions (Examples 4.3b and 4.3c) with switched argument and verb word order within the embedded clause. This is possible in Italian if the number agreement between the verb and the subject is carefully manipulated.

(4.3) a. Subject RC

*Questo è il poliziotto<sub>i</sub> che e<sub>i</sub> ha seguito i ladri*  
 This is the policeman that **e<sub>i</sub> has** pursued the robbers  
 'This is the policeman that pursued the robbers.'

b. Object RC with post-verbal subject

*Questo è il poliziotto<sub>i</sub> che hanno seguito i ladri e<sub>i</sub>*  
 This is the policeman that **have** pursued the robbers **e<sub>i</sub>**  
 'This is the policemen that the robbers pursued.'

c. Subject RC with pre-verbal object

*Questo è il poliziotto<sub>i</sub> che e<sub>i</sub> i ladri ha seguito*  
 This is the policeman that **e<sub>i</sub> the robbers** **has** pursued.  
 'This is the policeman that pursued the robbers.'

d. Object RC

*Questo è il poliziotto<sub>i</sub> che i ladri hanno seguito e<sub>i</sub>*  
 This is the policeman that the robbers **have** pursued **e<sub>i</sub>**  
 'This is the policemen that the robbers pursued.'

In (4.3), Conditions a and d correspond to the regular SR and OR in (4.1). The word order within the embedded clause remains Subject-Verb-Object. The embedded subject in Condition a and the embedded object in Condition d have undergone relativization. Compare them with the two new types of RC, the embedded subject *i ladri* “the robbers” in (b) is post-verbal which creates a word order of Verb-Subject-Object. Condition c has a preverbal embedded object. In this case, it is Subject-Object-Verb within the RC. Object RCs with post-verbal subject are only possible under certain condition. The number feature of the verb agrees with the embedded subject but not with the relativized object, namely in (b) the plural auxiliary verb *hanno* “have” is incompatible with the singular RC head *poliziotto* “policeman”. Similarly, when the first noun phrase within the RC is plural and the embedded verb is singular, that pre-verbal argument will be identified as the object, as in (4.3)c. This experimental design creatively forms two pairs of RCs, such that only one word differs between members of each pair – namely, the embedded verb.

Di Domenico and Di Matteo (2009) conducted two self-paced reading experiments to test those four conditions separately. Figure 4.1 found the subject advantage both at the embedded verb and the embedded object between (4.3)a and (4.3)b. The processing delay at the first word after “che” (W4) in Condition b (OR with post-verbal subject) reflects a structural reanalysis motivated by the number mismatch between the head noun and this auxiliary verb in number. At W6, Condition b was also read slower which could be due to the extra memory burden when integrating the filler and the gap.

The results of their experiment 2 is shown in Figure 4.2. No subject advantage is found. In fact, SRs were read much slower than ORs at W6. Di Domenico

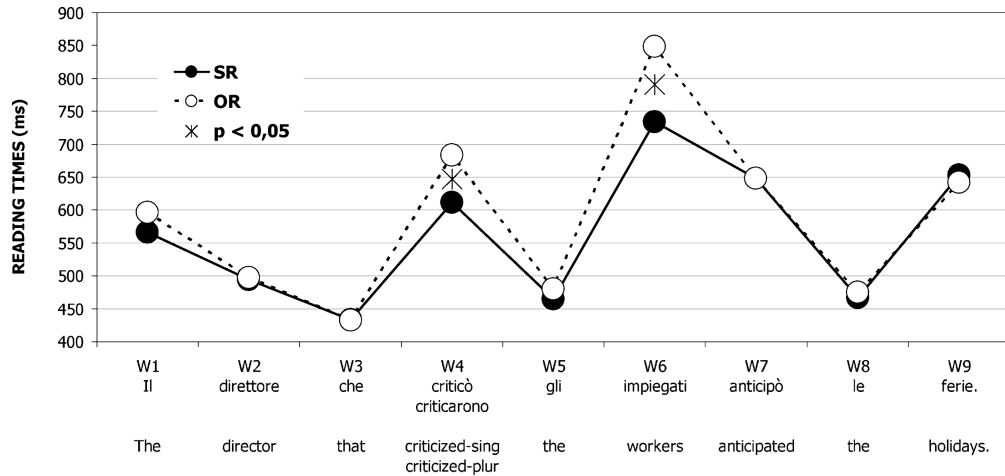


Figure 4.1: Self-paced reading times of Italian RC conditions (4.3)a-b in experiment 1 of Di Domenico and Di Matteo (2009)

and Di Matteo (2009) explained the effect at W6 using the Active Filler Theory (Clifton Jr & Frazier, 1989; Frazier & Flores D’Arcais, 1989; De Vincenzi, 1991) such that when the filler looks for its landing site (or GAP) at the embedded verb, the sentence structure in SRs (condition c) must be revised when the verb number information clarifies the grammatical function. This clearly is a plausible explanation. However, the possibility of memory-based approaches should not be excluded by the OR advantage seen at W6, simply because the structural reanalysis could well cause much longer reading delays in the SR condition, which overshadows the additional integration costs in ORs.

ER work in this chapter does not model the two additional conditions, i.e., (4.3)b and c. To compare the results directly with RCs in other languages and for the simplicity of animacy modeling, it instead focuses on the RC types with the regular Subject-Verb-Object sequence, as in (4.3)a and d. Table 4.1 summarizes the word-by-word reading times of the three-word RC region. Those numbers

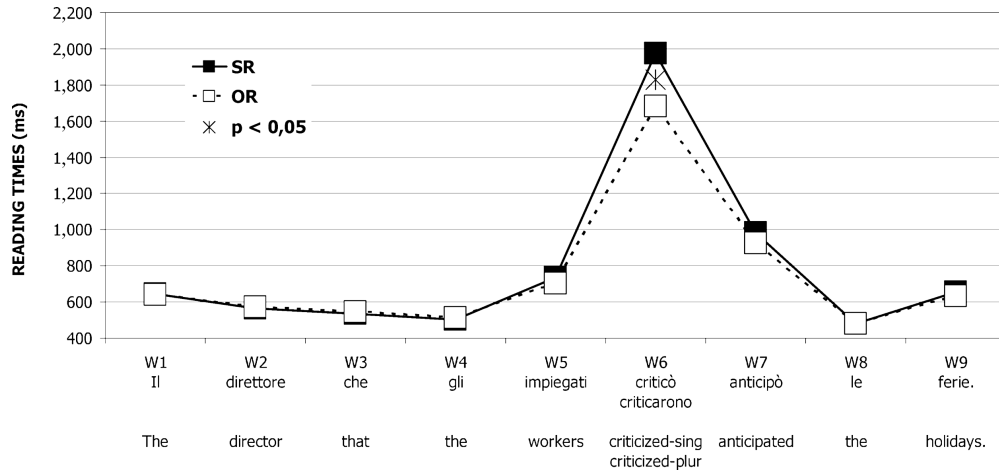


Figure 4.2: Self-paced reading times of Italian RC conditions (4.3)c-d in experiment 2 of Di Domenico and Di Matteo (2009)

	W4	W5	W6
SR	610	470	740
OR	500	700	1650

Table 4.1: Approximate reading times (ms) of the RC region in Italian SRs and ORs (Conditions (4.3)a and d), adopted from Di Domenico and Di Matteo (2009)

are read from the Figures 4.1 and 4.2. At W4, ORs were read slightly faster than SRs. At the following two words, W5 and W6, a robust subject advantage was found.

## 4.2.2 The effect of animacy

In contrast to the consistent subject advantage found in Italian experimental studies, the research on animacy shows some contradictory results. Two representative studies are discussed below in order. First, Arosio et al. (2011) asked 51

Italian-speaking children to participate in a self-paced listening task with four RC conditions, as listed in (4.4).

(4.4) Four Italian RC conditions in Arosio et al. (2011)

a. Subject RC with animate embedded object (SR++)

*Il pasticcere osserva i gatti che stanno rincorrendo il  
The baker watches the cats that are chasing the  
topolino  
mouse  
'The baker watches the cats that are chasing the mouse.'*

b. Subject RC with inanimate embedded subject (SR+–)

*Il pasticcere osserva i gatti che stanno rincorrendo il  
The baker watches the cats that are chasing the  
gomitolo  
ball-of-yarn  
'The baker watches the cats that are chasing the ball of yarn.'*

c. Object RC with post-verbal animate embedded subject (OR++)

*Il pasticcere osserva il topolino che stanno rincorrendo i  
The baker watches the mouse that are chasing the  
gatti  
cats  
'The baker watches the mouse that the cats are chasing.'*

d. Object RC with post-verbal inanimate embedded subject (OR–+)

*Il pasticcere osserva il gomitolo che stanno rincorrendo  
The baker watches the ball-of-yarn that are chasing  
i gatti  
the cats  
'The baker watches the ball of yarn that the cats are chasing.'*

Among the four conditions, there are two SRs and two ORs. Both ORs have post-verbal embedded subjects. The RC subject for all four conditions are animate while the animacy of the RC object varies. We use the abbreviation “SR++”

(4.4a) to represent subject RCs with animate head and animate embedded object. Similarly, “OR-+” (4.4d) stands for object relatives with inanimate head and animate embedded subject. Arosio et al reported a general subject advantage across four conditions, Italian-speaking children spent less time listening to SRs than ORs. More importantly, they found that the subject advantage was most significant at the gerundive verb *rincorrendo* “chasing”. This was driven by the longer listening time of object relatives with an animate object (4.4c: OR++), i.e. an animate head noun.

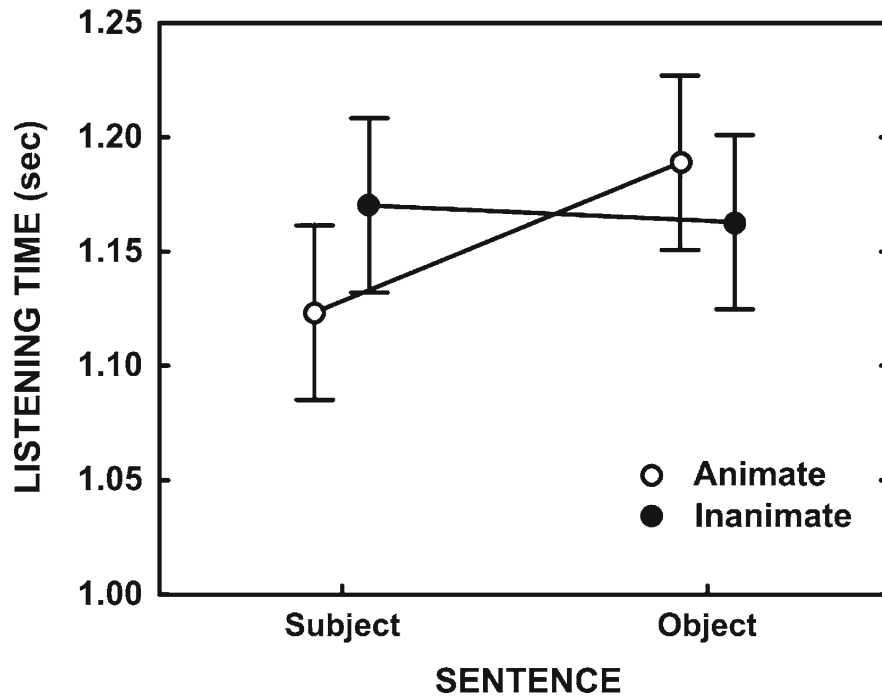


Figure 4.3: The subject advantage increases when ORs have animate heads (Arosio et al., 2011).

The interaction between animacy and the RC type becomes clearer in Figure 4.3. When the RC object is inanimate (black circle), the processing of Italian SRs and ORs do not differ in terms of listening times ( $LT_{OR-+} - LT_{SR+-} = 0.008sec$ ). However, with an animate RC object (white circle), ORs were pro-

cessed much slower ( $LT_{OR++} - LT_{SR++} = 0.065sec$ ).

According to Arosio et al., Italian object relatives with animate head are processed slower. To some extent, this conclusion is expected because it is largely compatible with studies in other languages such as (Traxler et al., 2002; Mak et al., 2002). However, Belletti and Chesi (2011) reported a quite different result in a sentence production study. They asked a similar question: will the change of head noun animacy influence the production of Italian ORs? In both experiments in the study, subjects listened to a minimal pair of cue sentences and were asked to answer a question based on the knowledge of two situations described. This kind of question is called “elicitation question” because it serves only to elicit an utterance from the participant. These utterances are then analyzed to see whether their production is affected by cue sentences. Belletti and Chesi’s study included two experiments, each of which has four conditions. In Experiment 1, as shown below in (4.5), the verb in two elicitation sentences varies. There were three items per condition which led to 12 experimental items in total, along with 28 fillers interspersed between the experimental items.

(4.5) Experiment 1: verb change (Belletti & Chesi, 2011)

a. head NP [+animate]; embedded NP [+animate]

(1) Cue sentences:

I poliziotti salutano un ragazzo      ‘The policemen greet a child.’

I poliziotti rincorrono un ragazzo      ‘the policemen chase a child.’

(2) Elicitation question:

tu quale ragazzo vorresti incontrare?      ‘Which child would you rather meet?’

“vorrei incontrare il ragazzo ...”      ‘I would rather meet the child ...’

b. head NP [+animate]; embedded NP [-animate]

(1) Cue sentences:

I secchi sbilanciano un imbianchino    *'The buckets unbalance a decorator.'*

I secchi sporcano un imbianchino    *'The buckets dirty a decorator.'*

(2) Elicitation question:

Tu quale imbianchino vorresti aiutare?    *'Which decorator would you rather help?'*

“vorrei aiutare l'imbianchino ...”    *'I would rather help the decorator ...'*

c. head NP [-animate]; embedded NP [+animate]

(1) Cue sentences:

I giornalisti scrivono un articolo    *'The journalists write an article.'*

I giornalisti copiano un articolo    *'The journalists copy an article.'*

(2) Elicitation question:

Tu quale articolo vorresti leggere?    *'Which article would you rather read?'*

“vorrei leggere l'articolo ...”    *'I would rather read the article ...'*

d. head NP [-animate]; embedded NP [-animate]

(1) Cue sentences:

I camini riscaldano un appartamento    *'The fireplaces warm an apartment.'*

I camini affumicano un appartamento    *'The fireplaces smoke an apartment.'*

(2) Elicitation question:

Tu quale appartamento vorresti scegliere?    *'Which apartment would you rather choose?'*

“vorrei scegliere l'appartamento ...”    *'I would rather choose the apartment ...'*

In contrast, Experiment 2, shown in (4.6), manipulated the subjects of two elicitation sentences. 24 subjects were tested with the verb change elicitation condition and 28 subjects with the subject change elicitation condition.

(4.6) Experiment 2: subject change (Belletti & Chesi, 2011)

a. head NP [+animate]; embedded NP [+animate]

(1) Cue sentences:

I poliziotti rincorrono un ragazzo    *'The policemen chase a child.'*

I commercianti rincorrono un ragazzo    *'The shopkeepers chase a child.'*

(2) Elicitation question:

tu quale ragazzo vorresti incontrare? *'Which child would you rather meet?'*

"vorrei incontrare il ragazzo ..." *'I would rather meet the child ...'*

b. head NP [+animate]; embedded NP [-animate]

(1) Cue sentences:

I secchi sporcano un imbianchino *'The buckets dirty a decorator.'*

I pennelli sporcano un imbianchino *'The paintbrushes dirty a decorator'*

(2) Elicitation question:

Tu quale imbianchino vorresti aiutare? *'Which decorator would you rather help?'*

"vorrei aiutare l'imbianchino ..." *'I would rather help the decorator ...'*

c. head NP [-animate]; embedded NP [+animate]

(1) Cue sentences:

I giornalisti scrivono un articolo *'The journalists write an article.'*

I pubblicitari scrivono un articolo *'The publicists write an article.'*

(2) Elicitation question:

Tu quale articolo vorresti leggere? *'Which article would you rather read?'*

"vorrei leggere l'articolo ..." *'I would rather read the article ...'*

d. head NP [-animate]; embedded NP [-animate]

(1) Cue sentences:

I camini riscaldano un appartamento *'The fireplaces warm an apartment.'*

I termosifoni affumicano un appartamento *'The heaters warm an apartment.'*

(2) Elicitation question:

Tu quale appartamento vorresti scegliere? *'Which apartment would you rather choose?'*

"vorrei scegliere l'appartamento ..." *'I would rather choose the apartment ...'*

Most answers elicited were object relative clauses. Other produced structures include subject relatives and prepositional phrases. The authors counted the number of produced ORs. These elicited answers included not just regular active ORs as in (4.1) but also passive ORs. OR variants with post-verbal

Head NP	+anim	+anim	-anim	-anim
Embedded NP	+anim	-anim	+anim	-anim
OR	3	2	4	2
OR with post-verbal subject	12	2	4	5
OR with <i>pro</i> subject	8	3	4	4
Total	23	7	12	11

Table 4.2: Aggregated counts of active OR with different animacy patterns in two elicitation tasks of Belletti and Chesi (2011)

subjects or covert subjects (subject *pro*-drop) were also found and taken into consideration. Table 4.2 summarizes their results on active ORs.

Surprisingly, the number of active ORs with inanimate head noun was not bigger than those with animate heads. In fact, Italian speakers produced more ORs with animate head and animate subject (OR++) than any of the other three types. As Belletti and Chesi (2011) put it:

“It is clear that the animacy (mis)match does not play any role in favoring or disfavoring the production of (active) ORs, in the adopted experimental conditions.”

To conclude the experimental literature reviewed in this section, Italian RCs allow more flexibility than their English counterparts, for example, post-verbal subjects and pre-verbal objects are allowed under certain conditions. Despite this flexibility, similar to other languages, the subject advantage also holds in Italian RCs. On the other hand, whether the noun phrase animacy is a strong cue in Italian RC comprehension remains unclear. In particular, the sentence production results of Belletti and Chesi (2011) seem to contradict with the what was observed in the comprehension experiment (Arosio et al., 2011).

### 4.3 Modeling Italian RC processing

This section presents the modeling of Italian relative clauses with a goal of accounting for the subject-object asymmetry as well as the effect of animacy. Using formalized grammar fragments weighted by detailed corpus studies, CCPC examines all possible syntactic variants given a sentence prefix. Enumerating those structural alternatives helps us better understand the reading difficulties at each word, especially those driven by disambiguations.

#### 4.3.1 Italian grammars

The derivation of Italian relative clauses involves movement. Depending on the nature of moved element, syntacticians have proposed two major classes of analysis (Utzeri, 2007). The modeling work on Italian RCs described in this section implemented both of them within the MG framework. The first proposal, which we refer as “the adjunction analysis”, argues that the head noun adjoins the RC (Cinque, 1982; Chomsky, 1986, 1995)<sup>2</sup>. Figure 4.4 provides an OR example generated by the prepared MG under the adjunction analysis. It follows the word order of (4.1b) but only with abstract part of speech tags, as in “Det Noun Che Det Noun Vt Vt Det Noun”. The relative pronoun *che* (DRelP) moves from embedded object position to the Spec of CRelP. The CP subsequently adjoins to the head noun phrase as a postmodifier.

Table 4.3 lists a number of lexical entries in the adjunction MG which are

---

<sup>2</sup>Italian RCs, in some studies, are derived as a result of relative operator movement from the embedded position to the Spec of CP. Nevertheless, the head noun phrase adjoins the CP.

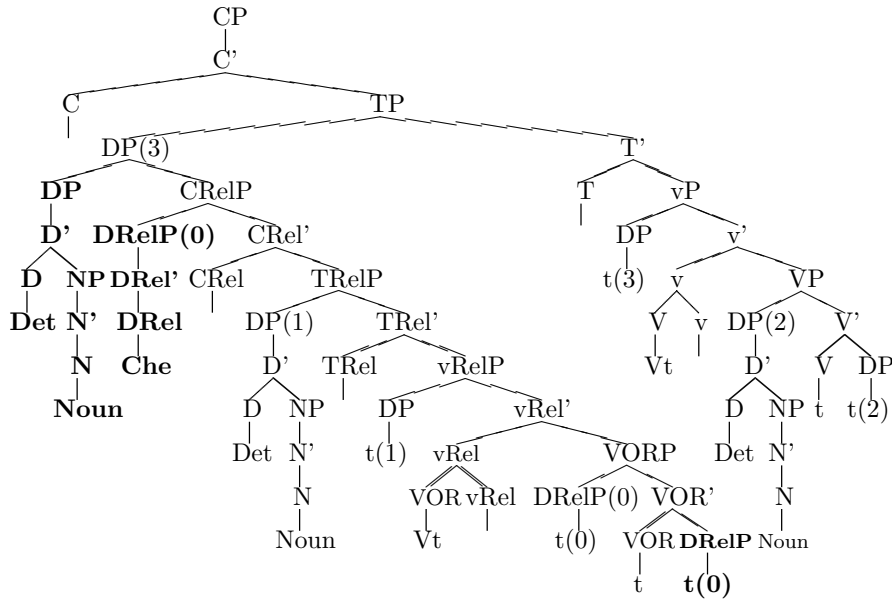


Figure 4.4: Italian object relative clause derived by an “adjunction” grammar

used to generate an OR tree like the one in Figure 4.4. The relative pronoun “che” has a *-wh* feature, which triggers *wh*-movement when it is checked by the *+wh* feature on line 7. Line 8 is the adjunction rule which links the RC with its head. The direction of the double arrow ensures that the head is to the left of CRel, i.e., head-initial.

	Terminal	Syntactic feature	Note
1	Noun	N	Noun phrase
2	Det	=N, D, -case	Determiner Phrase
3	<b>Che</b>	<b>DRel, -case, -wh</b>	<b>Relative pronoun <i>che</i> has -wh feature</b>
4	Vt	=DRel, +case, VOR	OR verb selecting a relativizable DP as object
5		=>VOR, =D, vRel	OR verb selecting a DP as subject
6		=vRel, +case, TRel	tense in the RC
7		=TRel, +wh, CRel	<i>+wh</i> feature hoisting noun phrase
8		<b>D &lt;&lt; CRel</b>	<b>RC left adjoins the head DP</b>

Table 4.3: Sample of MG lexical entries under the adjunction analysis for Italian relative clauses

The alternative “promotion analysis” supposes that the head noun itself originates from within the embedded clause to the left periphery area of the CP (Vergnaud, 1985; Kayne, 1994; Bianchi, 2002). Illustrated by Figure 4.5, the head noun “NP(0)” is part of the DRelP relativized from the embedded object position to the Spec of CP.

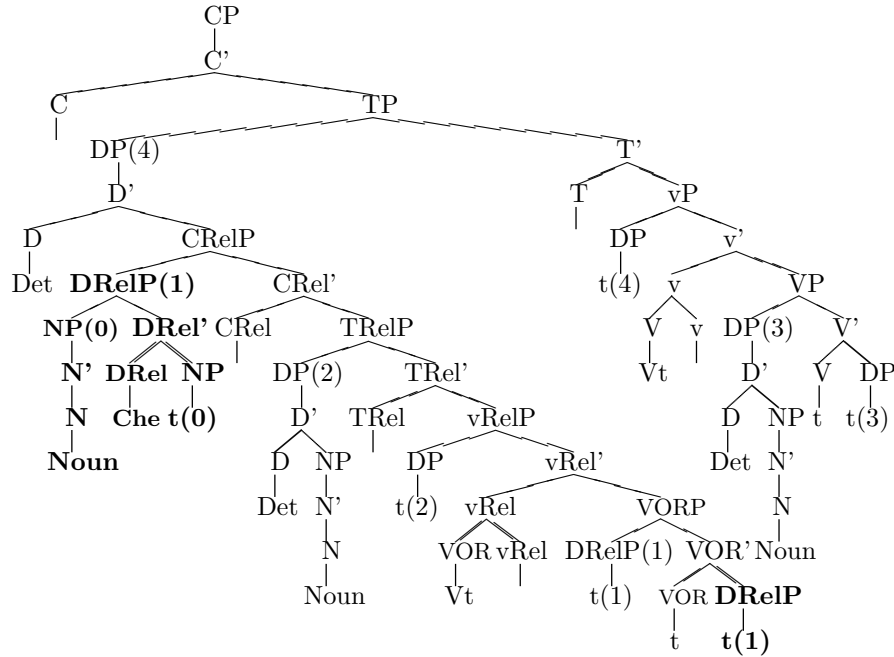


Figure 4.5: Italian object relative clause derived by a “promotion” grammar

The  $-f$  feature of *che* in Line 4 of Table 4.4 promotes the head noun (Line 2) to the Spec of CP along with the relative pronoun itself. Compare it with the counterpart in Table 4.3. With this promotion feature, the adjunction between RC and the head it modifies is no longer necessary.

In addition to *che*-relatives, the Italian MG fragments prepared for ER modeling also cover a number of relevant structures, including RCs with *cui* as a relative pronoun, subject *pro*-drop, and complement clauses. The grammars

	Terminal	Syntactic feature	Note
1	Noun	N	Noun phrase
2	Noun	N, $-f$	Noun phrase with promotion feature $-f$
3	Det	=N, $-case$	Determiner Phrase
4	<b>Che</b>	=N, $+f$ , <b>DRel</b> , <b><math>-case</math></b> , $-wh$	$+f$ <b>promotes NP to form a DP with <math>-wh</math> feature</b>
5	Vt	=DRel, $+case$ , VOR	OR verb selecting a relativizable DP as object
6		=>VOR, =D, vRel	OR verb selecting a DP as subject
7		=vRel, $+case$ , TRel	tense in the RC
8		=TRel, $+wh$ , CRel	$+wh$ feature hoisting noun phrase

Table 4.4: Sample of MG lexical entries under the promotion analysis for Italian relative clauses

also cover three different verb categories, namely intransitive, transitive and ditransitive verbs. Figure 4.6 shows an example of *cui*-relative with ditransitive verb (Vdi). Note that all Italian MGs adopt the general word order of Subject-Verb-Object. This renders the grammars more comparable across English and Chinese. In the future, the MG fragments can naturally be extended to cover RC variants in Italian such as ORs with post-verbal subjects. The major obstacle is that expressions have just one linear precedence interpretation, as derived by merge and move in Stabler (1997). However, in the style of Generalized phrase structure grammar (GPSG) one might separate out linear precedence for purposes of modeling freer word order (e.g. Kathol (2000)). An practical alternative would be to use movement to derive all possible word orders. This alternative presupposes that there are important syntactic or interpretative differences between the alternative word orders, a point that remains to be shown.

In order to test the impact of animacy information on models of Italian RC processing, this chapter also considers an MG fragment in which this information is represented in an enriched system of category symbols. This grammar

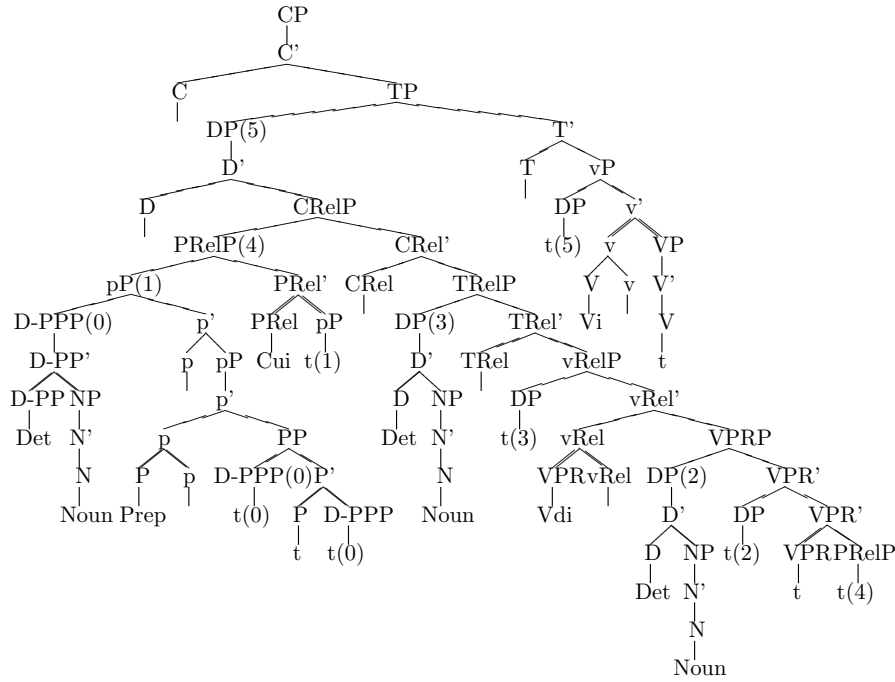


Figure 4.6: Italian *cui*-relative clause derived by a “promotion” grammar

adopts the promotional analysis of  $RC^3$ . A handful of MG lexical items exemplified in Table 4.5 are encoded with specific animacy. It can be used to generate an OR with animate subject and inanimate head noun, as in “Det InaniN Che Det AniN Vt ...”.

### 4.3.2 Counting structures in Italian treebanks

To reflect distributional factors, the MG fragment needs to be weighted. The MCFG translated from the MG defines a set of parameters that distinguishes one construction from others, like those listed in Table 4.6. The numbers asso-

<sup>3</sup>The ER modeling results of adjunction and promotion analysis are found to be similar, as to be discussed in Section 4.3.3.

	Terminal	Syntactic feature	Note
1	AniN	aniN	animate noun
2	InaniN	inaniN, <i>-f</i>	relativizable inanimate noun
3	Det	=aniN, aniD, -case	determiner phrase for animate N
4	Det	=inaniN, inaniD, -case	determiner phrase for inanimate N
5	<b>Che</b>	<b>=inaniN, +f, inaniDRel, -case, -inaniwh</b>	relative pronoun <i>che</i> for an inanimate head with <i>-inaniwh</i> feature
6	Vt	inaniDRel, +case, VOR	OR verb selecting its inanimate object
7		=>VOR, =aniD, vRel	OR verb selecting its animate subject
8		=vRel, +case, TRel	tense in the RC
9		=TRel, +inaniwh, inaniCRel	+ <i>inaniwh</i> hoisting inanimate DP

Table 4.5: Animacy-encoded MG fragment derives an Italian OR with animate subject and inanimate object

ciated with each construction are attestation counts from the Turin University Treebanks (TUT) (Bosco, Lombardo, Vassallo, & Lesmo, 2000). TUT is comprised of 2,860 Italian sentences and has a version in Penn Treebank format. A handful of parameters can therefore be estimated from the corpus counts: the rate of alternative RC types including *cui* relatives; the frequency of different types of arguments; the choice among transitive, intransitive and ditransitive verbs in both main and subordinate clauses, and the rate of subject omission. The selection of this parameterization, rather than some other one, is a consequence of the grammar fragment itself.

Table 4.6 suggests that Italian SRs are much more frequent than ORs, a result that is compatible with other languages. In addition, Italian subjects are very likely to be omitted. *pro* is even more well-attested than noun phrases in subject position.

Weighting the animacy-encoded grammar requires an in-depth corpus study of Italian noun phrases. . Since TUT has not been annotated with animacy, we

<i>cui</i> relative	134
<i>che</i> subject relative	430
<i>che</i> object relative	43
noun with determiner	2503
relative clause with determiner	70
complement clause with determiner	1
intransitive verb	1396
transitive verb	2268
ditransitive verb	991
subject <i>pro</i> with transitive verb	1314
subject <i>pro</i> with intransitive verb	771
transitive RC verb	277
intransitive RC verb	215
<i>pro</i> as matrix subject	1739
noun as matrix subject	1188

Table 4.6: Attestation counts from the Turin University Treebank are used to weight the grammar for Italian RCs

estimate the crucial distribution of animacy-related parameters using the Siena University Treebank (Chesi, Lebani, & Pallottino, 2011). SUT is a bilingual treebank with both Italian (ITA) and Italian Sign Language (LIS). In this work, only the ITA portion of the SUT is used. The XML-structured SUT is first transformed to the Penn Treebank phrase structure format so that it is searchable using Tregex (Levy & Andrew, 2006).

Tables 4.7 and 4.8 present the noun phrase animacy distribution in matrix clauses and relative clauses, respectively. Table 4.9 investigates the animacy of three noun phrases in a *cui* relative clause, namely the head noun, the embedded subject and the embedded object. Note that the total number of SRs and ORs here are different from those reported in Belletti and Chesi (2011) because they are retrieved using a stricter query, i.e. the head noun has to be co-indexed with the trace within the RC.

Verb	Subject	Object	Count
transitive	+anim	+anim	3
	+anim	-anim	42
	-anim	+anim	4
	-anim	-anim	73
	pro	+anim	1
	pro	-anim	16
intransitive	+anim	N/A	52
	-anim	N/A	179
	pro	N/A	16

Table 4.7: Corpus counts of different matrix clause types in Italian

	Head-NP	Embedded-NP	Count	Total
Subject Relative	+anim	N/A	14	80
	-anim	N/A	18	
	+anim	-anim	13	
	+anim	+anim	0	
	-anim	+anim	4	
	-anim	-anim	31	
Object Relative	+anim	pro	3	10
	-anim	pro	2	
	+anim	-anim	0	
	+anim	+anim	0	
	-anim	+anim	4	
	-anim	-anim	1	

Table 4.8: Corpus counts of Italian RCs with various animacy patterns

### 4.3.3 Italian prediction I: the Subject Advantage

Two explicit grammar fragments which cover the Italian SRs, ORs (Example 4.1), and relevant structures are prepared. One grammar adopts the “adjunc-

	Head-NP	Embedded-Subj	Embedded-Obj	Count	Total
<i>cui</i> relative	+anim	pro	-anim	2	4
	+anim	-anim	-anim	1	
	+anim	+anim	-anim	1	
	-anim	<i>pro</i>	N/A	3	9
	-anim	+anim	N/A	3	
	-anim	-anim	N/A	1	
	-anim	-anim	-anim	2	

Table 4.9: Corpus counts of Italian *cui* relatives with various animacy patterns

tion analysis” while the other implements the “promotion analysis”, both of which are discussed in Section 4.3.1. Those two grammars are then weighted by construction frequencies retrieved from the Turin University Treebank. The Entropy Reduction hypothesis, in conjunction with the grammars in Section 4.3.1 and the frequency facts presented in Section 4.3.2, correctly derives the widely-observed subject relative advantage holds in Italian and the predictions based on two grammars are similar.

Using self-paced reading, Di Domenico and Di Matteo (2009) observed the subject advantage in Italian. This effect was localized to the three-word RC region. A direct comparison between those two types RCs can be found in Table 4.1. The subject advantage starts from the second set of words in the RC, namely the determiner in SRs and the noun in ORs. The subject advantage effect was more significant at the third set of words, the processing of the OR embedded verb was much slower.

Based on two grammars with different syntactic analyses, Figures 4.7 and 4.8 illustrate the predicted disambiguation costs at each word, quantified by

reductions in Entropy. The comparison between SRs and ORs suggests early processing difficulty in ORs at the determiner of the embedded subject, as compared to the embedded verb in SRs. This is roughly compatible with the English modeling results using Surprisal (Hale, 2001; Levy, 2008). The predicted reading difficulties between SRs and ORs is bigger at the last set of words in the RC region, consistent with the result of Di Domenico and Di Matteo (2009).

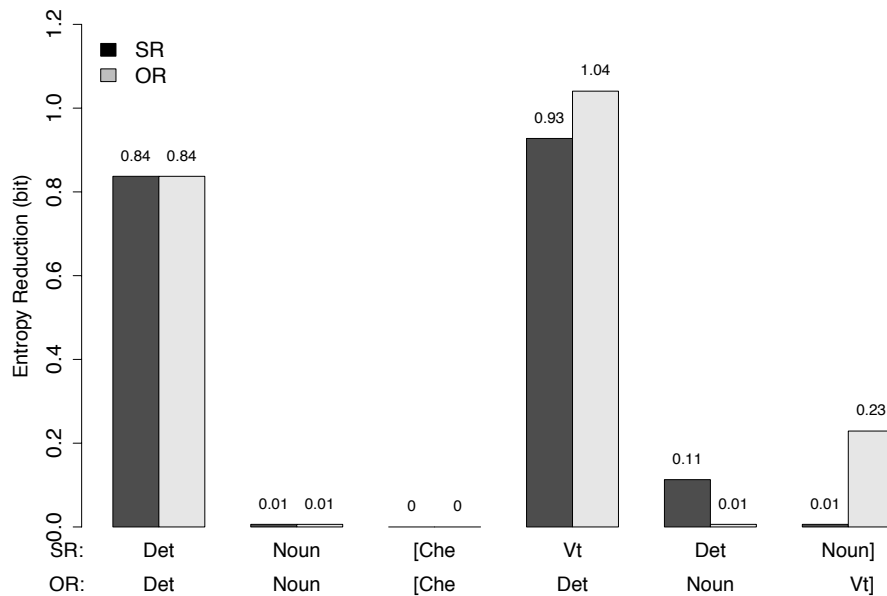


Figure 4.7: Predicted word-by-word reading difficulties of Italian RCs using an “adjunction” grammar

The CCPC makes it possible to understand this contrast in terms of specific syntactic disambiguation decisions. For example, Figure 4.9 examines how the uncertainty level is reduced at the first set of words, i.e., 0.93 bits at the SR verb and 1.06 bits at the determiner of OR subject. The Italian SR prefix “Det Noun Che Vt” (the bottom left box) is slightly more ambiguous than the OR prefix because the next word after the embedded SR verb can either be a determiner (of the embedded object) or a verb (highlighted in red). The later case falls into the category of OR, although with omitted RC subjects. By contrast, the

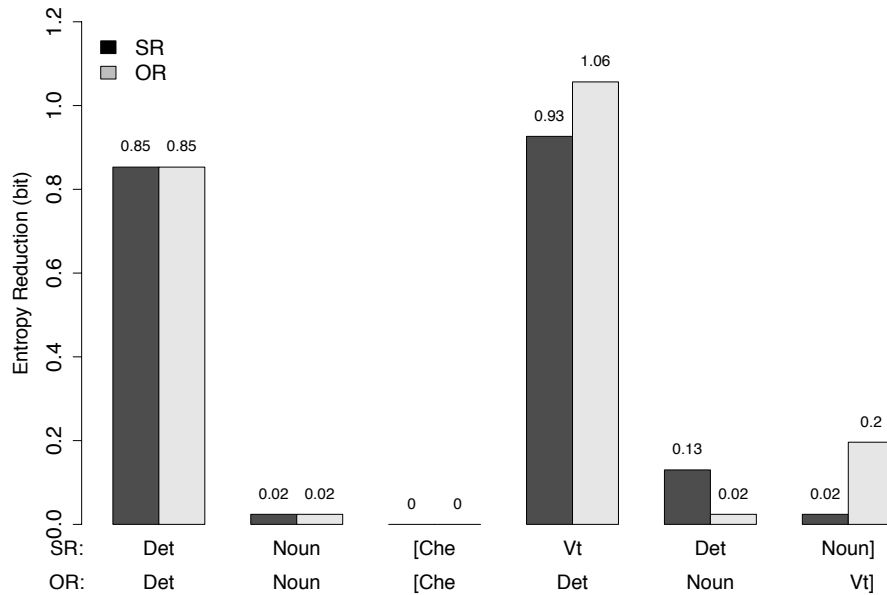


Figure 4.8: Predicted word-by-word reading difficulties of Italian RCs using a “promotion” grammar

alternatives given the OR prefix “Det Noun Che Det” in the bottom right box are less ambiguous because the following word after “Det” is most likely to be a noun. Therefore, less disambiguation work has been done in the SR than in the OR.

The predicted subject advantage in Italian RCs is more pronounced at the last word in the RC region. This modeling result is compatible with the observation in the self-paced reading experiment in which ORs were read much slower at the exact same word. Figure 4.10 explores the disambiguation processes in SRs and ORs. In the SR prefix “Det N Che Vt Det”, shown in the top left boxes, the next word “N” is predicted undoubtedly, which means the sentence has already been disambiguated. On the other hand, given the OR prefix “Det N Che Det N”, it is possible to continue the sentence with a relative pro-

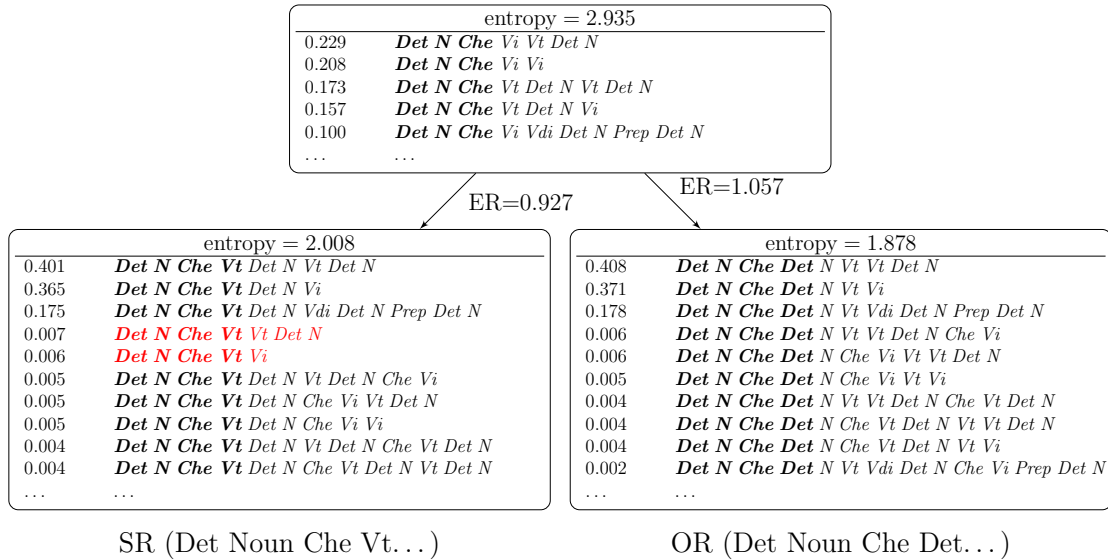


Figure 4.9: Transitions at the first set of words within RC region in Italian

noun *che*. This second *che* in the sentence leads to a double-embedded structure (highlighted in red). The uncertainty degree reduces when the parser has made its choice between “Vt” or “*che*”, in this case a “Vt” (the bottom right box).

#### 4.3.4 Italian prediction II: the head animacy effect

The generality of our approach allows a theorist to explore the role of “formalist” as well as “functionalist” features, such as animacy, in sentence processing. As we have discussed in Section 1.3, the animacy feature is interesting because of its differential influence on RC comprehension across languages. For example, Traxler et al. (2002) report that the difficulty of processing English ORs is reduced when the head noun of the OR is inanimate (OR-+ or OR--).

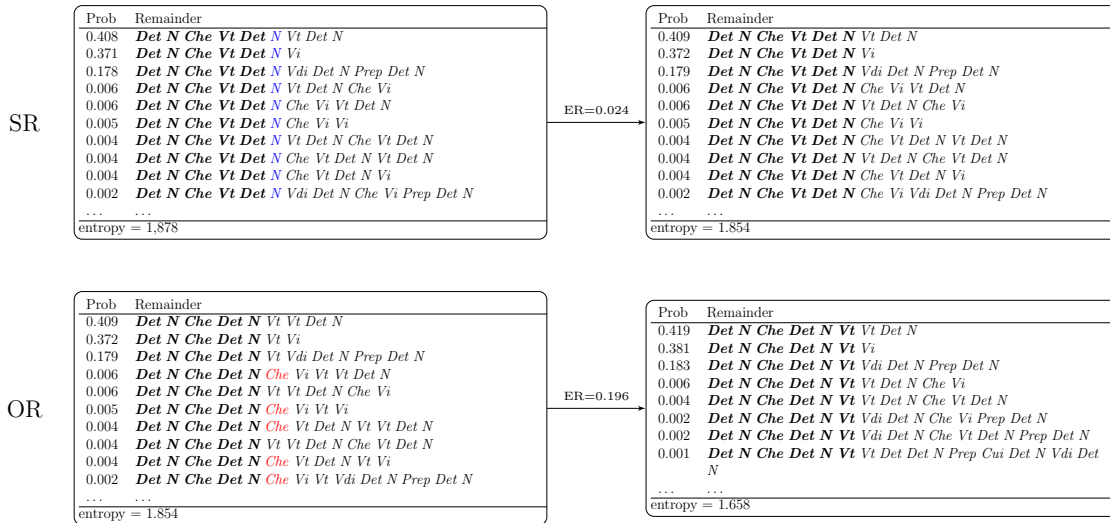


Figure 4.10: Transitions at the third set of words within RC region in Italian

Section 4.2 has reviewed two representative studies on this topic using Italian RCs. Similar to Traxler et al.'s result, an animacy effect was also reported by Arosio et al. (2011), they argued that object relatives were more taxing with an animate head noun. However, Belletti and Chesi (2011) found that participants of a Italian sentence production task are immune to the change of head noun animacy in ORs. Switching the head noun animacy from animate to inanimate will not result in more production of ORs.

Using CCPC and the modeling procedure discussed above, ER metric is able to correctly derive both observed results from a combination of grammatical assumptions and corpus distributions. First, Table 4.10 compares predictions from Entropy Reduction to the RC production data from Belletti and Chesi (2011). ER predictions are consistent with the observed result such that Italian ORs with animate heads are not processed slower than those with inanimate heads. In-

deed, among the four conditions, there were more ORs produced with animate subjects and animate heads.

Head NP	+anim	+anim	-anim	-anim
Embedded NP	+anim	-anim	+anim	-anim
number of OR production	23	7	12	11
ER prediction (bit)	1.67	1.67	1.75	1.75

Table 4.10: ER predictions of Italian ORs with different animacy patterns compared against the production study results of Belletti and Chesi (2011)

In addition, the modeling predictions similarly offer an explanation of the self-paced listening results of Arosio et al. (2011). As shown in Table 4.11, the subject advantage, quantified by the difference in second or in ER, is larger when comparing the two leftmost columns in which both nouns in the RC are animate. If the animacy pattern of an RC is canonical, i.e. animate subject with inanimate object, the subject advantage is less pronounced.

	SR	OR	Diff.	SR	OR	Diff.
Head NP	+anim	+anim		+anim	-anim	
Embedded NP	+anim	+anim		-anim	+anim	
Listening time (sec)	1.12	1.19	<b>0.07</b>	1.16	1.17	<b>0.01</b>
ER prediction (bit)	1.35	1.67	<b>0.32</b>	1.51	1.75	<b>0.24</b>

Table 4.11: ER predictions of Italian ORs with different animacy patterns compared against the self-paced listening results of Arosio et al. (2011)

Although to some extent the experimental data on the animacy effect in Italian RCs seem to be contradictory, ER modeling derives the most consistently-observed aspects of the processing difficulty profile. Continuing empirical work is important because the field has not reached a consensus. For example, Belletti and Chesi (2011) did not consider finer argument-predicate frequencies in their experiment design. Recall that in their Experiment 1, the verb of elicitation sen-

tences were altered so that listeners were expected to produce ORs with different animacy patterns. The number of each condition in the experiment remain constant. This kind of design is convenient for experimental analyses but may not correctly reveal how frequency distribution of Italian relatives and the noun phrase animacy influence readers' production.

In addition to modeling published data, CCPC can be used to make reading time predictions on conditions to be tested in the future. For example, Table 4.12 compares reading time differences between SRs and ORs (Traxler et al., 2002). With an inanimate head, ORs become less difficult. Therefore, the subject advantage effect is reduced.

	Head-NP	Embedded-NP	Total Time (ms)	Diff.
SR	+anim	-anim	881	<b>294</b>
OR	+anim	-anim	1175	
SR	-anim	+anim	847	<b>39</b>
OR	-anim	+anim	886	

Table 4.12: The subject advantage shrinks when OR has an inanimate head noun (Traxler et al., 2002)

Following this experimental design, Table 4.13 compares ER predictions at the RC region for those four RC conditions, suggesting that the processing difficulty of ORs will not be reduced with an inanimate head noun phrase. Although this prediction agrees with the Belletti and Chesi's claim such that Italian speakers are immune to the animacy in producing ORs, further investigations are still needed, especially in sentence comprehension.

To summarize, Chapter 4 examines the processing of Italian RCs. ER predictions are made to confirm both the subject advantage and the seemingly contradictory results on the effect of animacy. Chapter 5 presents the modeling of

	Head-NP	Embedded-NP	ER (bits)	Diff.
SR	+anim	-anim	1.51	<b>0.16</b>
OR	+anim	-anim	1.67	
SR	-anim	+anim	1.43	<b>0.32</b>
OR	-anim	+anim	1.75	

Table 4.13: Predicted processing difficulty for the RC region

Chinese RCs.

## CHAPTER 5

### CHINESE RELATIVE CLAUSES

This chapter provides the third modeling example on relative clause processing. The target language, Chinese, is structurally different from two languages discussed in previous chapters, namely English and Italian, because relative clauses in Chinese are prenominal. In other words, the head noun phrase follows the RC. Languages with these kinds of head-final RCs have drawn considerable attention in recent discussions, mostly because they present a challenge to previous theories proposed for relative clause processing in languages with postnominal RCs, in particular, English. As a start, this chapter examines the predictions of Chinese RC comprehension difficulty made by different classes of sentence processing theory. In particular, it will focus on two leading ideas which have garnered considerable attention in recent years: working-memory based accounts such as the Dependency Locality Theory (Gibson, 1998, 2000) and based on probabilistic grammars, including Entropy Reduction and Surprisal. Despite conflicting reports in the literature, the most recent evidence decisively disconfirms a working-memory account of Chinese RC processing. This chapter will instead support the idea that processing decisions made during Chinese RC comprehension are derived by readers' expectations, both structurally and non-structurally. Evidence for the expectation-based theories came from the Subject Advantage observed in "disambiguated" RCs as well as the facilitation of noun phrase animacy in Chinese RC processing. Entropy Reduction, as an expectation-based metric, is able to predict processing difficulties found in Chinese RCs, including the effect of head noun animacy. The model-

ing results made by CCPC are consistent with both recent experimental reports and other modeling efforts.

In this chapter, Section 5.1 serves as a basic introduction to Chinese RCs, with an emphasis on temporary ambiguities involved in their comprehension. Section 5.2 provides a comprehensive literature review of Chinese RC processing issues. This Section addresses the failure to find a Subject Advantage that was characteristic of early work on Chinese RC processing. However, the puzzling picture becomes much clearer once the impacts of temporary ambiguities are taken out (Jäger et al., Submitted). This section also pays special attention to recent studies on the role of noun phrase animacy in Chinese RC processing (Wu, 2009; Wu et al., 2012), an effect that is comparable to English (Traxler et al., 2002). Section 5.3 applies Entropy Reduction to derive a processing difficulty pattern, consistent with the Subject-Object asymmetry that has been observed experimentally with “disambiguated” stimuli. It also successfully mirrors the head noun animacy effect reported by Wu et al. (2012). Comparing results based on grammars with or without animacy information, the model shows that the animacy cue strongly impacts the processing of Chinese sentences.

## 5.1 Ambiguities in Chinese RCs

The most distinctive feature of Chinese relatives is that they are prenominal, as compared to English and Italian discussed in Chapters 3 and 4 respectively. To see the prenominal character of these RCs, consider the examples in (5.1) from F. Hsiao and Gibson (2003). In both examples, the noun phrase *guanyuan* “official” is relativized to the right end of the relative clause instead of its left-end.

Those two example sentences also highlight the empty category (or gap) which the head noun phrase left in the RC after relativization. Head-final RCs distinguish themselves from their head-initial counterparts because the distance (both linear and structural) between the empty category and head noun is longer in subject relatives than in object relatives. As the next section shows in detail, the reversed filler-gap distance between subject relatives and object relatives poses a challenge to theories of RC processing based on the working-memory burden. On the contrary, grammar-based proposals such as Surprisal and Entropy Reduction are not necessarily affected by this word order change: readers still have a stronger expectation for SRs than for ORs, as attested by the higher frequency of SRs in Chinese corpora.

(5.1) a. Chinese Subject-modifying Subject relative

邀请	富豪	的	官员	心怀不轨
[ e <sub>i</sub>	Yaoqing	fuhao	de ]	guanyuan <sub>i</sub> xinhuaibugui
invite	tycoon	DE	official	have.bad.intentions

*'The official who invited the tycoon has bad intentions.'*

b. Chinese Subject-modifying Object relative

富豪	邀请	的	官员	心怀不轨
[ Fuhao	yaoqing	e <sub>i</sub>	de ]	guanyuan <sub>i</sub> xinhuaibugui
tycoon	invite	DE	official	have.bad.intentions

*'The official who the tycoon invited has bad intentions.'*

Gibson and Wu (2013) argue that temporal ambiguities could confound the comprehension of Chinese RCs. Because RCs precede their head noun phrases in Chinese, readers have few clues to identify the left-boundary of an RC. In fact, there exist a number of sentence-medial temporary ambiguities that could mislead the comprehension of RCs. For example, when reading an SR as in (5.1a),

the sentence-initial words *Yaoqing fuhao* “invite tycoon” can be parsed as a main clause with a dropped subject (or subject *pro*), meaning “someone invited the tycoon”. This ambiguity is strong because subject *pro*-drop is highly acceptable in Chinese. *Pro*-drop is particularly felicitous when a preceding context is presented. The beginning of a Chinese SR can also be a complement clause with the main clause predicate, such as *Yaoqing fuhao bu changjian* “It is not common to invite a tycoon”. At the next word, the relativizer *de*, both ambiguities still exist because *de* can also be part of a possessive structure, as in *fuhao de mishu* “tycoon’s secretary”. Therefore, in the most typical Chinese SR structure, such as (5.1a), alternative parses competing with the RC are sometimes only resolved semantically when reaching to the main predicate. Example (5.2) highlights this temporary ambiguity between SR and complement clause structures. The sentence only gets disambiguated at the main predicate, i.e., *xinhuaibugui* “have bad intentions” in (5.2a) and *buchangjian* “uncommon” in (5.2b).

(5.2) a. Subject relative clause

	邀请	富豪	的	秘书	心怀不轨
e <sub>i</sub>	Yaoqing	fuhao	de	mishu <sub>i</sub>	xinhuaibugui
	invite	tycoon	DE	secretary	have.bad.intentions

*‘The secretary who invited the tycoon has bad intentions.’*

b. Complement clause with subject *pro*-drop

	邀请	富豪	的	秘书	不常见
<i>pro</i>	Yaoqing	fuhao	de	mishu	buchangjian
	invite	tycoon	DE	secretary	uncommon

*‘Inviting the tycoon’s secretary is uncommon.’*

Chinese object-extracted RCs, as in (5.1b), start with the RC subject *fuhao* “tycoon” followed by the RC verb *yaoqing* “invite”. However, since the canoni-

cal word order in Chinese is Subject-Verb-Object, this Noun+Verb sequence can also be interpreted as the beginning of a canonical main clause, which would be a complete sentence if a direct object were to follow the transitive verb, e.g., *Fuhao yaoqing pengyou* “The tycoon has invited friends”. This main clause analysis becomes impossible when reaching the next word, namely the relativizer *de*. Similarly, the complement clause reading also exists, for example *Fuhao yaoqing bu changjian*. “It is not common that the tycoon invited someone.”, although corpus studies have show that object-*pro* drop is rare in Chinese. Note that there is no ambiguity caused by possessive structures in Chinese ORs since *de* follows the RC verb instead a noun phrase. The possessor noun phrase in Chinese cannot be omitted.

Relative clauses sometimes modify the matrix subject. In that situation, ambiguities also exist. In object-modifying SRs (5.3a), the verb and object sequence within the RC could be analyzed as the subject of a sentential complement, similar to the subject-modifying counterpart. However, whether this parse is possible depends on the choice of the main clause verb. If the RC verb+RC-object combination follows certain verbs that can take sentential complements, such as *tingshuo* “hear” or *zhidao* “know”, the complement clause reading would be allowed.

Another classic example which has been studied previously is the main clause illusion in object-modifying ORs. The RC subject (*fuhao* in 5.3b) can be interpreted as the direct object of the preceding main clause verb (*daibule* “arrested”). Therefore, the main clause subject (*jingcha* “police”), the main clause verb and the RC subject could be read as a canonical transitive main clause (*Jingcha daibule fuhao*; “The police arrested the tycoon”). This reading becomes

impossible upon reaching the RC verb (*yaoqing*). It has been argued that this local ambiguity leads to the garden-path effect of structural reanalysis at the RC verb (C.-J. C. Lin & Bever, 2011; Z. Chen, Grove, & Hale, 2012; Gibson & Wu, 2013)<sup>1</sup>.

(5.3) a. Chinese object-modifying subject relative

警察 逮捕了 邀请 富豪 的 官员  
 Jingcha daibule [ e<sub>i</sub> yaoqing fuhao de ] guanyuan<sub>i</sub>  
 Police arrested invite tycoon REL official  
 'The police arrested the official who had invited the tycoon.'

b. Chinese object-modifying object relative

警察 逮捕了 富豪 邀请 的 官员  
 Jingcha daibule [ fuhao yaoqing e<sub>i</sub> de ] guanyuan<sub>i</sub>  
 Police arrested tycoon invite REL official  
 'The police arrested the official who the tycoon had invited.'

Lastly, Chinese RCs can be headless. The ambiguity between an RC with overt or covert head noun phrase exists in all types of RCs. As shown in Example (5.4), both SRs and ORs can modify an covert matrix subject. In addition, a corpus study suggests that Chinese ORs are more likely to be headless than SRs (Z. Chen, Jäger, & Hale, 2012).

(5.4) a. Chinese headless SR

邀请 富豪 的 心怀不轨  
 [ e<sub>i</sub> Yaoqing fuhao de ] e<sub>i</sub> xinhuaibugui  
 invite tycoon DE have.bad.intentions  
 'Someone who invited the tycoon has bad intentions.'

b. Chinese headless OR

<sup>1</sup>See Mazuka and Itoh (1995) for a discussion of garden-path sentences in Japanese.

富豪 邀请 的 心怀不轨  
 [ Fuhao yaoqing e<sub>i</sub> de ] e<sub>i</sub> xinhuaibugui  
 tycoon invite DE have.bad.intentions  
*'Someone who the tycoon invited has bad intentions.'*

To some extent, the head-final feature of Chinese relativized structures provides an environment for those ambiguities to co-exist with the RC parse in incremental sentence comprehension, especially when there is limited help from the context. Indeed, previous conflicting results on this topic are likely due to local ambiguities in the experimental materials, as discussed by a number of authors (C.-J. C. Lin & Bever, 2006, 2011; Qiao, Shen, & Forster, 2012; Vasishth, Chen, Li, & Guo, 2013; Y. Hsiao, Li, & MacDonald, 2014). Reducing the impact of local ambiguities becomes necessary in experiments of Chinese RC comprehension.

## 5.2 Experimental studies of Chinese RC processing

Chinese is not the only language with prenominal relative clause structures. Other languages in this category, such as Japanese and Korean, attest the Subject Advantage. The evidence on this point is strong, and it comes from a wide variety of methodologies, such as self-paced reading, eye-tracking, and ERP (Japanese: Miyamoto & Nakamura, 2003; Ueno & Garnsey, 2008; Korean: Kwon, Polinsky, & Kluender, 2006; Kwon, Lee, Gordon, Kluender, & Polinsky, 2010; Kwon, Kluender, Kutas, & Polinsky, 2013). Among different theories which have been proposed to account for the processing ease of SRs in Japanese and Korean, the working-memory based theories fail to furnish a sound explanation. By contrast, models based on readers structural expectations are able

to account for the subject advantage in both languages (Yun, Whitman, & Hale, 2010; Yun et al., 2015).

While the processing literature on Japanese and Korean is univocal about the Subject Advantage in those languages, the Chinese RC processing literature is inconsistent<sup>2</sup>. F. Hsiao and Gibson (2003) were the first to test the processing of Chinese RCs and argued that Chinese ORs were actually easier to process as compared to SRs. This surprising result was different from findings in all other languages. The evidence for their claim came from a 2×2 self-paced reading experiment with different RC types (SRs vs ORs) and number of embeddings (single vs double).

In singly-embedded RCs, ORs were read more slowly than SRs at the relative clause region, namely the RC verb and RC object in SRs and the RC subject and the RC verb in ORs. At the relativizer *de* and subsequent words including the head noun, no OR advantage was found. In doubly-embedded conditions, readers spent less time in ORs than SRs at the region containing two relative clauses up to the relativizer *de* of the outer RC. F. Hsiao and Gibson's result is particularly interesting not only because it was the first time that an Object Advantage was reported, but also because the authors interpreted this finding as evidence for a memory-based account, i.e., the Dependency Locality Theory (Gibson, 2000), discussed above in Section 3.1. The DLT includes two processing difficulty metrics, namely the storage and integration costs. The storage cost assumes that the processing difficulty increases if more syntactic heads are predicted at a given word. The integration cost, on the other hand, is defined as a

---

<sup>2</sup>Carreiras, Duñabeitia, Vergara, de la Cruz-Pavía, and Laka (2010) have also reported that Basque also represents an exception to the universal subject-relative preference. However, the object-relative advantage found there could relate to the unique ergativity in Basque. See also for recent evidence supporting the Subject Advantage in ergative Mayan languages (Clemens et al., in press).

function of the number of new discourse referents (such as noun phrases) that intervene between the phrase that is currently being processed and the constituent(s) to which the dependency links. In the case of RCs, the integration cost metric reflects the number of new discourse referents between the head noun phrases and the empty category (or the RC verb) within the RC. In languages like English and Italian, the integration cost at the embedded verb is higher in ORs than in SRs. For languages with prenominal relative clauses like Chinese, Japanese, and Korean, the DLT predicts longer reading times in SRs, compared to ORs. This inverse pattern for prenominal RCs is predicted to occur at the head noun, where the filler-gap dependency is finally resolved.

The original F. Hsiao and Gibson (2003) study runs counter to subsequent experimental findings involving Chinese RCs, which found a subject preference. Among those results that support an universal subject preference, Lin & Bever present a series of self-paced reading studies which all provide counterevidence against Hsiao and Gibson (C.-J. C. Lin & Bever, 2006, 2007; C.-J. C. Lin, 2008; C.-J. C. Lin & Bever, 2011). For example C.-J. C. Lin and Bever (2006) considered the same types of RCs as Hsiao and Gibson, along with two other types in which the head noun occupies a matrix object position. These two types appear on Page 87 as example (5.3). Figure 5.1 plots the reading time per word across four types of RCs and shows that SRs are read faster than ORs in Chinese in both subject-modifying and object-modifying conditions, although the subject advantage was only statistically significant when the object of the main clause was modified by an RC. Although this result plainly supports the idea of a Chinese Subject Advantage, Gibson and Wu (2013) suggested that Lin et al's result actually reflected a methodological flaw. Lin and Bever's stimuli, while otherwise well-controlled, nevertheless allow for the main clause illusion discussed

above in Section 5.1.

However, as pointed out, it is possible that the amplified subject advantage in object-modifying conditions relates to other local ambiguities, such as the main clause illusion in object-modifying ORs.

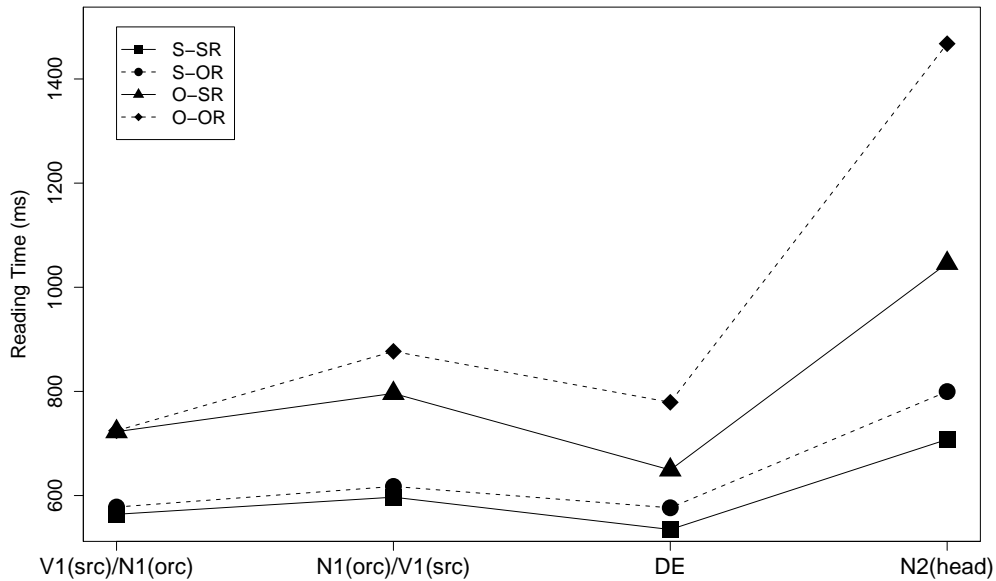


Figure 5.1: Reading times of each region in Chinese RCs (C.-J. C. Lin & Bever, 2006)

Other experimental studies have reported either an SR advantage (Wu, 2009; Wu et al., 2012; Vasishth et al., 2013; Xu, 2014; Jäger et al., Submitted) or an OR advantage (Su, Lee, & Chung, 2007; B. Chen, Ning, Bi, & Dunlap, 2008; Y. B. Lin & Garnsey, 2011; Gibson & Wu, 2013). Nevertheless, after the initial attempt by F. Hsiao and Gibson (2003), the weight of the evidence seems to suggest that Chinese is not as exceptional as first suggested. However, a final determination awaits further investigation, especially when temporary ambiguities could confound the processing effect of RCs, no matter whether it is a subject advantage or an object one.

One natural way to solve the problem of unwanted temporary ambiguity is

to introduce a context preceding the critical sentence that promotes the expectation of an upcoming RC. Gibson and Wu (2013) conducted such an experiment on subject-modifying singly-embedded RCs<sup>3</sup>. In their design, the target sentences were similar to the experimental stimuli of F. Hsiao and Gibson (2003). In addition, those target sentences were preceded by a context which introduces the action described in the RC. The preceding context also asked a question eliciting either the agent or the patient of the action. The answer to this question needed to be the target sentence itself, i.e., either an SR or an OR. Gibson and Wu found an OR advantage at the relativizer *de* and this effect reached statistical significance at the next word, the head noun. This finding contradicts the null result in singly-embedded RC conditions of F. Hsiao and Gibson (2003). Gibson and Wu interpreted this finding as evidence for the integration cost metric. However, ORs were not read slower at the RC region preceding the relativizer *de* where a significant effect was reported by the initial Hsiao and Gibson study.

Vasishth et al. (2013) cast doubts on whether the working memory-based DLT could correctly predict the processing difficulties in Chinese RCs after they found that the Object Advantage at the head noun in Gibson and Wu's paper was driven entirely by 13 extreme data points in the SR condition that exceeded 2,300 ms. In addition, Vasishth and colleagues were unable to replicate both the F. Hsiao and Gibson (2003) and the Gibson and Wu (2013) studies. In particular, different from Hsiao and Gibson's original finding, they instead reported an Subject Advantage using the same experimental materials and design. Vasishth et al. also attempted a replication using the items and fillers of the Gibson and Wu paper. Although ORs were found to be read slower than SRs, this Object

---

<sup>3</sup>Gibson and Wu (2013) did not test object-modifying conditions perhaps because the main-clause illusion caused by the first three words (a Noun-Verb-Noun sequence) in object-modifying ORs cannot easily be eliminated even with a preceding context. See C.-J. C. Lin and Bever (2011) for a detailed discussion on this garden-path ambiguity.

Advantage effect started from and was only significant at the relativizer *de*, which cannot be explained by either DLT integration or storage cost metrics.

C.-J. C. Lin (2014) investigated the puzzling results of Gibson and Wu (2013) from a different angle. He argued that the Object Advantage observed there might reflect a syntactic priming from the preceding context rather than a lower DLT integration cost in ORs. Indeed, a stronger Object Advantage was found when the thematic order in the preceding context was similar to the one of an OR. In other word, the familiarity of the thematic pattern facilitated the comprehension of ORs.

These conflicting results may be due, as Gibson and Wu originally suggested, to local ambiguities. If this is true, then what we are really seeing is garden-pathing that happens not to be present in head-initial RCs like English. Therefore, a clean and precise experimental design needs to be in place to unveil the natural of Chinese RC processing. Unlike the biasing context provided in Gibson and Wu (2013), Jäger et al. (Submitted) used an experimental design where RC-initial substrings are disambiguated from matrix clauses. In such contexts, readers strongly anticipate a relative clause as opposed to other structures. The disambiguation is accomplished using extra words that help guide readers towards some RC interpretation while still leaving the specific gap site unspecified.

(5.5) a. “Disambiguated” Chinese SR

那个	昨天		邀请	富豪	的	官员	...
na-ge	zuotian	[ e <sub>i</sub>	yaoqing	fuhao	de ]	guanyuan <sub>i</sub>	...
that-CI	yesterday	[	invite	tycoon	DE ]	official	...

*‘The official who invited the tycoon yesterday ...’*

b. “Disambiguated” Chinese OR

那个 昨天 富豪 邀请 的 官员 ...  
na-ge zuotian [ fuhao yaoqing e<sub>i</sub> de ] guanyuan<sub>i</sub> ...  
that-Cl yesterday [ tycoon invite DE ] official ...  
‘The official who the tycoon invited yesterday ...’

In example (5.8), the demonstrative-classifier combination *na-ge* at the onset of the relative clause encourages readers to expect a noun phrase. However, the following word is a temporal phrase, such as *zuotian* “yesterday”, which has to be attached a verb phrase. This design therefore leads the reader to only foresee an upcoming RC-modified noun phrase by ruling out the *pro*-drop analysis. Jäger et al. (Submitted) tested these “disambiguated” RC stimuli in both self-paced reading and eye-tracking experiments. They reported that SRs are consistently read faster than ORs in the RC region (*Vt N* or *N Vt*, respectively) and at the head noun. A Subject Advantage was also found after the head noun, potentially a spillover effect from previous regions. This result can be explained by the expectation-based theories because readers would have higher expectation on a subject-extracted structure than an object-extracted one, as evidenced by the higher frequency of SRs in the corpus. Surprisal was used as a processing difficulty metric and derived the Subject Advantage observed in the experiments.

Studies have shown that theories based on structural expectations could well explain the processing difference between Chinese SRs and ORs. But the subject-object asymmetry was not the only evidence for these expectation-based accounts. Wu and colleagues (Wu, 2009; Wu, Kaiser, & Andersen, 2011; Wu et al., 2012) have conducted a series of experiments investigating how the frequency distribution of noun phrase animacy would impact the processing of Chinese RCs. The results indicate that animacy serves as an important cue for thematic role assignment and affects how ambiguities are resolved during RC

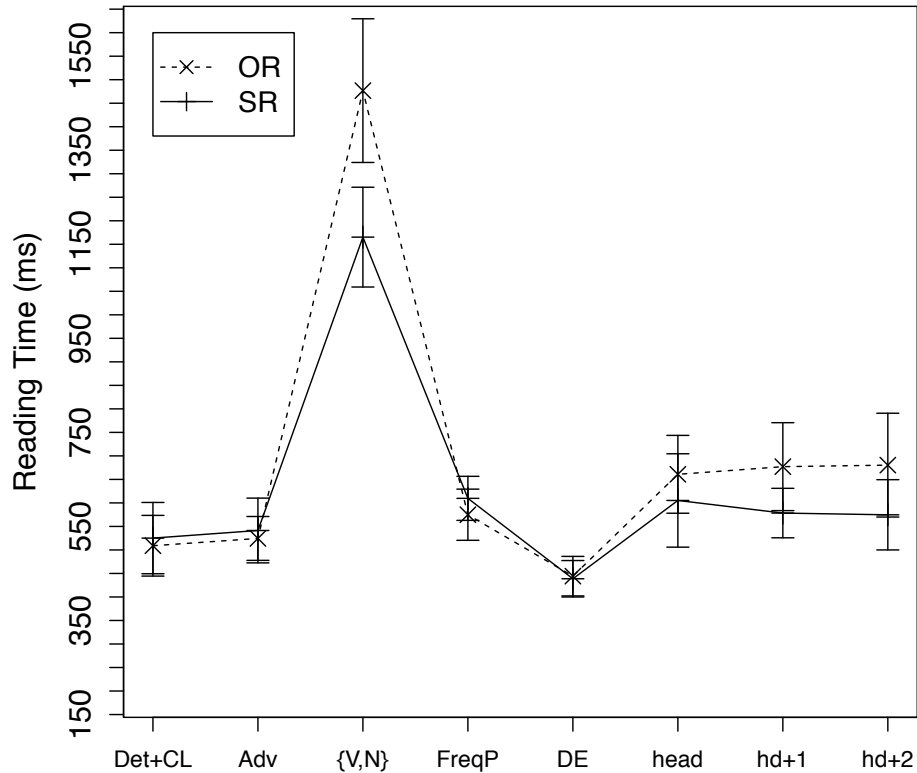


Figure 5.2: Reading times of each region in Subject-modifying Chinese RCs (Jäger et al., Submitted, Experiment 1)

processing, which is similar to those reported in other languages (Mak et al., 2002, 2006; Traxler et al., 2002, 2005), as discussed in Sections 1.3 and 3.2.2.

Based corpus analyses (Pu, 2007; Wu, 2009), observations in functional linguistics (Givón, 1983; Croft, 2003) and in cognitive psychology (Just & Carpenter, 1992; Clifton et al., 2003; Gennari & MacDonald, 2008, among others), Wu et al. (2012) suggested that a set of “Animacy Preference Constraints” could explain the patterns of RC structures with different animacy manipulations observed cross-linguistically, including Chinese.

(5.6) Animacy Preference Constraints (Wu et al., 2012, Page 1495)

- i. Subjects tend to be animate;
- ii. Head nouns in object-extracted RCs tend to be inanimate;
- iii. As a joint consequence of (i) and (ii), a contrastive animacy configuration tends to occur in object-extracted RCs with inanimate heads and in subject-modifying, subject-extracted RCs with animate heads.

With respect to the specific processing effects in Chinese RCs, Wu and colleagues found a facilitation effect of head noun animacy. SRs with animate heads and ORs with inanimate heads were easier to comprehend. They also reported the effect of animacy in embedded noun phrases (RC-objects in SRs and RC-subjects in ORs). This effect was marginal in SR conditions and was only seen at the third word after the head in OR conditions. Comparing SRs and ORs with two nouns that contrast animacy, the animacy effect of head nouns was also significant. Those experimental observations were consistent with Chinese corpus studies and further indicated a strong relationship between sentence processing patterns and comprehenders' expectations on structures and lexical semantics.

To conclude, Section 5.2 has reviewed previous works investigating how Chinese relativized structures are processed. A number of temporary ambiguities which coexist inside the RC prefix have led to conflicting experimental results. Removing the impact of syntactic alternatives with special experimental controls allows the Subject Advantage to reemerge in Chinese RCs. Together with the animacy effect at the head noun, the processing of Chinese RCs provides additional evidence for expectation-based sentence processing theories. The next section describes how expectation-based ER derive the Chinese RC

processing patterns.

### 5.3 Modeling Chinese RC processing

This section demonstrates the modeling of Chinese relative clause processing. As in the discussion of English and Italian in Chapters 3 and 4, the focus is on the Subject Advantage and noun animacy in RCs. Using the modeling procedure described in Section 2, Entropy Reduction derives a Subject Advantage in Chinese RCs, a results that is consistent with experimental evidence on “disambiguated” Chinese RC structures Jäger et al. (Submitted). With a subcategorized grammar encoding the frequency distribution of animacy in argument and verb relations, the modeling results mirror the self-paced reading times in (Wu et al., 2012) and are compatible with the predictions made by a connectionist model (Y. Hsiao & MacDonald, 2013).

#### 5.3.1 The Chinese Grammars

To calculate structural uncertainties at each word in a sentence, CCPC requires a weighted grammar. Different from modeling works using context-free analyses of Chinese RCs (Z. Chen, Grove, & Hale, 2012; Y. Hsiao & MacDonald, 2013), the Minimalist Grammar formalism allows us to analyze the filler-gap dependency as feature-driven movement. The grammar fragment used in this work supposes that, in relativization, an argument NP moves to become the head of an RC. Additionally, when the relativization is driven by an animate NP, processing is easier. Including syntactic categories that differentiate between + and

– animate NPs allows a modeler to combine syntactic and non-syntactic cues in the same formal grammar.”

Two Chinese MG fragments have been prepared for the modeling work described in this section. The first grammar covers the “disambiguated” experimental conditions in Jäger et al. (Submitted). It allows the determiner-classifier (“Det Cl”) combination attaching to a noun phrase. Clauses can also take a temporal phrase (“Time”) as an adjunct. The second grammar is capable of describing sixteen types of Chinese RCs with various animacy manipulation. Eight of those are subject-modifying RCs used by Wu et al. (2012) as experimental conditions while the rest are object-modifying ones modeled in Y. Hsiao and MacDonald (2013). We introduce the details of those two grammars in order.

Table 5.1 lists a sample of eight MG lexical items used in the first Chinese grammar, which are essential in constructing a Chinese relative clause, namely a sequence of “Vt Noun de Noun” as SRs or “Noun Vt de Noun” as ORs. As always, all grammars used in this dissertation work use abstract lexical items such as “Noun” or “Vt” so that entropy calculations based on it reflect only to structural uncertainty, as opposed to word-choice uncertainty.

Here both Chinese grammars follow Aoun and Li (2003) in the sense that an NP rather than a DP is raised to the RC head position (c.f. Huang, Li, and Li (2009)). In this way, the determiner and the classifier modifying the head noun can be outside of the RC. The MG focuses on the argument NP relativization. At this moment, they do not cover relativization involving resumptive pronouns or adjuncts. Both Chinese grammar fragments employ the promotion analysis of Kayne (1994) such that the RC head is moved directly from the position of the gap into RC head position. In a typical derivation, driven by the *+wh* feature in

line 7, the *wh*-movement will raise an argument noun phrase headed by the lexical entry in line 2 out of the CP (in this case, C-RelP). In addition, line 8 of Table 5.1 follows Ou (2007) and treats the relativizer *de* as a LINKER (den Dikken & Singhapreecha, 2004) to derive the correct word order for Chinese RCs<sup>4</sup>. The *+f* feature of *de* motivates a sentence (here TP) to the Spec of a functional head (FP). In other words, the *-f* feature licenses a kind of movement that puts RCs on the left hand side of the NP they modify.

	Terminal	Syntactic feature	Note
1	Noun	headful, -case	noun
2	Noun	headful, -case, -wh	relativizable noun with <i>-wh</i> feature
3		=headful, Arg	relativizable noun is also an argument
4	Vt	Arg, +case, V	verb selecting an argument as object
5		=>V, =Arg, v	verb selecting an argument as subject
6		=v, +case, T, -f	tense with <i>-f</i> feature
7		=T, +wh, Crel	<i>+wh</i> feature hoisting noun phrase
8	de	=Crel, +f, F	<i>+f</i> feature selects RC to the left of relativizer <i>de</i>

Table 5.1: Sample of MG lexical entries used to construct a Chinese relative clause

Figure 5.3 illustrates how a Chinese object relative is derived using the MG fragment in Table 5.1. Starting from a noun with *-wh* feature chosen as the RC object, the Chinese grammar builds the RC structure step by step. It accurately records how *wh* and *f* feature-checkings (highlighted in bold) allow the relativization of noun phrase and derive the correct head-final word order in Chinese.

Section 5.1 highlights a variety of local ambiguities in Chinese RC prefixes. In order to test whether the Subject Advantage also holds in Chinese RCs, Jäger et al. (Submitted) uses the sentence-initial determiner-classifier-adverb sequence to encourage an expectation for relative clause structures (as opposed to

<sup>4</sup>*de* can also be treated as a modification marker according to Cheng and Sybesma (2005).



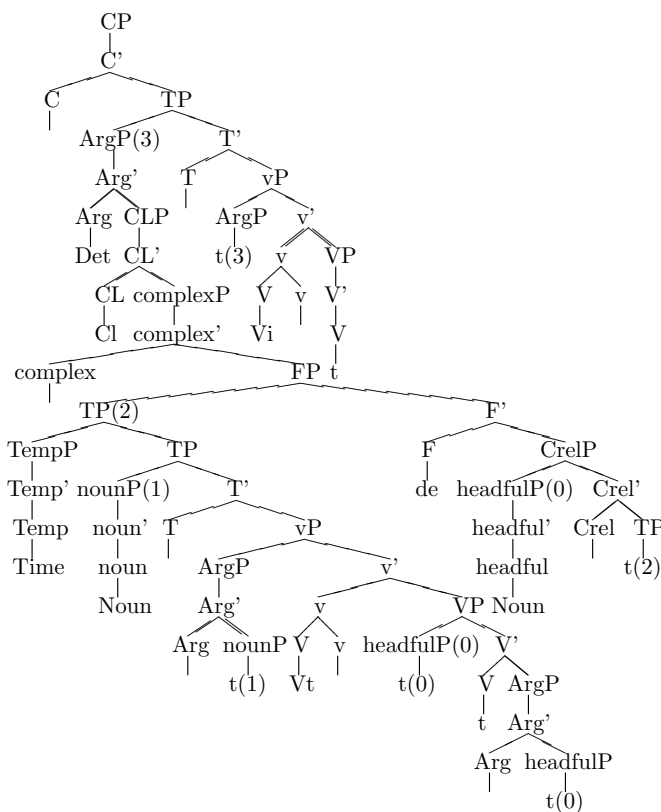


Figure 5.4: X-bar tree of “disambiguated” Chinese ORs in (5.8b) built through syntactic derivations in Figure 5.3

Terminal	Syntactic feature	Note	
1	AniN	aniheadful, -case	animate noun
2	InaniN	inaniheadful, -case, -inaniwh	relativizable inanimate noun
3		=aniheadful, Arg-ani	animate argument
4		=inaniheadful, Arg-rel-inani	relativizable inanimate argument
5	Vt	Arg-rel-inani, +case, V-OR	OR verb selecting its inanimate object
6		=>V-OR, =Arg-ani, v-OR	OR verb selecting its animate subject
7		=v-OR, +case, T-OR, -f	tense with <i>-f</i> feature
8		=T-OR, +inaniwh, Crel-OR-inani	+ <i>inaniwh</i> hoisting the inanimate noun
9	de	=Crel-OR-inani, +f, F-OR-inani	+ <i>f</i> feature keeping the RC head final

Table 5.2: Animacy-encoded MG fragment derives a Chinese OR with animate subject and inanimate object

The fragment of the second Chinese MG in Table 5.2 is enough to derive a Chinese OR with animate subject and inanimate object. Example 5.7 provides such an example with Chinese words.

- (5.7) 记者 绕开 的 大门 坏了  
 [ Jizhe raokai e<sub>i</sub> de ] damen<sub>i</sub> huale  
 reporter bypass DE gate was broken  
*'The gate which the reported bypassed was broken.'*

For simplicity and consistency with experimental conditions in Wu et al. (2012), the second Chinese MG with animacy does not include lexical entries for “disambiguation” materials used in the first Chinese grammar, such as “Det” and “Cl”. Compare the lexical entry in Line 2 with its counterpart Line 2 in Table 5.1, the generic *wh*-feature in the animacy grammar is divided into two categories (*-aniwh* and *-inaniwh*) depending on the animacy nature of the noun phrase. Line 5 suggests that a transitive predicate can be named as “V-OR” if it selects a relativizable inanimate noun phrase as its object. The noun phrase’s *-inaniwh* feature is checked by the *+inaniwh* feature of OR tense and hence projects an inanimate OR (or OR with inanimate head, “Crel-OR-inani”), as shown in Line 8. A derived X-bar tree of the target OR structure is illustrated in Figure 5.5.

The two MG fragments introduced above present us an opportunity to model reading difficulties observed in two recent Chinese RC experiments. But before calculating the Entropy Reduction values at each word, both grammars need to be weighed. The next section explains how weights of crucial parameters in the grammar are obtained from a Chinese corpus.

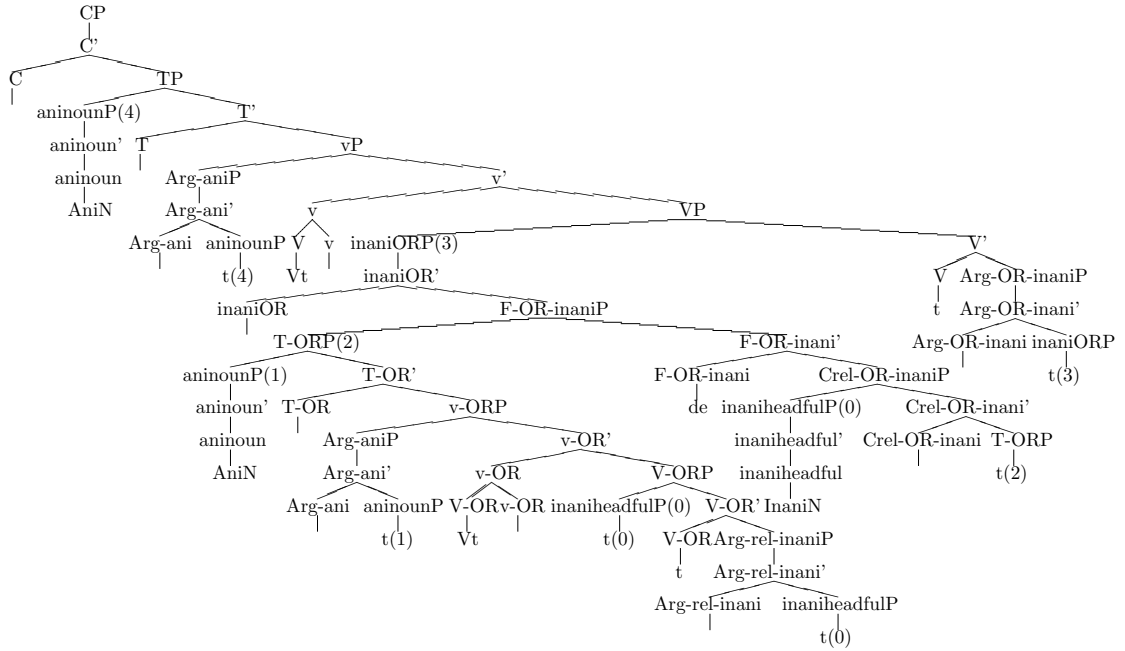


Figure 5.5: X-bar tree of inanimate Chinese ORs with animate RC subject in (5.7) built through syntactic derivations in Figure 5.2

### 5.3.2 Counting constructions in the Chinese treebank

The ER complexity metric derives processing difficulty, in part, from probabilities. This means we need to weight the prepared grammars with help from language resources. For the first Chinese MG, we obtain attestation counts from Chinese Treebank 7 (Xue, Xia, Chiou, & Palmer, 2005) which contains 51,447 fully parsed sentences or 1,196,329 words. These yield the weights shown below in Table 5.3. Note that during the transformation from MG to MCFG, the grammar itself determines the parameterization by defining the set of rewrites that are “contested.” These become choicepoints whose relative weights determine the probability of alternative subtrees of the derivation. This is done by adding up the products of attestation counts and rule applications (Chi, 1999). The “dis-

ambiguated” RCs shown in Example (5.8) on Page 108 motivate a somewhat richer set of choice points in the formal grammar fragment, which obligates us to estimate weights for a longer list of parameters than in English and Italian.

noun with a demonstrative modifier	2916
complex NP with a demonstrative modifier	345
noun in argument position	8133
complex NP in argument position	2316
possessive phrase in argument position	1866
headful SR	2281
headless SR	280
headful OR	830
headless OR	304
noun in subject position	3830
noun with a demonstrative modifier in subject position	167
<i>pro</i> in subject position	6385
noun in object position	3766
noun with a demonstrative modifier in object position	123
<i>pro</i> in object position	2
subject <i>pro</i> with transitive verb	5054
subject <i>pro</i> with intransitive verb	1331
subject NP with transitive verb	17250
subject NP with intransitive verb	4377
noun as OR subject	185
noun with a demonstrative modifier as OR subject	12
<i>pro</i> as OR subject	162
matrix modified by temporal adjunct	343
matrix not modified by temporal adjunct	16852
SR not modified by temporal adjunct	2532
OR not modified by temporal adjunct	1124
RC modified by temporal adjunct	39
relative clause	3695
complement clause	852

Table 5.3: Attestation counts from Chinese Treebank 7 are used to weight the grammar for “disambiguated” Chinese RCs

Table 5.3 reveals several key facts of Chinese RCs. It shows that SRs are much frequent than ORs in Chinese, similar to other languages including English and Italian. Interestingly, Chinese ORs are more likely to have a covert head than

Chinese SRs. As expected, subject-*pro* is extremely frequent in Chinese while object-*pro* is rare (Huang et al., 2009). *pro*-drop is also popular within the RC. An adequate grammar must allow for an OR's embedded subject to be omitted.

To properly weight the second Chinese grammar, a more fine-grained investigation of noun phrase animacy is necessary. Wu (2009) and Y. Hsiao and MacDonald (2013) hand-coded the animacy of the head noun and the embedded noun phrase in a subset of RCs taken from the Chinese Treebank. The Chinese animacy modeling work in this dissertation adopts key corpus counts from CTB 7 reported by Y. Hsiao and MacDonald (2013) in order to compare ER values directly with results based on the first Chinese grammar using CTB 7 counts. Table 5.4 is borrowed from Y. Hsiao and MacDonald (2013). It confirms that SRs are indeed more frequent than ORs in Chinese. In addition, head nouns and embedded nouns prefer to have contrastive animacy in both SRs and ORs. SRs tend to have more animate head whereas OR heads are more likely to be inanimate. Those results are consistent with the corpus counts in Wu (2009).

Y. Hsiao and MacDonald (2013) have considered the animacy of nouns in both RCs and simple sentences. In order to best model reading difficulties observed in Wu et al. (2012), this dissertation work also takes into account several additional animacy patterns, in particular, the animacy of nouns in possessive structures as well as the relationship between head noun animacy and the type of matrix verb in subject-modifying conditions.

Table 5.5 exemplifies how noun phrase animacy varies in Chinese possessive structures. Possessives are common in Chinese and can have the word *de* in between the possessor and the possessee. The "Noun *de* Noun" sequence is also a substring of an SR potentially confounding RC processing, as shown in

Modification Type	RC Type	RC verb	Head NP	Embedded NP	Count
Subject-modifying	SR	Vt	+anim	+anim	61
			+anim	-anim	209
			-anim	+anim	68
			-anim	-anim	31
	Vi	+anim	-	14	
		-anim	-	56	
OR	Vt	+anim	+anim	11	
		+anim	-anim	0	
		-anim	+anim	163	
		-anim	-anim	23	
Object-modifying	SR	Vt	+anim	+anim	30
			+anim	-anim	62
			-anim	+anim	27
			-anim	-anim	76
	Vi	+anim	-	34	
		-anim	-	74	
OR	Vt	+anim	+anim	10	
		+anim	-anim	0	
		-anim	+anim	79	
		-anim	-anim	23	

Table 5.4: Tokens of Chinese RCs with varies animacy patterns (Y. Hsiao & MacDonald, 2013, Table 3, Page 6)

Example 5.2 on Page 85. Therefore, it is possible that the animacy of nouns in possessive structures also affect readers' expectations at the embedded noun and the head noun of SRs. In order to gather precise counts of noun animacy in possessives, a total of 732 possessives with only bare nouns are extracted from CTB 7, as compared to 1866 of those with noun phrases in Table 5.3. Among four types of possessives with various animacy, the one with two inanimate noun phrases is most frequent. The second noun phrase, also the head of possessive structure, tends to be inanimate.

Besides the animacy of possessives, the present work also explores the re-

Example	Possessor	Possessee	Count
父母的朋友 parents DE friend <i>a friend of parents</i>	+anim	+anim	10
人民的自由 people DE freedom <i>people's freedom</i>	+anim	-anim	182
当地的村民 local DE villager <i>the local villager</i>	-anim	+anim	45
语言的艺术 language DE art <i>the art of language</i>	-anim	-anim	495

Table 5.5: Animacy pattern of Chinese possessive structures

lation between RC head animacy and the type of matrix verb, as required by the fine-grained Chinese MG fragment with animacy. Table 5.6 serves as a complement to the corpus counts of simple sentences with noun phrase arguments (see Table 2 of Y. Hsiao and MacDonald (2013)). It identifies the fact that an RC-modified head noun tends to have an intransitive main verb rather than a transitive one.

Matrix Verb	RC Type	Head NP	Count
Vt	SR	+anim	66
		-anim	40
	OR	+anim	13
		-anim	8
Vi	SR	+anim	101
		-anim	101
	OR	+anim	20
		-anim	3

Table 5.6: Animacy pattern of Chinese RC heads with different types of matrix verb

Using the corpus counts of various structures from Chinese Treebank 7,

the two Chinese MG fragments are weighted. The probabilistic grammar reflects Chinese readers' expectation on how to construct a phrase or a sentence, this time, with both syntactic and non-syntactic knowledge. CCPC can subsequently calculate the readers' sentence-medial expectations, incorporating both syntactic and non-syntactic knowledge. The fluctuation of the uncertainty degrees, formalized by ER, mirrors the incremental reading difficulty at each word. Sections 5.3.3 and 5.3.4 below present the Chinese modeling results.

### 5.3.3 Chinese prediction I: the Subject Advantage

Entropy Reduction is first used to model the subject-object asymmetry in "disambiguated" Chinese RCs such as Example (5.5), repeated below as (5.8). The regions of interest include the RC region (the V+N/N+V in SRs and ORs respectively) and the head noun. The derived Subject Advantage in Chinese RCs is shown in Figure 5.6. Comparing ER predictions with observed reading times in Figure 5.2 suggests that ORs are correctly categorized as harder to process, in both regions.

(5.8) a. "Disambiguated" Chinese SR

那个	昨天	邀请	富豪	的	官员	...
na-ge	zuotian	[ e <sub>i</sub> yaoqing	fuhao	de ]	guanyuan <sub>i</sub>	...
that-Cl	yesterday	[ invite	tycoon	DE ]	official	...

*'The official who invited the tycoon yesterday ...'*

b. "Disambiguated" Chinese OR

那个	昨天	富豪	邀请	的	官员	...
na-ge	zuotian	[ fuhao yaoqing	e <sub>i</sub> de ]	guanyuan <sub>i</sub>	...	...
that-Cl	yesterday	[ tycoon invite	DE ]	official	...	...

*'The official who the tycoon invited yesterday ...'*

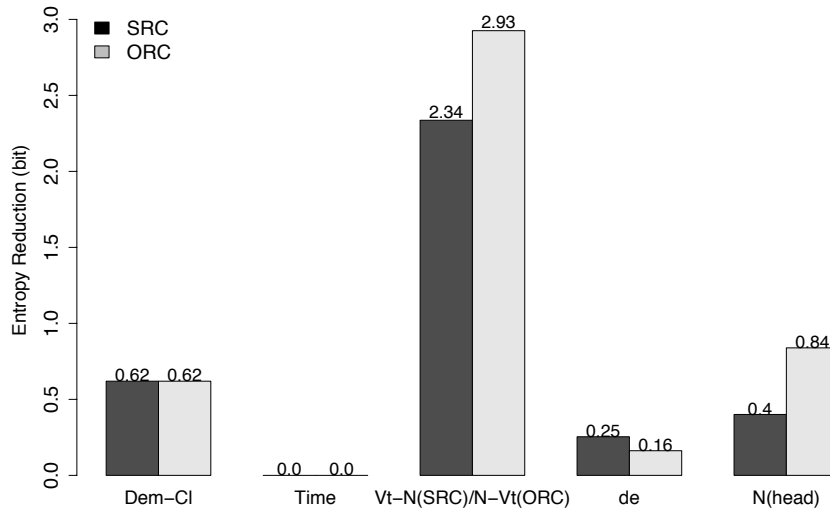


Figure 5.6: ER predictions of “disambiguated” Chinese RCs

The Subject Advantage here could also be explained by the Tuning Hypotheses (Mitchell et al., 1995; Jurafsky, 1996). The explanation would be that SRs occur more frequently than ORs and therefore are more expected and easier to process. However, this intuitive idea only goes so far. The Entropy Reduction values calculated by CCPC, as an instantiation of a particular expectation-based theory, furnish linguistic-driven explanations behind different levels of processing difficulty.

ER measures the reduction of uncertainty, in other words, the cognitive effort expended on disambiguation. Therefore, it is natural to ask how it predicts the correct processing pattern for already “disambiguated” Chinese RCs. The answer is that sentences like (5.5) still have residual ambiguities. Although the presence of the determiner-classifier and the temporal phrase serve to reduce ambiguity greatly, there still exist other RC-like structures that could be viable alternatives to the globally correct analysis, for example, RCs with dif-

ferent types of verb. The ER account is that RTs reflect work that is done to incrementally decide between these alternatives. Because this disambiguation is defined over weighted syntactic alternatives, the overall proposal can be seen as a very particular type of structural frequency theory.

The CCPC allows psycholinguists to look at the probability distribution on syntactic alternatives at each prefix. Figure 5.7 illustrates how the Subject Advantage is predicted in Chinese RCs region-by-region. After processing the two-word RC region, entropy reduces by 2.34 bits in the SR case and 2.93 bits in the OR case. Looking at the boxes in the “RC region” row, we focus on the two initial substrings *Dem Cl Time Vt N* and *Dem Cl Time N Vt*.

The greater Entropy Reduction in the OR case as compared to the SR case quantifies the intuitive idea that more comprehension work is called for in the RC region of one item as compared to the other. This numerical contrast reflects the fact that, by the end of the RC region, the OR parser state is more organized than the corresponding SR state. Table 5.7 quantifies this degree of organization by counting the remainders at various probability thresholds. More peaked probability distribution of remainders leads to lower uncertainty level in ORs. Starting from the same entropy level (7.448 bits in the first row of Figure 5.7) but arriving at a lower “destination” level (second row of Figure 5.7) in just the OR means that the OR-initial substring goes farther in reducing this ambiguity. In other words, since more work has been done, the prediction of greater processing effort in ORs will be made via the Entropy Reduction hypothesis.

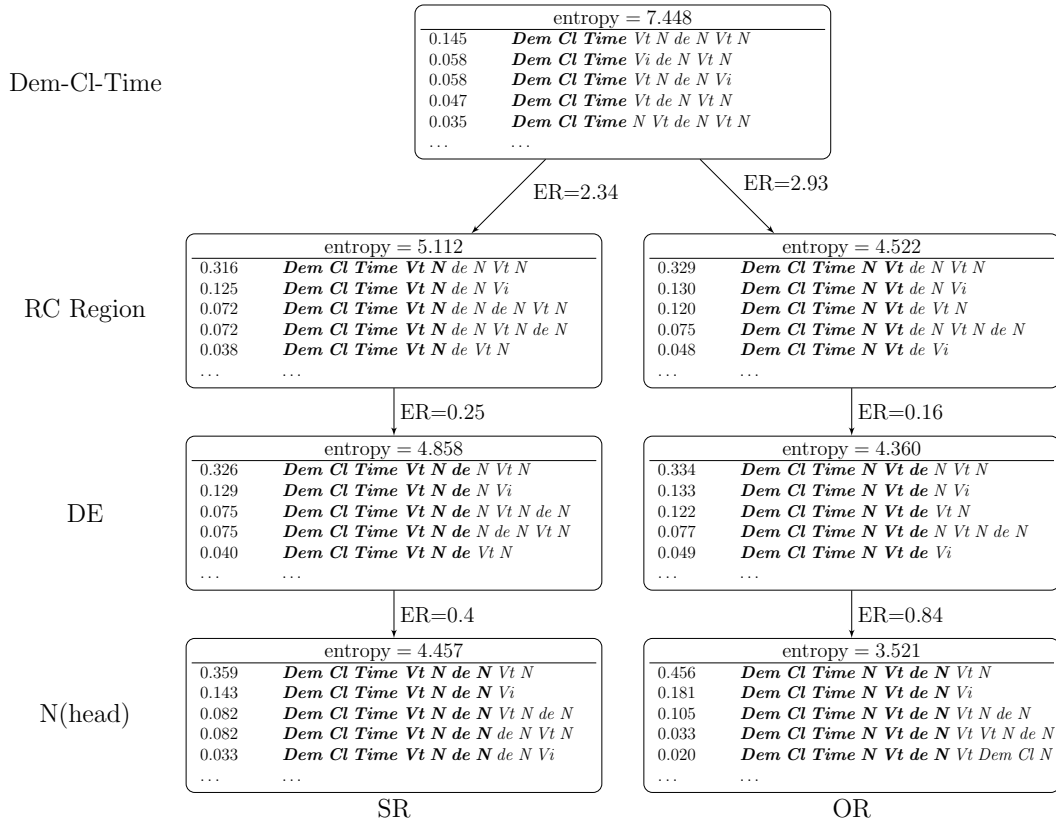


Figure 5.7: Derivations and their conditional probabilities in “disambiguated” Chinese RCs

Probability	Type	No. of Remainders	Total Probability	Uncertainty
> 0.01	SR	10	0.722	High
	OR	8	0.768	Low
> 0.001	SR	64	0.881	High
	OR	54	0.908	Low

Table 5.7: The frequency distributions of syntactic alternatives after reading the RC region is different between SRs and ORs (second row of Figure 5.7).

Earlier modeling work on Chinese RCs using Surprisal did not derive the Subject Advantage at the RC head noun (Z. Chen, Grove, & Hale, 2012); however, the present work with Entropy Reduction does this. At the head noun, ER

value of SRs is 0.44 bits smaller than that of ORCs. This prediction reflects, in part, the possibility of headless RCs in Chinese. Examining the pre-head syntactic remainders in Figure 5.8, the grammar defines contrasting probabilities for headlessness in SR vs OR contexts. In ORCs, as shown in the bottom left box, there is at least a 20.8% chance that the initial substring will continue as a headless RC. On the other hand, it is less likely (5.6%) that an SRC-initial substring will be headless (top left box). The information-value of the headlessness question itself is different across the two constructions. It is also consistent with the corpus counts in Table 5.3 where ORCs are found to be more likely to have an omitted head and therefore contributes to the prediction of a Subject Advantage at the head word in Chinese.

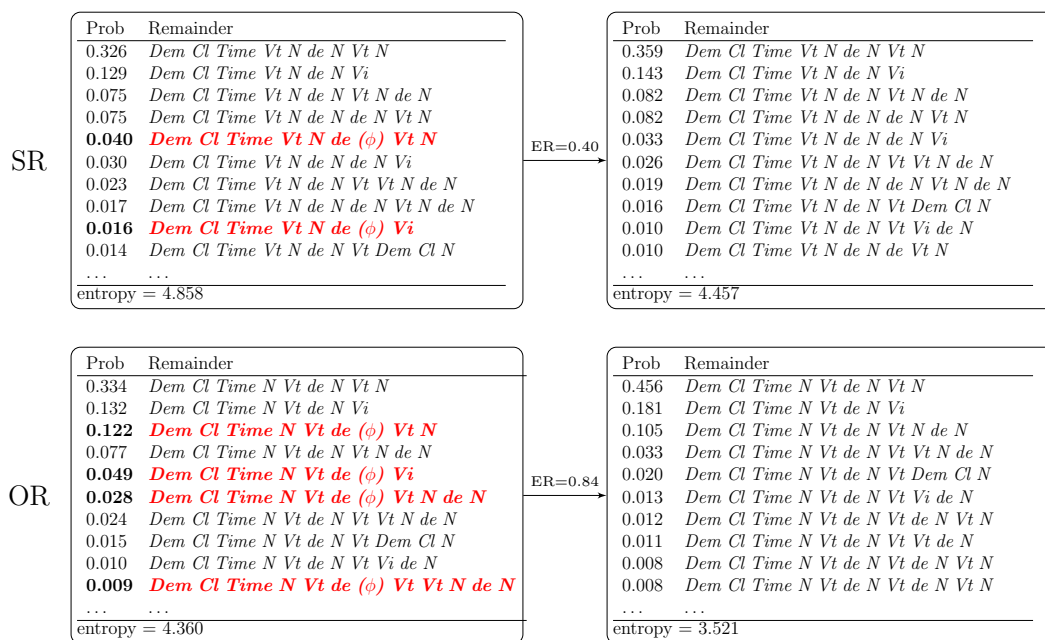


Figure 5.8: ER at the head noun in Chinese RCs; detail of last two rows in Figure 5.7. Here, expectations that the RC will be headless are highlighted in red.

### 5.3.4 Chinese prediction II: the head noun animacy

Section 5.3.3 above provides a successful example of using CCPC to model the Subject Advantage in Chinese RCs when competing temporary ambiguities are taken out. ER predictions are consistent with the reading time pattern observed in Jäger et al. (Submitted) in both the RC region and the head noun. This section takes the modeling of Chinese RC processing one step further with a focus on the animacy effect, which has been precisely described as a set of Animacy Preference Constraints in (5.6), repeated below.

(5.9) Animacy Preference Constraints (Wu et al., 2012, Page 1495)

- i. Subjects tend to be animate;
- ii. Head nouns in object-extracted RCs tend to be inanimate;
- iii. As a joint consequence of (i) and (ii), a contrastive animacy configuration tends to occur in object-extracted RCs with inanimate heads and in subject-modifying, subject-extracted RCs with animate heads.

The animacy constraints above are generalized from corpus count observations (Pu, 2007; Wu, 2009), tested in reading experiments (Wu, 2009; Wu et al., 2012), and recently modeled by a connectionist simple recurrent network (Y. Hsiao & MacDonald, 2013). The modeling work reported in this section contributes to the ongoing investigation by integrating corpus counts, the “formalist” grammar and the complexity metric Entropy Reduction. The advantage of CCPC, as previous chapters have repeatedly shown, is to provide insights into how legitimate syntactic alternatives sharing the common prefix lead to

different uncertainty levels when processing each word in the sentence. The animacy-encoded Chinese grammar, in a way similar to the English and Italian grammars, further allows us to take non-syntactic information into account.

Table 5.8 presents ERs at both the RC region and the head noun in Chinese RCs with eight different animacy patterns. CCPC calculates those predicted reading difficulties based on an MG fragment (Table 5.2) which specifies noun phrase animacy and is weighted by corpus counts in Tables 5.5 and 5.6 in addition to those in Y. Hsiao and MacDonald (2013). The animacy pattern of Chinese possessives and the interaction between head animacy and the type of matrix verb will partially account for the modeling results reported below. ER predicts that SRs with animate head and ORs with inanimate head (highlighted in red) are easier at the head noun than their counterparts. It also suggests an interaction between the head animacy and the animacy of embedded noun in SRs, namely those with animate head and inanimate RC object are the easiest across four SR conditions.

RC Type	Head NP	Embedded NP	ER (bits)	
			RC region	RC Head
SR	+anim	+anim	2.27	1.45
	+anim	-anim	2.67	1.19
	-anim	+anim	2.27	1.47
	-anim	-anim	2.67	1.57
OR	+anim	+anim	2.35	1.46
	+anim	-anim	2.48	1.44
	-anim	+anim	2.35	1.28
	-anim	-anim	2.48	1.26

Table 5.8: ER predictions of Chinese RCs with different animacy patterns

Table 5.9 details ER differences between the animate head and the inanimate head in SRs. This diagram shows parser states arising during the processing of an SR with an inanimate object. This corresponds to rows 2 and 4 of the SR

category in Table 5.8. These two constructions share a common initial substring, namely “Vt InaniN de”. At this point, both conditions have the same level of uncertainty, 4.087 bits. After reading the next word, the syntactic alternatives for the SR prefix with animate head (the bottom left box) become more diversified. This leads to a smaller ER value of 1.19 bits such that less disambiguation work has been done. In contrast, when the SR head is inanimate, ER is larger, i.e., 1.57 bits, because the competition among possible remainders is weaker (the bottom right box). Looking into different alternatives in that box explains how the lower level of uncertainty with an inanimate head is calculated. As discussed in Section 5.1, the Chinese relativizer “de” can also occur in a possessive structure. In addition, corpus counts in Table 5.5 reveal that a possessive structure with two inanimate noun phrases are very frequent. Consequently, the surface string of the most probable syntactic derivation given the prefix “Vt InaniN de InaniN” is the string itself, a subject-dropped simple sentence with a possessive object. This is recorded as the top one in the bottom right box with a probability of 0.557. Since none of probabilities of other alternatives are larger than 0.2, an SR with an inanimate head is less ambiguous.

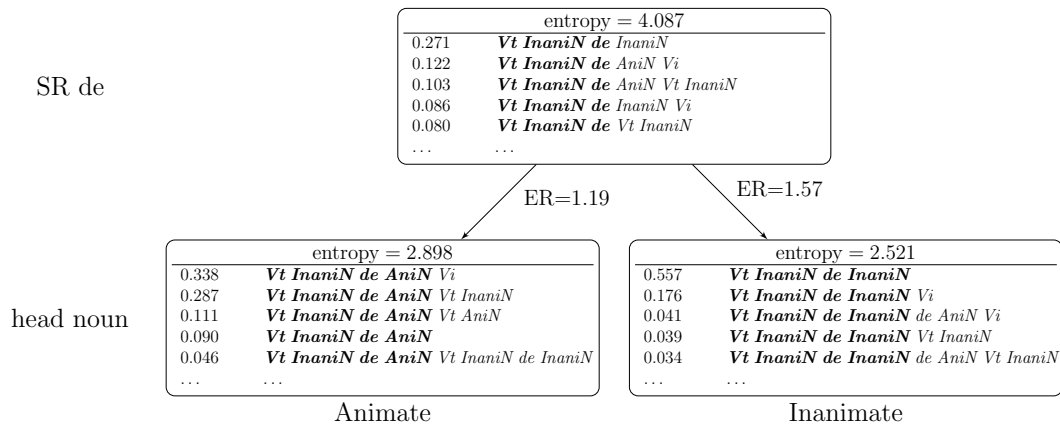


Figure 5.9: SR transitions at heads with different animacy

In addition to the animacy of Chinese possessives, this modeling work also explores the frequency distribution between head animacy and the main verb type. This issue in fact partly explains the animacy effect at the OR head. Figure 5.10 illustrates the reductions of uncertainty before and after two OR heads with different animacy, which correspond to Rows 1 and 3 of the OR category in Table 5.8. Different from the pattern predicted for SRs, integrating an inanimate head with the OR prefix is easier. From the same start point “AniN Vt de”, the OR with inanimate head has a flatter distribution of syntactic alternatives, resulting in a higher uncertainty degree of 1.828 bits. This is because ORs with inanimate head, rather than those with animate head, are more likely to be subject of a transitive main verb, as attested in the Chinese Treebank (Table 5.6). Therefore the competition between “Vt” or “Vi” continuations is stronger in the inanimate condition, as exemplified by the top two candidates in the bottom right box.

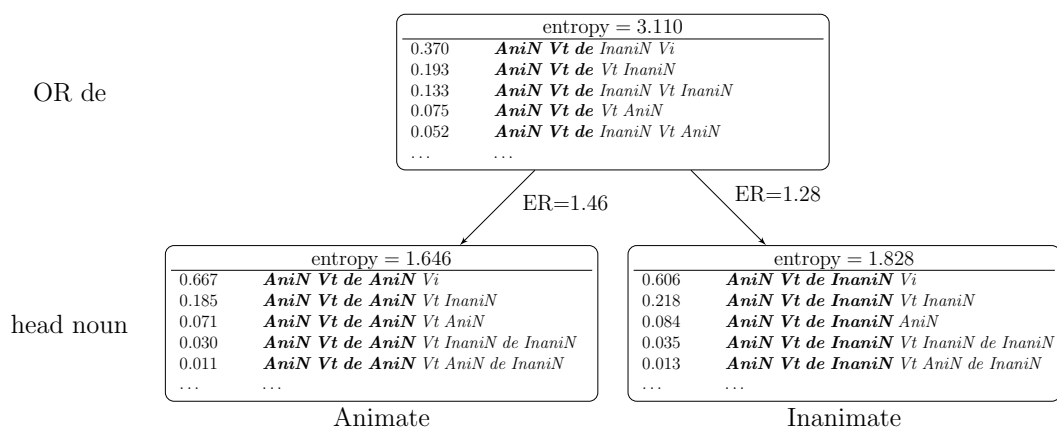


Figure 5.10: OR transitions at heads with different animacy

Comparing the modeling results with the reading time data of Wu et al. (2012), ER accurately predicts the animacy effect at the head noun in both Chinese SRs and ORs. SRs with animate heads and ORs with inanimate heads are easier. However, the modeling work here also has its limits. One interesting observation in Wu et al. (2012) was that the effect of head noun animacy sometimes showed up early in the sentence, often before reaching the RC head. The modeling work in its current form is not able to predict the effect before the head noun simply because ER values are the same given the same prefix. The effect of the head noun can only be evaluated after knowing the animacy of it. The modeling described in this section also does not predict any animacy effect at the embedded noun. We think the reason may be due to how structures are derived by the MCFG with subcategorized features. For example, a pair of MCFG rules below describes how the OR embedded verb selects its relativizable object. The probability varies between choosing an inanimate object and an animate one. However, *wh* features with different animacy make the left-hand side of those rules different. Therefore, with the same 1.0 probability, those two rules are incomparable.

1.0 (: +case V-OR; -case -**aniwh**) → (:: =Arg-rel-ani +case V-OR) (: Arg-rel-ani; -case -**aniwh**)  
 1.0 (: +case V-OR; -case -**inaniwh**) → (:: =Arg-rel-inani +case V-OR) (: Arg-rel-inani; -case -**inaniwh**)

This issue may be solved in the future by employing a more complicated probability model for Minimalist Grammars (Hunter & Dyer, 2013).

## CHAPTER 6

### CONCLUSIONS

Chapters 3 to 5 have demonstrated that the reading difficulty pattern found in relativized structures of English, Italian, and Chinese are connected with the reductions of structural uncertainty between words. In addition, the results fulfill the research objective of modeling the animacy effect at the head noun in all three languages. Most importantly, weighting grammar rules with the frequency distribution of various constructions in animacy-annotated treebanks, the CCPC system examines the “remainder set” of syntactic alternatives at each word. This advantage has allowed me to propose many specific and plausible explanations to account for the observed animacy effect. It therefore moves the general frequency-driven approach (Roland et al., 2007; Reali & Christiansen, 2007) forward towards finer-grained linguistic analyses.

In this Chapter, Section 6.1 compares the role that the animacy cue plays cross-linguistically by computing its cue weight in three languages, Section 6.2 discusses directions for future research, and Section 6.3 concludes.

#### **6.1 Compare the animacy cue weight in three languages**

As discussed in Chapter 1, one of the motivations of this dissertation is the pioneering work of MacWhinney et al. (1984). At the beginning of that paper, they expressed concern about evaluating sentence processing principles solely based on the study of English:

“There is a danger that we may find ourselves proposing generaliza-

tions about languages as a whole that turn out to be nothing more than facts about English.”

This concern is well-taken by following studies in the field of psycholinguistics, as we see more and more of them comparing results across languages.

In terms of understanding the role that a linguistic cue plays in sentence processing, one natural approach would be to put it in a cross-linguistic setting. This section compares the importance of the animacy cue in the processing of English, Italian, and Chinese relative clauses. The ranking of cues proposed by MacWhinney et al’s competition model successfully depicts that the rankings of cues are not the same across languages. However, this model has its limits. First, the cue validity invoked in their model are not quantified numerically. Second, the rankings are language-specific and thus not directly comparable between languages. Not to mention that their results were based on off-line experiments which may be quite different from those coming out of online sentence processing studies. Therefore, a new schema needs to be set up that will allow us to compare the processing effects in different languages driven by the same cue, for example, animacy.

Fortunately, the computed ER values in previous chapters have paved a way to achieve this goal. The idea is simple. The language-specific animacy cue weight can be quantified by comparing ER results calculated based on two weighted grammars, one of which specifies the noun phrase animacy in its grammar rules. An example from English is given below. Chapter 3 employs ER to model the processing difficulty profile in two set of English RCs, repeated here as (6.1) and (6.2).

- (6.1) a. Subject relatives (SRs)  
 The reporter<sub>*i*</sub> who *e<sub>i</sub>* attacked the senator left the room.
- b. Object relatives (ORs)  
 The reporter<sub>*i*</sub> who the senator attacked *e<sub>i</sub>* left the room.
- (6.2) Experimental conditions of Traxler et al. (2002)
- a. SR, +animate, –animate  
 The director that watched the movie received a prize.
- b. OR, +animate, –animate  
 The director that the movie pleased received a prize.
- c. SR, –animate, +animate  
 The movie that pleased the director received a prize.
- d. OR, –animate, +animate  
 The movie that the director watched received a prize.

The animacy cue weight  $W_{animacy}$  reflects the magnitude of animacy effect, as illustrated in the following formula (6.3). Consider the English example, the two conditions in (6.1) do not specify the animacy whereas (6.2) includes four conditions with contrastive animacy between the head NP and the embedded NP. If all possible animacy manipulations are considered, there should be ten animacy conditions in total, as shown in Table 6.1. Formula (6.3) computes the weighted arithmetic mean for ER values of two sets of conditions separately. For English, there are  $spec = 10$  animacy conditions and  $unspec = 2$  animacy-unspecified conditions. The ratio between two weighted average ERs gives us a hint on how ER changes before and after adding the animacy information,

which mirrors how the knowledge of animacy will help processing an RC construction.

$$W_{animacy} = \frac{\sum_1^{spec} P_{spec} \times ER_{spec}}{\sum_1^{unspec} P_{unspec} \times ER_{unspec}} \quad (6.3)$$

The same process could be done for Italian and Chinese as well. Table 6.1 records the complete picture on the animacy effect in this study. In this table, the numbers under the ER column are the sum of ER values at both the RC region and the head noun. The head noun ERs are included since they are proven to be distinctive in prenominal Chinese RCs. Note that the ERs for animacy-unspecified Chinese RCs are not what have been reported in Section 5.3. These values (2.64 for SRs and 2.59 for ORs) in Table 6.1 reflect the processing difficulty of standard Chinese RC stimuli without disambiguation so that they are comparable to ERs calculated for animacy-specified Chinese RCs. For the purpose of computing the animacy cue weight, this table uncovers more animacy conditions that are not included in the discussion of previous chapters. For example, SR verbs can be intransitive such that there is no embedded object available. Subject *pro*-drops are allowed in Italian and Chinese ORs. In the table, each condition is accompanied by its relative weight in the fraction form. Note that due to the size of Italian treebank, there are a few RC types with zero occurrence. An extremely low weight are assigned in those cases.

The rightmost column of Table 6.1 lists the computed animacy cue weight in each of the three target languages. English has the highest animacy cue weight whereas Italian has the lowest one. Although it seems to be inconsistent with what MacWhinney and colleagues originally argued, the result is not entirely surprising given that the animacy cue is evaluated in different scenar-

Language	Type	Head NP	Embedded NP	ER	Relative Freq.	Weight
English	SR	unspecified	unspecified	1.70	6817 / 10027	1.609
	OR	unspecified	unspecified	3.86	3210 / 10027	
	SR	+anim	+anim	4.25	25 / 2274	
		+anim	-anim	3.40	79 / 2274	
		-anim	+anim	4.19	33 / 2274	
		-anim	-anim	3.33	43 / 2274	
		+anim	N/A	4.32	351 / 2274	
		-anim	N/A	4.45	552 / 2274	
	OR	+anim	+anim	3.58	130 / 2274	
		+anim	-anim	4.40	1 / 2274	
-anim		+anim	3.46	930 / 2274		
-anim		-anim	3.33	130 / 2274		
Italian	SR	unspecified	unspecified	1.96	430 / 473	1.426
	OR	unspecified	unspecified	2.15	43 / 473	
	SR	+anim	+anim	2.22	1 / 93	
		+anim	-anim	2.39	13 / 93	
		-anim	+anim	2.88	4 / 93	
		-anim	-anim	3.04	31 / 93	
		+anim	N/A	2.54	14 / 93	
		-anim	N/A	3.20	18 / 93	
	OR	+anim	+anim	2.55	1 / 93	
		+anim	-anim	2.55	1 / 93	
		-anim	+anim	3.20	4 / 93	
		-anim	-anim	3.20	1 / 93	
+anim		<i>pro</i>	1.47	3 / 93		
-anim		<i>pro</i>	2.27	2 / 93		
Chinese	SR	unspecified	unspecified	2.64	2561 / 3695	1.450
	OR	unspecified	unspecified	2.59	1134 / 3695	
	SR	+anim	+anim	3.72	91 / 1214	
		+anim	-anim	3.85	271 / 1214	
		-anim	+anim	3.74	95 / 1214	
		-anim	-anim	4.23	107 / 1214	
		+anim	N/A	4.27	48 / 1214	
		-anim	N/A	5.08	130 / 1214	
	OR	+anim	+anim	3.81	21 / 1214	
		+anim	-anim	3.92	1 / 1214	
		-anim	+anim	3.63	242 / 1214	
		-anim	-anim	3.74	46 / 1214	
+anim		<i>pro</i>	2.50	12 / 1214		
-anim		<i>pro</i>	2.68	150 / 1214		

Table 6.1: Predicted ERs at the “RC+head” region and the relative frequency of various RC conditions derive the animacy cue weights for English, Italian, and Chinese.

ios. MacWhinney et al. employed an offline experiment asking subjects to state which of the two nouns was the actor of a simple sentence. On the contrary, the modeling results reported here are based on online reading experiments using relative clauses. The fact that experiments in English such as Traxler et al. (2002, 2005) have demonstrated that readers do rely on their animacy knowledge to understand an RC matches the larger animacy cue weight computed for English. If the assumption of English is correct, the Italian result can also be justified since there exists contradictory experimental evidence on whether the animacy cue could facilitate RC processing. The animacy weight calculated based on Chinese RC processing is slightly higher than Italian, but lower than English, suggesting that the animacy cue could still be effective during the processing of Chinese RCs. However, further investigations on processing structures other than RCs are necessary.

## 6.2 Future Work

Although the dissertation research accurately derives both the subject-object asymmetry and the animacy effect at the head noun for all three languages, there are a few testable hypotheses remain unexplored. Some of those have already been discussed in previous chapters. This section adds a few more items to the list of future work.

A simplified and accurate weighting method needs to be in place for the CCPC system. As briefly mentioned in Chapter 2, CCPC currently allows automatic weighting for an MCFG. However, its present form is unable to deal with ambiguous sentences in the “mini-treebank”. If there is an ambiguous string in

the training dataset, the parser will only consider one of its syntactic analyses and therefore estimate incorrect weights for certain grammatical rules. In order to obtain accurate weighted grammars in this work, relative frequency of a construction is manually added to the MCFG file. However, this process is time-consuming and requires deep knowledge on the MG and MCFG formalisms. In order to speed up the modeling process and allows more applications of CCPC for the psycholinguistic community, the system should have the flexibility to read in a training dataset with unambiguous tree-formatted constructions rather than surface strings.

In English, experimental evidence has supported a structural-frequency driven account for the subject and object asymmetry. One representative work comes from the research on pronominal relative clauses (Reali & Christiansen, 2007). In general, subject relatives are more frequent than object relatives. However, according to Reali and Christiansen's corpus study, object relative clauses are more frequent when the embedded NP is pronominal. They manipulated the embedded NP with four different types of pronoun, namely the second-person pronoun "you", the singular first-person pronoun "I", the plural third-person pronoun "they", and the impersonal pronoun "it". In all four self-paced reading experiments, object relatives with a pronominal subject were found to be easier than their SR counterparts. They argued that the pronominal object relative advantage is due to the higher frequency of "that pronoun verb" chunks than "that verb pronoun" chunks. The English modeling work in Chapter 3 has already taken the pronoun into consideration, both in the grammar and in the corpus counts. For example, the corpus study results in Table 3.6 on Page 39 include the frequency distribution of different verb categories given a pronoun subject. To model reading time data of Reali and Christiansen (2007), the new

MG fragment has to specify the person feature of a pronoun and the corpus study needs to be refined accordingly.

One problem that occurs during the modeling of Italian relatives is its relatively free order, at least in the relative clauses. In particular, the embedded subjects can be placed either before or after the embedded verb. Indeed, Di Domenico and Di Matteo (2009) tested two types of Italian OR in their study, as repeated below in (6.4). In fact, they called Condition b “the canonical” version of Italian ORs <sup>1</sup>.

(6.4) b. Object RC with post-verbal subject

*Questo è il poliziotto<sub>i</sub> che hanno seguito i ladri e<sub>i</sub>*  
This is the policeman that **have** pursued the robbers **e<sub>i</sub>**  
*‘This is the policemen that the robbers pursued.’*

d. Object RC

*Questo è il poliziotto<sub>i</sub> che i ladri hanno seguito e<sub>i</sub>*  
This is the policeman that the robbers **have** pursued **e<sub>i</sub>**  
*‘This is the policemen that the robbers pursued.’*

The main goal of introducing ORs with post-verbal subjects was to create a comparable condition to Italian SRs because they share a similar prefix with only the number-feature difference at the auxiliary verb. Utzeri (2007) also tested Italian ORs with post-verbal subjects in a sentence production study. She reported that Italian children produced more ORs with pre-verbal subjects easier than ORs with post-verbal subjects. She attributed this result to the Minimal Chain Principle of De Vincenzi (1991) because the former involve only one dependency whereas the latter involves two with an extra dependency between the *pro* at

---

<sup>1</sup>Contrastively, we ask three Italian speakers to judge whether post-verbal subjects are allowed in Italian RCs. None of them think this condition is possible.

the pre-verbal position and the post-verbal subject. To model those experimental data on this particular Italian ORs requires a grammar which offers special treatment for NP movement within the RC and additional corpus enquiries.

Lastly, the ER model can be extended to discuss the broader role of “discourse” as a potential explanation for supposed processing universals. Roland, Mauner, O’Meara, and Yun (2012) argue for the importance of discourse information used in processing RCs and sentence processing in general. They proposed a different account to explain the inverse processing pattern found in pronominal RCs (Reali & Christiansen, 2007). Roland and colleagues claimed that the distributional frequency of word chunks in RCs (i.e. NPs’ pronominal status) is in fact a result of discourse influences (i.e. NPs’ the informational status). ORs tend to have a “discourse-old” referent (e.g. pronoun) as the embedded subject whereas SRs tend to have a “discourse-new” referent as the embedded object. Comprehenders have “discourse expectations” on the RC structure. A violation of expectations will lead to a higher degree of reading difficulty. The background theory behind this “discourse expectation” idea is the corpus study of spoken conventional English in Fox and Thompson (1990). They argued that ORs are more likely to have a pronoun rather than a full NP is because the embedded NP refers back to a referent in the ongoing discourse, a process known as “grounding”. The “discourse-new/old” idea of Roland et al. (2012) is also compatible with the “Topichood hypothesis” on RC processing proposed for Dutch relatives (Mak et al., 2002). Leaving the details behind, the conclusion is that there exists a limitation on the referent of this discourse-old NP/pronoun. It could not be any individual mentioned in the context. It is more likely to be the topic on the ongoing discussion. All these studies mentioned above are rooted in Givenness, which is similar to the aforementioned priming effect in

functional linguistics (Givón, 1983).

ER, as an expectation-based processing model, has the flexibility to allow the discourse information as an ingredient. As long as the expectation of ongoing discourse referents can be quantified as uncertainty degree. Using the same technique, we anticipate no particular difficulty using ER to derive the processing pattern explained by discourse expectation hypothesis. As for specific context-based studies, such as (Roland et al., 2012), the modeling needs to consider the co-occurrence between the embedded NP type and the context type (i.e. topic, mention, neutral) in the corpus. This co-occurrence information could either be encoded as more detailed subcategorized grammar rules or some kind of penalty during parsing, that is, as long as there is no violation of discourse expectation, there will be no extra burden in processing. Otherwise, a penalty has to be added.

### **6.3 Conclusion**

This dissertation tackles one of the most fundamental research topics in the field of human sentence processing, that is, the interaction between syntactic and non-syntactic cues in incremental sentence comprehension. It examines the role that a non-syntactic cue, animacy, plays in processing relative clauses. It develops ideas first introduced in the Competition Model of MacWhinney et al. in the late 1980s such that the weight of a linguistic cue varies among languages. This work addresses this problem from an information-theoretic perspective. Based on corpus studies of the animacy frequency distribution using a computational system for modeling incremental reading difficulties, the complexity metric, En-

tropy Reduction, predicts both the subject-object asymmetry and the animacy effect observed at the head noun of RCs in English, Italian and Chinese. The results are consistent with the intuitive structural frequency idea, namely that a frequent structure is easier to comprehend. However, it takes this idea further by highlighting the particular disambiguation decisions that contribute to predicted difficulty. This line of modeling work not only provides finer-grained explanation for the animacy effect in relative clause processing reported in previous literature, but also contributes to link an important functional notion, animacy, to the formal linguistic framework in a unique way.

## APPENDIX A

### LIST OF ABBREVIATIONS

AH	Accessibility Hierarchy
CCPC	Cornell Conditional Probability Calculator
DLT	Dependency Locality Theory
ER	Entropy Reduction
MCFG	Multiple Context Free Grammar
MG	Minimalist Grammar
OR	Subject Relative Clause
PCFG	Probabilistic Context-Free Grammar
RC	Relative Clause
SR	Subject Relative Clause

## REFERENCES

- Adani, F., van der Lely, H. K., Forgiarini, M., & Guasti, M. T. (2010). Grammatical feature dissimilarities make relative clauses easier: A comprehension study with Italian children. *Lingua*, *120*(9), 2148–2166.
- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological review*, *111*(4), 1036.
- Aoun, J., & Li, A. (2003). *Essays on the representational and derivational nature of grammar: the diversity of wh-constructions*. Cambridge, MA: MIT Press.
- Arosio, F., Adani, F., & Guasti, M. T. (2009). Grammatical features in the comprehension of Italian relative clauses by children. *Merging Features: Computation, Interpretation, and Acquisition*, Oxford University Press, Oxford, 138–155.
- Arosio, F., Guasti, M. T., & Stucchi, N. (2011). Disambiguating information and memory resources in children's processing of Italian relative clauses. *Journal of psycholinguistic research*, *40*(2), 137–154.
- Bar-Hillel, Y., Perles, M., & Shamir, E. (1964). On formal properties of simple phrase structure grammars. In *Language and information: Selected essays on their theory and application* (pp. 116–150). Reading, Massachusetts: Addison-Wesley.
- Bates, E., McNew, S., MacWhinney, B., Devescovi, A., & Smith, S. (1982). Functional constraints on sentence processing: a cross-linguistic study. *Cognition*, *11*, 245–299.
- Belletti, A., & Chesi, C. (2011). Relative clauses from the input: syntactic considerations on a corpus-based analysis of Italian. In *STiL-Studies in Linguistics* (Vol. 4).

- Belletti, A., & Contemori, C. (2009). Intervention and attraction. on the production of subject and object relatives by italian (young) children and adults. In *Language acquisition and development. proceedings of gala* (pp. 39–52).
- Belletti, A., Friedmann, N., Brunato, D., & Rizzi, L. (2012). Does gender make a difference? comparing the effect of gender on children's comprehension of relative clauses in hebrew and italian. *Lingua*, 122(10), 1053–1069.
- Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 279–360). John Wiley.
- Bianchi, V. (2002). Headed relative clauses in generative syntax. part ii. *Glott International*, 6(8), 1–13.
- Blachman, N. (1968). The amount of information that y gives about x. *IEEE Transactions on Information Theory*, IT-14(1), 27–31.
- Bosco, C., Lombardo, V., Vassallo, D., & Lesmo, L. (2000). Building a Treebank for Italian: a Data-driven Annotation Schema. In *Proceedings of the 2nd LREC* (pp. 99–105).
- Bowman, S. R., & Chopra, H. (2012). Automatic animacy classification. In *Proceedings of the 2012 conference of the north american chapter of the association for computational linguistics: Human language technologies: Student research workshop* (pp. 7–10).
- Carreiras, M., Duñabeitia, J. A., Vergara, M., de la Cruz-Pavía, I., & Laka, I. (2010). Subject relative clauses are not universally easier to process: Evidence from basque subject relative clauses are not universally easier to process: Evidence from basque. *Cognition*, 115(1), 79–92.
- Chen, B., Ning, A., Bi, H., & Dunlap, S. (2008). Chinese subject-relative clauses are more difficult to process than the object-relative clauses. *Acta Psychologica*, 129(1), 61–65.

- Chen, Z., Grove, K., & Hale, J. (2012). Structural expectations in Chinese relative clause comprehension. In J. Choi, E. A. Hogue, J. Punske, D. Tat, J. Schertz, & A. Trueman (Eds.), *Proceedings of the 29th West Coast Conference on Formal Linguistics (WCCFL-29)* (pp. 29–37). Somerville, MA: Cascadilla Proceedings Project.
- Chen, Z., Jäger, L., & Hale, J. (2012). Uncertainty reduction as a predictor of reading difficulty in Chinese relative clauses. In Y.-O. Biq & L. Chen (Eds.), *Proceedings of the 13th International Symposium on Chinese Languages and Linguistics (IsCLL-13)* (pp. 245–261). Taipei, Taiwan.
- Cheng, L. L.-S., & Sybesma, R. (2005). A Chinese Relative. In H. C. Broekhuis, N. Huybregts, R. Kleinhenz, U. Kleinhenz, & J. Koster (Eds.), *Organizing Grammar: Studies in Honor of Henk van Riemsdijk* (pp. 69–76). Berlin and New York: Mouton de Gruyter.
- Chesi, C., Lebani, G., & Pallottino, M. (2011). A Bilingual Treebank (ITALLIS) suitable for Machine Translation: what Cartography and Minimalism teach us. In *STiL-Studies in Linguistics* (Vol. 2, p. 165-185).
- Chi, Z. (1999). Statistical properties of probabilistic context-free grammars. *Computational Linguistics*, 25(1), 131–160.
- Chomsky, N. (1965). *Aspects of the theory of syntax* (No. 11). MIT press.
- Chomsky, N. (1977). On wh-movement. In P. Culicover, T. Wasow, & A. Akmajian (Eds.), *Formal syntax* (pp. 71–132). New York: Academic.
- Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. Greenwood Publishing Group.
- Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.
- Cinque, G. (1982). On the theory of relative clauses and markedness. *The linguistic review*, 1(3), 247–294.

- Clemens, L. E., Coon, J., Pedro, P. M., Morgan, A., Polinsky, M., Tandet, G., & Wagers, M. (in press). Ergativity and the complexity of extraction: A view from mayan. *Natural Language and Linguistic Theory*.
- Clifton, C., Juhasz, B., Ashby, J., Traxler, M. J., Mohamed, M. T., Williams, R. S., ... Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language*, 49, 317-334.
- Clifton Jr, C., & Frazier, L. (1989). Comprehending sentences with long-distance dependencies. In *Linguistic structure in language processing* (pp. 273–317). Springer.
- Comrie, B., & Kuteva, T. (2005). Relativization strategies. *The World Atlas of Language Structures*, 494–501.
- Contemori, C., & Belletti, A. (2011). Disentangling the mastery of object relatives in children and adults. evidence from italian. In *STiL-Studies in Linguistics* (Vol. 4, pp. 48–81).
- Croft, W. (2003). *Typology and universals*. Cambridge University Press.
- Demberg, V., & Keller, F. (2009). A Computational Model of Prediction in Human Parsing: Unifying Locality and Surprisal Effects. In *Proceedings of the 29th meeting of the Cognitive Science Society*. Amsterdam, Netherlands.
- den Dikken, M., & Singhapreecha, P. (2004). Complex noun phrases and linkers. *Syntax*, 7(1), 1–54.
- De Vincenzi, M. (1991). *Syntactic parsing strategies in italian: The minimal chain principle* (Vol. 12). Springer.
- Di Domenico, A., & Di Matteo, R. (2009). Processing Italian relative clauses: working memory span and word order effects on RTs. *Journal of General Psychology*(136), 387-406.

- Dixon, R. M. (1979). Ergativity. *Language*, 59–138.
- Dryer, M. S. (2005). Order of relative clause and noun. In M. S. Dryer & M. Haspelmath (Eds.), *Feature 90a*.
- Fox, B. A., & Thompson, S. A. (1990). A discourse explanation of the grammar of relative clauses in English conversation. *Language*, 297–316.
- Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies*. Unpublished doctoral dissertation, University of Connecticut.
- Frazier, L., & Clifton, C. (1996). *Construal*. Cambridge, MA: MIT Press.
- Frazier, L., & Flores D'Arcais, G. B. (1989). Filler driven parsing: A study of gap filling in Dutch. *Journal of memory and language*, 28(3), 331–344.
- Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14(2), 178–210.
- Gennari, S. P., & MacDonald, M. C. (2008). Semantic indeterminacy in object relative clauses. *Journal of memory and language*, 58(2), 161–187.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68, 1–76.
- Gibson, E. (2000). Dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, Language, brain: Papers from the First Mind Articulation Project Symposium*. Cambridge, MA: MIT Press.
- Gibson, E., & Wu, H.-H. I. (2013). Processing Chinese relative clauses in context. *Language and Cognitive Processes*, 28(1-2), 125-155.
- Givón, T. (1983). *Topic continuity in discourse: A quantitative cross-language study* (Vol. 3). John Benjamins Publishing.
- Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic

- input for sentential complexity. *Cognitive Science*, 29(2), 261–290.
- Guillaumin, M. (2005). *Conversions between mildly context sensitive grammars* (Tech. Rep.). Ecole Normale Supérieure and UCLA.
- Hale, J. (2001). A probabilistic early parser as a psycholinguistic model. In *Proceedings of the second meeting of the north american chapter of the association for computational linguistics on language technologies*.
- Hale, J. (2003). The information conveyed by words in sentences. *Journal of Psycholinguistic Research*, 32(2), 101–123.
- Hale, J. (2006). Uncertainty about the rest of the sentence. *Cognitive Science*, 30(4).
- Harkema, H. (2001). *Parsing minimalist grammars*. Unpublished doctoral dissertation, University of California, Los Angeles.
- Hawkins, J. A. (2004). *Efficiency and complexity in grammars*. Oxford University Press.
- Hsiao, F., & Gibson, E. (2003). Processing relative clauses in Chinese. *Cognition*, 90, 3–27.
- Hsiao, Y., Li, J., & MacDonald, M. (2014). Ambiguity affects Mandarin relative clause processing. In *The 27th Annual CUNY Conference on Human Sentence Processing*. Columbus, OH.
- Hsiao, Y., & MacDonald, M. C. (2013). Experience and Generalization in a Connectionist Model of Mandarin Chinese Relative Clause Processing. *Frontiers in Psychology*, 4(767).
- Huang, C.-T. J., Li, Y.-H. A., & Li, Y. (2009). *The syntax of chinese*. Cambridge University Press.
- Hunter, T. (2013). User documentation for ccpc [Computer software manual]. Cornell University.

- Hunter, T., & Dyer, C. (2013). Distributions on minimalist grammar derivations. In *The 13th meeting on the mathematics of language*.
- Jäger, L., Chen, Z., Li, Q., Lin, C.-J. C., & Vasishth, S. (Submitted). The subject-relative advantage in Chinese: Evidence for expectation-based processing.
- Johnson, M. (1998). PCFG models of linguistic tree representations. *Computational Linguistics*, 24, 613–632.
- Joshi, A. K. (1985). Tree adjoining grammars: How much context-sensitivity is required to provide reasonable structural descriptions? In D. Dowty, L. Karttunen, & A. M. Zwicky (Eds.), *Natural language parsing: Psychological, computational and theoretical perspectives* (pp. 206–250). Cambridge University Press.
- Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognition*, 20, 137–194.
- Jurafsky, D., & Martin, J. H. (2008). *Speech and language processing: An introduction to natural language processing, computational linguistics and speech recognition* (2nd ed.). McGraw Hill.
- Just, M. A., Carpenter, P., Keller, T., Eddy, W., & Thulborn, K. (1996). Brain Activation Modulated by Sentence Comprehension. *Science*, 274(5284), 114.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological review*, 99, 122–149.
- Kathol, A. (2000). *Linear syntax*. Oxford University Press.
- Kayne, R. S. (1994). *The antisymmetry of syntax*. Cambridge, MA: MIT Press.
- Keenan, E. L., & Comrie, B. (1977). Noun phrase accessibility and universal grammar. *Linguistic Inquiry*, 8(1), 63–99.

- Keenan, E. L., & Hawkins, S. (1987). The psychological validity of the Accessibility Hierarchy. In E. L. Keenan (Ed.), *Universal grammar: 15 essays* (pp. 60–85). London: Croom Helm.
- King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, 30(5), 580–602.
- King, J., & Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPS to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, 7(3), 376–395.
- Kwon, N., Kluender, R., Kutas, M., & Polinsky, M. (2013). Subject/object processing asymmetries in Korean relative clauses: Evidence from ERP data. *Language*, 89, 537–585.
- Kwon, N., Lee, Y., Gordon, P. C., Kluender, R., & Polinsky, M. (2010). Cognitive and linguistic factors affecting subject/object asymmetry: An eye-tracking study of pre-nominal relative clauses in Korean. *Language*, 86(3), 546–582.
- Kwon, N., Polinsky, M., & Kluender, R. (2006). Subject Preference in Korean. In D. Baumer, D. Montero, & M. Scanlon (Eds.), *Proceedings of the 25th West Coast Conference on Formal Linguistics (WCCFL-25)* (p. 1-14). Somerville, MA: Cascadilla Proceedings Project.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106, 1126-1177.
- Levy, R., & Andrew, G. (2006). Tregex and Tsurgeon: tools for querying and manipulating tree data structures. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 06)*.
- Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29, 1-45.
- Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of

- working memory in sentence comprehension. *Trends in cognitive sciences*, 10(10), 447–454.
- Li, P., Bates, E., & MacWhinney, B. (1993). Processing a language without inflections: A reaction time study of sentence interpretation in Chinese. *Journal of Memory and Language*, 32, 169-192.
- Lin, C.-J. C. (2008). The processing foundation of head-final relative clauses. *Language and Linguistics*, 9.4, 813-839.
- Lin, C.-J. C. (2014). Effect of thematic order on the comprehension of Chinese relative clauses. *Lingua*, 140, 180-206.
- Lin, C.-J. C., & Bever, T. G. (2006). Subject preference in the processing of relative clauses in Chinese. In *Proceedings of the 25<sup>th</sup> WCCFL* (p. 254-260).
- Lin, C.-J. C., & Bever, T. G. (2007). Processing doubly-embedded head-final relative clauses. In *Interdisciplinary Approaches to Relative Clauses*. Cambridge, UK.
- Lin, C.-J. C., & Bever, T. G. (2011). Garden path and the comprehension of head-final relative clauses. In H. Yamashita, Y. Hirose, & J. L. Packard (Eds.), *Processing and Producing Head-final Structures*, Studies in Theoretical Psycholinguistics (p. 277-297). Springer.
- Lin, Y. B., & Garnsey, S. M. (2011). Animacy and the resolution of temporary ambiguity in relative clause comprehension in Mandarin. In H. Yamashita, Y. Hirose, & J. L. Packard (Eds.), *Processing and Producing Head-final Structures*, Studies in Theoretical Psycholinguistics (pp. 241–276). Springer.
- Lowder, M. W., & Gordon, P. C. (2012). The pistol that injured the cowboy: Difficulty with inanimate subject–verb integration is reduced by structural separation. *Journal of Memory and Language*, 66(4), 819–832.
- MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working mem-

- ory: A reply to Just and Carpenter and Waters and Caplan. *Psychological Review*, 109(1), 35–54.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological review*, 101(4), 676.
- MacWhinney, B. (1977). Starting points. *Language*, 53, 152-168.
- MacWhinney, B. (1982). Basic syntactic processes. In S. Kuczaj (Ed.), *Language Acquisition* (Vol. 1). Hillsdale, NJ: Lawrence Erlbaum.
- MacWhinney, B., & Bates, E. (Eds.). (1989). *The crosslinguistic study of sentence processing*. New York: Cambridge University Press.
- MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German and Italian. *Journal of Verbal Learning and Verbal Behavior*, 23, 127–150.
- Maiden, M., & Robustelli, C. (2007). *A reference grammar of modern italian* (Second ed.). McGraw Hill.
- Mak, W., Vonk, W., & Schriefers, H. (2002). The Influence of Animacy on Relative Clause Processing. *Journal of Memory and Language*, 47(1), 50–68.
- Mak, W., Vonk, W., & Schriefers, H. (2006). Animacy in processing relative clauses: the hikers that rocks crush. *Journal of Memory and Language*, 54, 466-490.
- Marcus, M. P., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: the Penn treebank. *Computational Linguistics*, 19(2), 313–330.
- Mazuka, R., & Itoh, K. (1995). Can japanese speakers be led down the garden path. In R. Mazuka & N. Nagai (Eds.), (pp. 295–329). Mahwah, NJ: Lawrence Erlbaum.
- Michaelis, J. (2001). *On formal properties of Minimalist Grammars*. Unpublished

- doctoral dissertation, University of Potsdam, Potsdam, Germany.
- Miller, G. A. (1962). Some psychological studies of grammar. *American Psychologist*, *17*, 748-762.
- Mitchell, D. C., Cuetos, F., Corley, M. M. B., & Brysbaert, M. (1995). Exposure-based models of human parsing: Evidence for the use of coarse-grained (nonlexical) statistical records. *Journal of Psycholinguistic Research*, *24*, 469-488.
- Miyamoto, E., & Nakamura, M. (2003). Subject/object asymmetries in the processing of relative clauses in Japanese. In *The 22nd West Coast Conference on Formal Linguistics (WCCFL-22)* (pp. 342-355). University of California, San Diego.
- Narayanan, S., & Jurafsky, D. (2002). A Bayesian Model Predicts Human Parse Preference and Reading Time in Sentence Processing. In T. G. Dietterich, S. Becker, & Z. Ghahramani (Eds.), *Advances in Neural Information Processing Systems* (Vol. 14, pp. 59-65). Cambridge, MA: MIT Press.
- Nederhof, M.-J., & Satta, G. (2008). Computation of distances for regular and context-free probabilistic languages. *Theoretical Computer Science*, *395*, 235-254.
- O'Grady, W. (1997). *Syntactic Development*. The University of Chicago Press.
- Ou, T.-S. (2007). *Suo* relative clauses in Mandarin Chinese. *Language and Linguistics*, *8*(4), 913-937.
- Pu, M.-M. (2007). The distribution of relative clauses in Chinese discourse. *Discourse Processes*, *43*(1), 25-53.
- Qiao, X., Shen, L., & Forster, K. (2012). Relative clause processing in Mandarin: Evidence from the maze task. *Language and Cognitive Processes*, *27*(4), 611-630.

- Reali, F., & Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language*, 57(1), 1–23.
- Roland, D., Dick, F., & Elman, J. L. (2007). Frequency of basic english grammatical structures: A corpus analysis. *Journal of Memory and Language*, 57, 348–379.
- Roland, D., Mauner, G., O'Meara, C., & Yun, H. (2012). Discourse expectations and relative clause processing. *Journal of Memory and Language*, 66(3), 479–508.
- Seki, H., Matsumura, T., Fujii, M., & Kasami, T. (1991). On multiple context-free grammars. *Theoretical Computer Science*, 88(2), 191-229.
- Sheldon, A. (1974). On the role of parallel function in the acquisition of relative clauses in English. *Journal of Verbal Learning and Verbal Behavior*, 13, 272–281.
- Silverstein, M. (1976). Hierarchy of features and ergativity. In R. M. W. Dixon (Ed.), *Grammatical categories in Australian languages* (p. 112-171). Canberra: Australian Institute of Aboriginal Studies.
- Stabler, E. (1997). Derivational minimalism. In C. Retoré (Ed.), *Logical aspects of computational linguistics*. Springer-Verlag.
- Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition*, 116(1), 71–86.
- Stromswold, K., Caplan, D., Alpert, N., & Rauch, S. (1996). Localization of syntactic comprehension by positron emission tomography. *Brain and Language*, 52, 452–473.
- Su, Y., Lee, S., & Chung, Y. (2007). Asyntactic thematic role assignment by Mandarin aphasics: a test of the Trace-Deletion Hypothesis and the Double

- Dependency Hypothesis. *Brain and Language*, 101(1), 1–18.
- Traxler, M., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, 47, 69-90.
- Traxler, M., Williams, R., Blozis, S., & Morris, R. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language*, 53, 204-224.
- Trueswell, J., & Tanenhaus, M. (1994). Toward a lexical framework of constraint-based syntactic ambiguity resolution. *Perspectives on sentence processing*, 155–179.
- Ueno, M., & Garnsey, S. M. (2008). An ERP study of the processing of subject and object relative clauses in Japanese. *Language and Cognitive Processes*, 23(5), 646–688.
- Utzeri, I. (2007). The production and acquisition of subject and object relative clauses in Italian: a comparative experimental study. *Nanzan Linguistics*, 2.
- Vasishth, S., Chen, Z., Li, Q., & Guo, G.-L. (2013). Processing Chinese relative clauses: Evidence for the subject-relative advantage. *PLoS One*, 8(10)(e77006).
- Velupillai, V. (2012). *An introduction to linguistic typology*. John Benjamins Publishing.
- Vergnaud, J.-R. (1985). *Dépendances et niveaux de représentation en syntaxe* (Vol. 13). John Benjamins Publishing.
- Wanner, E., & Maratsos, M. (1978). An ATN approach in comprehension. In *Linguistic theory and psychological reality* (p. 119-161). Cambridge, MA: MIT Press.

- Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition*, 85(1), 79–112.
- Wilson, K., & Carroll, J. B. (1954). Applications of entropy measures to problems of sequential structure. In C. E. Osgood & T. A. Sebeok (Eds.), *Psycholinguistics: a survey of theory and research* (pp. 103–110). Indiana University Press.
- Wu, F. (2009). *Factors Affecting Relative Clause Processing in Mandarin: Corpus and Behavioral Evidence*. Unpublished doctoral dissertation, University of Southern California.
- Wu, F., Kaiser, E., & Andersen, E. (2011). Subject Preference, Head Animacy and Lexical Cues: A Corpus Study of Relative Clauses in Chinese. In H. Yamashita, Y. Hirose, & J. L. Packard (Eds.), *Processing and producing head-final structures* (pp. 173–193). Springer.
- Wu, F., Kaiser, E., & Andersen, E. (2012). Animacy effects in Chinese relative clause processing. *Language and Cognitive Processes*, 27(10), 1489–1524.
- Xu, Y. (2014). Evidence of the Accessibility Hierarchy in Relative Clauses in Chinese as a Second Language. *Language and Linguistics*, 15(3), 435–464.
- Xue, N., Xia, F., Chiou, F.-D., & Palmer, M. (2005). The Penn Chinese TreeBank: Phrase structure annotation of a large corpus. *Natural Language Engineering*, 11, 207–238.
- Yun, J., Chen, Z., Hunter, T., Whitman, J., & Hale, J. (2015). Uncertainty in processing relative clauses across East Asian languages. *Journal of East Asian Linguistics*.
- Yun, J., Whitman, J., & Hale, J. T. (2010). Subject-object asymmetries in Korean sentence comprehension. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32<sup>nd</sup> Annual Meeting of the Cognitive Science Society*.

Zaenen, A., Carletta, J., Garretson, G., Bresnan, J., Koontz-Garboden, A., Nikitina, T., ... Wasow, T. (2004). Animacy encoding in english: why and how. In *Proceedings of the 2004 acl workshop on discourse annotation* (pp. 118–125).