

COMPETITION IN NATURAL LANGUAGE MEANING: THE CASE OF ADJECTIVE
CONSTRUCTIONS IN MANDARIN CHINESE AND BEYOND

By

Yan Cong

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Linguistics – Doctor of Philosophy

2021

ABSTRACT

COMPETITION IN NATURAL LANGUAGE MEANING: THE CASE OF ADJECTIVE CONSTRUCTIONS IN MANDARIN CHINESE AND BEYOND

By

Yan Cong

Utterances compete with each other. Rational speakers choose an utterance that is true, informative, and relevant, and listeners reason about that choice. As a consequence, pragmatic listeners make inferences about other possible utterances (so-called *alternatives*). In the well-studied case of Scalar Implicature (henceforth SI), pragmatic enrichment yields the inference that more informative alternatives are false, or at least that the speaker doesn't believe them. A central question in the SI literature is what counts as an alternative of a given utterance, due to what is known as the *symmetry problem*: without constraints on alternatives, every potential alternative ψ has a symmetric partner (roughly, *not ψ*), whose existence preempts any SI about ψ . Consequently, theories of formal alternatives have been proposed (Katzir, 2007). However, relatively few studies concern Non-Scalar Implicature (henceforth NSI) (Rett, 2015). This dissertation argues that the interpretation of adjectival constructions in Mandarin Chinese involves non-scalar competition, that a kind of symmetry problem arises even for NSIs, and that standard (e.g., Katzirian) theories of formal alternatives do not solve the problem. I propose to associate gradient costs with structural alternatives to break symmetry.

In this dissertation, I provide a detailed theoretical, computational, and experimental investigation into the role of cost in implicature. The key finding is that cost influences pragmatic reasoning, in a way that compared to the costly alternative, rational cooperative language users are inclined to reason about the less-costly one. The dissertation involves a variety of novel contributions. Methodologically, it presents a detailed empirical investigation of Mandarin adjectives through a truth value judgment survey with explicit contexts. Theoretically, it provides formal pragmatic analysis disambiguation through competition, implemented computationally within the Rational Speech Act framework (Frank & Goodman, 2012a). And experimentally, it tests predictions of the

pragmatic model using the artificial language learning methodology (Adger et al., 2019; Buccola et al., 2018; Culbertson, 2012; Culbertson & Schuler, 2019; Motamedi et al., 2019). Overall, the dissertation sheds new light on the semantics and pragmatics of degree constructions in Mandarin and beyond, and on the nature of pragmatic competition beyond SI.

Copyright by
YAN CONG
2021

To my grandpa.

ACKNOWLEDGMENTS

One of my favorite things to do while reading a dissertation is to read the acknowledgments. I think it's special because you can visualize all of the stuff that the writer and their mentors and peers are going through behind the scenes. Now it's my turn to express my gratitude to my mentors and peers, with whom I have had a long conversation about linguistics. This dissertation would not have been possible without the help, support and guidance of them.

I would like to express my deepest appreciation to my advisor Brian Buccola, who knows everything about everything, understands everything. Brian provided me with encouragement and patience throughout the duration of this thesis. I don't think I can ever find enough ability or courage to write this thesis if not for him. I'm deeply indebted to Alan Munn. Alan advised my very first linguistics project. His syntax class is my all time favorite. To Cristina Schmitt, I aspire to be like her in her wisdom and kindness. To Karthik Durvasula for his constructive suggestions and emotional support. I'm extremely grateful to have them in my committee. I would also like to extend my deepest gratitude to Marcin Morzycki. I would never have known or pursued semantics if not for Marcin.

Devin McAuley and Alan Beretta helped me better understand linguistics from the perspective of psychology, neuroscience, and cognition. I'm incredibly grateful for their nurturing. To Deo Ngonyani, I owe invaluable inspirations on studying Non-Indo-European languages. Special thanks to Suzanne Wagner and Yen-Hwei Lin. Without them, I would not have the grants to live on in the final years of my graduate education. Thanks should also go to Alan Ke for listening to my research and giving me feedback whenever possible. To Yingfei Chen, Xiaoshi Li, and Wenying Zhou, who supported my first year assistantship and gave me opportunities to run the class. I'm extremely grateful to Parisa Kordjamshidi and Jiayu Zhou for kindly letting me join their group meetings and their classes.

Bridget Copley generously volunteered to mentor me, and ensured my survival and well-being when I was in my final year. Thank you Bridget! Thanks to Phillip Wolff, who spent a lot of

time exchanging thoughts with me about the meaning of meaning. I'm deeply indebted to Allyson Ettinger, who inspired me that linguists' insights can make a difference in many many ways. Thank you Hadas Kotek, Michael Terry, Coppe van Urk, and Linmin Zhang, for taking the time to speak with me. I'd like to also thank John Wakefield and Chu-Ren Huang for their guidance and nurturing.

Many thanks to Ryan Hasselbach, Tanner Schudlich, Benjamin Lampe, Jennifer Nelson, and Carly Kabel. I deeply value their help and hard work behind the scenes. Special thanks to all the Prolific participants and all the volunteers for completing various versions of the experiments, which tremendously helped improve the design. To Ying Cong, Tun Hao, and Jia Ming Sun who helped me find those volunteers, without their unwavering support, I would have no data to analyze. I also had great pleasure of working with Lisa Lipin, Caroline Zackerman, Toni Smith, and Lalchand Pandia. Special thanks to Ai Taniguchi for patiently answering my linguistics and non-linguistics questions during my very first year in the states. Thank you Xiao Yang for helping me out with job info in my final year.

Finally, I'd like to thank Lian-Hee Wee for helping me get prepared for grad school, intellectually, mentally, and financially. I will never forget the noodles that Lian-Hee and Winnie cooked for me in their house after spending hours and hours explaining OT, c-command, and Guqin. Thank you to my parents, my sister, and LZ for their company, which sweetened every bitter moment of my life.

TABLE OF CONTENTS

LIST OF TABLES	xi
LIST OF FIGURES	xiii
CHAPTER 1 INTRODUCTION AND PREVIEW OF THE PROPOSAL	1
1.1 General introduction	1
1.2 Core data and puzzles	4
1.2.1 Competition in natural language meaning	4
1.2.2 Canonical analysis of degree expressions: summary	5
1.2.3 Degree expressions in Chinese: core data	6
1.3 Previous approach to the puzzles	8
1.4 Proposed approach to the puzzles	9
1.5 The structure of the dissertation	11
CHAPTER 2 BACKGROUND ON DEGREE SEMANTICS	13
2.1 Degree semantics and cross-linguistic variation	13
2.1.1 The morphosyntactic realization of the comparative	13
2.1.2 Degree semantics parameter	18
2.2 Degree expressions in Chinese	21
2.2.1 General descriptions about Chinese gradable expressions	21
2.2.2 Bare predication <i>has to be</i> COMP	23
2.2.3 Bare predication <i>can be</i> POS	27
2.3 Chapter summary	35
CHAPTER 3 A MORE DETAILED LANDSCAPE OF THE DATA	36
3.1 Methodology for detecting ambiguity	36
3.2 Theoretical assumptions	37
3.2.1 Basic cases	37
3.2.2 A more complex picture - Introspective data	38
3.3 Empirical assumptions - Quantitative data	43
3.3.1 Method	44
3.3.2 Results	48
3.3.3 Discussion	51
3.4 Chapter summary	52
CHAPTER 4 COMPETITION: PREVIOUS STUDIES AND IMPLEMENTATION	54
4.1 Competition of utterances	54
4.1.1 Gricean pragmatics and (Scalar) Implicature	54
4.1.2 Non-scalar Implicature	57
4.2 Competition of interpretations	63
4.2.1 True on the strongest meaning	63
4.2.2 True on all meanings	65

4.3	More about the intention based view	66
4.4	Chapter summary	70
CHAPTER 5 A COMPETITION-BASED DISAMBIGUATION MODEL		71
5.1	Motivation for a competition-based proposal for NSIs	71
5.1.1	The symmetry problem in SIs	72
5.1.2	Structural alternatives in SIs	73
5.2	Basic proposal	75
5.2.1	The disambiguation model in a nutshell	75
5.2.2	Alternatives in non-scalar implicature	77
5.2.3	The symmetry problem in non-scalar implicatures	77
5.3	Refined proposal	79
5.3.1	Assumptions	80
5.3.2	Derivations	81
5.4	Chapter summary	83
CHAPTER 6 UNDERSTANDING <i>COST</i>		84
6.1	“Costly” with respect to semantic interpretation	84
6.1.1	Structural competition among alternatives	85
6.1.2	Interpreting alternatives <i>geng</i> and <i>hen</i>	88
6.1.3	Spelling out the semantics	92
6.2	“Costly” with respect to primitiveness	99
6.2.1	Cross-linguistic lexicalization	99
6.2.2	Conceptual alternatives	100
6.3	“Costly” with respect to frequency	101
6.4	Chapter summary	103
CHAPTER 7 SIMULATING THE PROPOSAL		105
7.1	Cost in RSA models	105
7.2	Stepwise implementation – vanilla RSA	106
7.3	RSA models as disambiguation simulation tools	114
7.4	Chapter summary	119
CHAPTER 8 EVIDENCE FROM ARTIFICIAL LANGUAGE LEARNING		122
8.1	Competition principle and consequence	122
8.2	Experiment: the speaker perspective	124
8.2.1	Task summary and hypothesis	126
8.2.2	Method	129
8.2.3	Results	138
8.2.4	Other findings	147
8.2.5	Discussion	150
8.2.5.1	Limitations and explanations	150
8.2.5.2	Implications	151
8.3	Chapter summary	153

CHAPTER 9	CONCLUSION AND FUTURE WORK	154
9.1	Circling back to the question about competition	154
9.2	Circle back to the question about degree	154
9.3	Circle back to the question about language universals	155
9.3.1	Expand the typology table	155
9.3.2	Universals and tendencies	156
9.3.3	Candidate languages	157
9.4	Chapter summary	159
BIBLIOGRAPHY		160

LIST OF TABLES

Table 2.1: Y.Zhang & Grano’s (2019) observation about how to interpret <i>hen</i> and bare adjectivals	27
Table 2.2: Implication of Zhang’s analysis (2019) (inspired by Liu (2010b) and Liu (2018))	29
Table 3.1: A more detailed landscape of the data: summary	51
Table 6.1: Zhang’s (2019) analysis	89
Table 6.2: The standard and differential involved in comparison (Only the marker of discourse salience and numerals are pronounced) summarized in Zhang (2019) and Zhang & Ling (2020)	91
Table 6.3: Morphosyntactic relationship between positive and comparative forms cross- linguistically (adapted from Grano (2012) and Grano & Davis (2018))	99
Table 6.4: Word frequency comparison (Source: Da, Jun 2005)	102
Table 6.5: Global (<i>blog, literature, news, tech, weibo</i>) word frequency from BLCU corpus .	102
Table 6.6: Word frequency and count from Chinese Lexical Database (CLD)	102
Table 7.1: How cost influences pragmatic reasoning under different RSA models	118
Table 8.1: Groups, Conditions and Parameters	128
Table 8.2: Literal semantics	128
Table 8.3: Group comparisons, phase comparisons, and predictions	128
Table 8.4: Calculation of trials and blocks	128
Table 8.5: Trials types in the Gaming part	132
Table 8.6: Nonce phrases used in the real experiment stimuli	132
Table 8.7: Descriptives	145
Table 8.8: Comparison (cross-group; cross-phase) Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1	146

Table 8.9: Tukey multiple comparisons of means 95% family-wise confidence level	147
Table 8.10: Gerken 2005:2 (columns: B; rows: A)	152

LIST OF FIGURES

Figure 3.1: <i>gui</i> (“expensive”) _POS	47
Figure 3.2: <i>gui</i> (“expensive”) _COMP	47
Figure 3.3: Truth value judgment survey on plain form	48
Figure 3.4: Truth value judgment survey on modified form	50
Figure 7.1: scenario r_1 : COMP true POS false	108
Figure 7.2: scenario r_2 : COMP true POS true	109
Figure 7.3: $C(\text{hen gao})=-1$; $C(\text{gao})=0$; $\alpha(\text{LI})=1$	112
Figure 7.4: $C(\text{hen gao})=-1$; $C(\text{bi gao})=-2$; $C(\text{gao})=0$	112
Figure 7.5: $C(\text{hen gao})=-1$; $C(\text{gao})=0$; $\alpha(\text{LI})=1$	113
Figure 7.6: $C(\text{hen gao})=-1$; $C(\text{bi gao})=-2$; $C(\text{gao})=0$	114
Figure 7.7: Franke and Bergen (2020:e81): Schematic representation of main conceptual differences in the speaker production part of the four models	116
Figure 8.1: Predicted reasoning process – rational speakers	125
Figure 8.2: Learning trial: ambiguous phrase	132
Figure 8.3: Learning trial: costly phrase	133
Figure 8.4: Learning trial: less-costly phrase	133
Figure 8.5: Testing trial: ctrl cases	134
Figure 8.6: Testing trial: critical cases	135
Figure 8.7: Experimental group: sequence of phases	136
Figure 8.8: Control group: sequence of phases	137

Figure 8.9: Zoom in: A flowchart of a learning phase in the artificial language learning experiment (<i>the sequence is presented for illustration purpose; trials are randomly interspersed in the actual experiment</i>)	138
Figure 8.10: Zoom in: A flowchart of a gaming phase in the artificial language learning experiment (<i>the sequence is presented for illustration purpose; trials are randomly interspersed in the actual experiment</i>)	138
Figure 8.11: Overall performance - Experimental group	139
Figure 8.12: Within-group comparison: ambiguous responses phase1	140
Figure 8.13: Within-group comparison: involving non-ambiguous responses phase1	142
Figure 8.14: Between-group comparison	143
Figure 8.15: Error bar represents standard error of the mean: critical cases condition-based (costly vs. less-costly intended) <i>cost</i> effect difference between group	144
Figure 8.16: Error bar represents standard error of the mean: critical cases condition-based (costly vs. less-costly intended) <i>cost</i> effect difference between phase	145
Figure 8.17: Within-group: ambiguous response (ctrl)	148
Figure 8.18: Within-group: non-ambiguous response (ctrl)	148
Figure 8.19: Error bar represents standard error of the <i>mean</i> : <i>cost</i> effect on group	149
Figure 8.20: Error bar represents standard error of the mean: <i>cost</i> effect on phase	149

CHAPTER 1

INTRODUCTION AND PREVIEW OF THE PROPOSAL

To set the stage, this chapter provides an overview for the dissertation and a preview of the core proposal. Specifically, the chapter starts with empirical data observation, puzzles are raised, followed by the proposal in a nutshell. A general summary of previous studies on the phenomenon is given to further narrow down the research questions, and to motivate the current proposal drew on. Potential challenges and possible directions are discussed briefly. At the end of the chapter, the structure of the dissertation is elaborated on.

1.1 General introduction

What we do not say matters. Outside the utterance itself, there are various factors determining which interpretation stands out as the most likely one in terms of the intended meaning when a person says something that can have multiple interpretations. Contextual factors involve utterances that the speaker could have chosen but did not, common sense knowledge, salient objects in the environment, the amount of effort used to say the utterance, etc. Systematically predicting these contextual factors is a major unsolved puzzle in pragmatic studies when it comes to natural language understanding. This dissertation attempts to explore the competing meanings of back-and-forth iterations from both the speaker's and listener's perspectives. This exploration starts with a focused case study about degree constructions in Mandarin Chinese.

Specifically, Chapter 2 sets the stage for the cases studies about degree expressions. Background literature on degree semantics is reviewed, with a focus on cross-linguistic variation (Hohaus, 2015, 2018; Hohaus & Bochank, 2020; Deal & Hohaus, 2019), and what the key parameters look like in Chinese (Xiang, 2005; Stojanovic, 2007; Grano, 2012; Zhang & Grano, 2019; Li, 2017; Liu, 2018; Zhang, 2019a). It turns out that there are a lot of debate in terms of the basic empirical picture: whether or not gradable bare adjectivals are ambiguous (in Chinese), how to interpret them, and under what contexts a particular interpretation stands out. To answer these questions, Chapter 3

provides a more detailed landscape of the data through truth value judgment surveys. These surveys investigate the extent to which Chinese plain-form and modified-form adjectives are ambiguous, relative to linguistic contexts, for example, upward and downward entailing environments. The truth-value judgment surveys show that bare adjectivals are indeed ambiguous, and the surfaced reading is dependent on linguistic context. This is an important empirical observation, because it's very reminiscent of the competition-based pragmatic account that has been widely proposed for (scalar) implicatures.

Both the theoretical intuitions and the empirical observations point toward competition-based pragmatic account: not only that the enriched meaning of the utterance is heavily related to contexts, but also that this utterance is competing with its alternatives, which are hallmarks of implicatures studies. Against this setting, the goal of Chapter 4 is to capture the Chinese adjectives phenomena using Gricean pragmatics' standard recipe. After carefully reviewing previous accounts of competition (Grice, 1975, 1989; Geurts, 2010; Rett, 2008a, 2014a; Dalrymple et al., 1994, 1998; Spector, 2017), I come to the conclusion that they do not seem to be able to explain the whole picture without making wrong predictions. This is mainly due to the fact that for the Chinese degree expressions in the current case studies, the utterance and its alternatives are not *scalemates*. Different from *some* and its alternative *all*, using a bare adjectival does not give rise to the negative inference that the speaker *believes* that its alternatives are *false*. All the inference we can draw from hearing a bare adjectival is that the speaker's *intended* meaning is the one corresponding to its *costly* alternative. I argue that this is a competition-based disambiguation inference. It has a lot to do with *intention* and *cost*.

This competition-based disambiguation model is extensively discussed in Chapter 5. I highlight the nature of the degree expressions puzzle: it's concerned with *non-scalar* implicature. Since it's relatively less studied, I discuss the distinct features of alternatives in non-scalar implicatures, and raised the question of the symmetry problem in non-scalar implicature. The core proposal is given with two attempts: basic proposal and refined proposal, in which I recast the degree expressions puzzle using Gricean pragmatics first, followed by probabilistic pragmatics terms so

that the analytical intuitions can get better quantified and visualized. Several assumptions are made in Chapter 5, among which the notion of *cost* is so critical that it deserves a separate chapter. This is due to the consideration that *cost* defines alternatives in (non-scalar) implicature, and more importantly, it breaks the symmetry of competitions in (non-scalar) implicature.

Chapter 6 therefore extends Chapter 5 and takes a close look at the concept of *cost*, from the perspective of semantic interpretation, primitiveness, and frequency. I present evidence showing that the *costly* alternative is costly because it has higher order functions involved in its composition, not only adding extra layers of complexity (Moracchini, 2018), but also making the meaning less primitive and hard to lexicalize across languages (Buccola et al., 2021), relative to the less-costly alternative. Even for languages that appear to lexicalize this costly message, the expression does not seem to correspond exactly to the message. Further, frequency is discussed through a mini-corpus study, suggesting that the costly alternative is costly since it has low frequency, relative to the less-costly alternative.

Earlier in Chapter 5, I briefly recast the calculation of non-scalar implicatures using probability pragmatics, with details of the derivations left unexplained. Chapter 7 therefore extends Chapter 5, and spells out the derivation under the Rational Speech Act framework (RSA) (Frank & Goodman, 2012a; Franke & Bergen, 2020). This Chapter also serves as a proof of concept. Bayesian probability simulates the reasoning process proposed by Chapter 5, and visualizes the extent to which *cost* influences the competition-based disambiguation model, which speaks to Chapter 6. Similar to Chapter 7, Chapter 8 is a proof of concept chapter. Under the artificial language learning paradigm (Buccola et al., 2018), I designed a behavioral experiment, in which participants are instructed to learn nonce phrases and to play a game with aliens on choosing phrase, relative to a given message. The goal of the experiment is to measure the extent to which cost influences reasoning in an artificial language learning setting. If the proposed competition principle is still observed in this setting, then there is potential that the competition principle can go beyond language. It provides a way to better understand human cognition.

Finally, Chapter 9 concludes the dissertation. Limitations are acknowledged, contributions are

summarized, and future work is briefly laid out. In particular, I circle back to the questions about competition based pragmatics, about degree semantics, and about language universals in general.

1.2 Core data and puzzles

The core data and puzzles concern adjectival constructions. They are novel and meaningful puzzles under the theoretical framework about competition in natural language understanding. Plain form adjectivals in Chinese appear to have multiple interpretations, corresponding to unambiguous specific degree expressions (i.e., alternatives). Depending on the context and on the available alternatives, the reading that stands out varies. The key question is to figure out how they vary and why they vary in the way they are. And this is very reminiscent of pragmatic accounts on competition of utterances and competition of interpretations.

1.2.1 Competition in natural language meaning

This dissertation concerns pragmatic enrichment of (degree) interpretation. There are two classes of pragmatic enrichment approaches. One is about competition between utterances, which is called implicature in the literature (Grice, 1975; Sauerland, 2004). If one particular thought can be conveyed by multiple utterances, then these utterances end up competing with each other, which gives rise to implicature. For example, “I will invite Alice or Bob” typically implies I won’t invite both, even though strictly speaking, the sentence is compatible with me inviting both. This inference (implicature) arises because *or* competes with *and*: if I were going to invite both, I would have said “I will invite Alice and Bob”, which is more informative. Thus, *or* is routinely enriched to mean *not both*. In other environments, however, namely downward entailing ones (e.g., “I won’t invite Alice or Bob”), *or* does not get enriched, because the alternative with *and* is now less informative. The other class concerns enrichment of an utterance that is ambiguous between two or more interpretations, so that speakers go for the strongest one (Dalrymple et al., 1994; Spector, 2017). Both of these classes of phenomena have been studied very well. Is there any interaction between the two? It turns out that Chinese degree expressions have the hallmark of both. I attempt to develop a disambiguation

competition model, which hopefully synthesizes the two classes of enrichment analyses.

Canonical analysis of degree constructions pay a lot of attention to their logical forms and composition. The proposed competition-based disambiguation model also provides a novel approach to understand degree expressions pragmatically. A closer scrutiny indicates that contextual factors as well as language users' reasoning play an interesting role in understanding "degree".

1.2.2 Canonical analysis of degree expressions: summary

Degrees are things that undergo comparison, which stand in contrast with entities or events. Gradable adjectives, as illustrated by the lexical entry *tall* of (1a) in (2), denote a relation between a point on a relevant scale (i.e., a degree) and an individual (Cresswell, 1976; von Stechow, 1984; Heim, 1985; Kennedy, 1999).

(1) a. Anna is tall.

b. Anna is taller than Kai.

(2) $\llbracket tall \rrbracket_{\langle d, et \rangle} \stackrel{\text{def}}{=} \lambda d_d. \lambda x_e. \mathbf{height}_{\langle e, d \rangle}(x) \geq d$

Gradable adjectives are canonically analyzed as relations of type $\langle d, et \rangle$. Thus (2) is read in plain English as x is d -tall, namely, x is tall to degree d . Here, **HEIGHT** is a measure function, mapping an individual x to a degree on a relevant scale (i.e., height). As to the matrix and *than*-clauses in (1b), they are typically analyzed as sets of degrees and the morphemes *-er/more* are canonically analyzed as relation between sets of degrees. Now consider the implementation of this analysis to a measurement phrase as in (3).

(3) a. John is two meters tall.

b. $\llbracket tall \rrbracket(\llbracket two\ meters \rrbracket)(\llbracket John \rrbracket) = \mathbf{height}(j) \geq 2m$

In degree-based analyses, both positive (henceforth, POS) and comparative (henceforth, COMP) uses of gradable adjectives are *derived* (Cresswell, 1976; Heim, 1985; Kennedy, 2007).

(4) a. John is tall.

(POSITIVE)

- b. $\llbracket \text{POS} \rrbracket = \lambda g_{\langle d, \langle e, t \rangle \rangle} \lambda x. \exists d [g(d)(x) \wedge d > d_c]$
 c. $\llbracket \text{POS} \rrbracket (\llbracket \text{tall} \rrbracket) (\llbracket \text{John} \rrbracket) = \exists d [\text{height}(j) \geq d \wedge d > d_c]$

(4b) is read as ‘There is some degree d such that John’s height meets or exceeds d and d exceeds a contextually determined threshold d_c .’.

- (5) a. John is taller than Bill. (COMPARATIVE)

- b. $\llbracket \text{COMP} \rrbracket = \lambda g_{\langle d, \langle e, t \rangle \rangle} \lambda x \lambda y. \exists d [g(d)(y) \wedge \neg g(d)(x)]$
 c. $\llbracket \text{COMP} \rrbracket (\llbracket \text{tall} \rrbracket) (\llbracket \text{than Bill} \rrbracket) (\llbracket \text{John} \rrbracket) = \exists d [\text{height}(j) \geq d \wedge \neg [\text{height}(b) \geq d]]$

(5c) is read as ‘There is some degree d such that John’s height meets or exceeds d and Bill’s height does not meet or exceed d .’. Chapter 2 takes a close scrutiny of how to implement this standard degree-based analysis cross-linguistically, the consequences of the implementation, the possible challenges as well as their solutions.

1.2.3 Degree expressions in Chinese: core data

The adjectival construction in Mandarin Chinese such as (6) has received a lot of attention in the literature due to the so-called Mandarin *hen* puzzle (7). (6) appears to have *only* a comparative interpretation ‘tall-er’, while (7) appears to have *only* a positive interpretation ‘tall’, despite the fact that (7) is morphologically more complex than (6). This is at odds with the cross-linguistically stable asymmetry between positive and comparative forms, according to which comparative forms are more complex (or at least not simpler) than positive forms, not the other way around.

- (6) *Anna gao.*
 Anna tall
 Anna is tall-er (than everyone else salient in the context). *comparative*
 NOT: Anna is (above average) tall. *positive*

Sybesma (1999a), Grano (2012)

- (7) *Anna hen gao.*
 Anna HEN tall

Anna is (above average) tall.

NOT: Anna is taller (than everyone else salient in the context).

Sybesma (1999), Grano (2012)

One may wonder, in general, why the comparative reading “tall-*er*” but not the positive “tall” surfaces when speaker uses (6), what’s the role of utterance (7) when listeners reason about (6), what about other similar constructions — how they behave and how they influence the interpretation of *gao* in (6). It turns out that the answers to these puzzles could shed light not only on degree semantics, but also on cognition and cross-linguistic variation. Exactly how these data and puzzles can advance our knowledge about degree semantics and implicatures? And how they will inform us on human cognition? We will need to start with the simple morpheme *gao* ‘tall’.

A truth-value judgment survey shows that *gao* ‘tall’ is ambiguous, and the surfaced meaning is actively interacting with linguistic context and the available alternatives. These are the hallmarks of competition-based pragmatic account. However, a close scrutiny indicates that even if we take the Gricean scalar implicature’s standard recipe, we cannot cover the whole picture. This is due to the fact that using *gao* ‘tall’ does not infer that its alternatives are believed to be false. The lack of such negative inference makes it hard to capture the *gao* ‘tall’ puzzle using the classic scalar implicature account.

I therefore recast the *gao* ‘tall’ puzzle from the perspective of Manner maxim (i.e., “avoid ambiguity”), and reduplicate the classical “symmetry” problem in the non-scalar implicature domain, which has been widely discussed by the scalar implicatures studies. Namely, without constraints on alternatives, anything can become an “alternative”. Consequently, we will not be able to get the desirable inference. Hence, all of the puzzles boil down to one question: what counts as an alternative and how to break the symmetry in non-scalar implicature.

To sum up, this dissertation examines the symmetry problem in non-scalar implicatures through case studies about ambiguity in Chinese degree expressions. These case studies view ambiguity as a violation of Grice Manner Maxim (i.e., “avoid ambiguity”), and thus analyze the inferences from the perspective of (non-scalar) Manner Implicature. The analysis provides a way to address

the following questions. First, how do we reason about ambiguity, and in particular how do costs influence the pragmatic reasoning process? Second, why do cooperative rational language users reason about manner inferences in the way it is? Third, how will the proposed models advance our knowledge about cognition?

1.3 Previous approach to the puzzles

The puzzles can be classified based on two aspects: the empirical aspect — how to interpret *gao* “tall”; the conceptual aspect — how to address the symmetry problem in non-scalar implicature. The two aspects are not mutually independent, rather they are closely related to each other. The explanation of one is better off involving a good explanation of the other. Previous accounts shed lights on both aspects.

Regarding the interpretation of degree expressions in Chinese, previous studies offer syntax-semantics analyses of these phenomena based on various empirical and theoretical assumptions. It’s mostly assumed that there is a strict distinction in where the positive reading is available and where the comparative one is available. And it’s syntax that ultimately determines which reading of the degree construction gets realized in a given sentence. However, among those studies, there is a lot of disagreement about what interpretations are available for degree constructions in (6) and (7). Grano (2012) maintains that (6) only has the comparative interpretation. In contrast, Li (2017) argues that (6) only has the positive interpretation. A corpus based study by Zhang (2019a) reports that (6) is interpreted as comparative, in contrast with (7) which only has the positive reading available. Yet Liu (2010b) proposes that bare AP in (6) can be interpreted in a positive way when interacting with certain syntactic operators.

One of the main contributions I attempt to make is to provide a more detailed landscape of the data that bear on the Chinese *hen* puzzle, and to argue that they point towards a pragmatic, competition-based account.

Regarding the symmetry problem in implicature studies, a lot of attention has been paid to scalar implicature (Geurts, 2010; Sauerland, 2004; Fox, 2007), in which the utterance and its alternative are

on the same scale, and one is logically stronger (i.e., more informative) than the other. Nevertheless, less has been said about non-scalar implicatures (Rett, 2014a).

When trying to capture the *hen/gao* puzzle as a phenomenon that the speech act is flouting the Manner maxim, I find it difficult to fully cover the picture because it concerns language users' intention but not belief, and “complexity” is involved in the derivation of the inferences in a different way than we have often observed in scalar implicatures (Katzir, 2007). The notion of “complexity” is proposed to be a gradient concept, as opposed to a categorical or binary one. In other words, Katzirian operations need a revision to account for the symmetry problem in non-scalar implicature domain. Addition still adds up cost, yet it does not mean that alternatives with “addition” can no longer enter into the competition. I propose that they can take part in the competition, as long as the *relative* cost is not significant.

1.4 Proposed approach to the puzzles

Empirically, the dissertation argues that the bare adjective phrases such as (8) show systematic ambiguity, and the interpretations that surface seem to be sensitive to linguistic environments, which is reminiscent of implicatures. The main theoretical claim I attempt to make is that the structure (8) is construed in two ways: “above average tall” (positive, i.e. POS) and “taller than contextually salient people” (comparative, i.e. COMP).

- (8) *Anna gao.*
Anna tall
Anna is tall (POS)/ taller (COMP)

The dissertation first provides a systematic investigation on both introspective and experimental data, showing (i) that adjectival constructions without *hen* can have both positive and comparative interpretations; (ii) that the pattern of surfaced interpretations seems to track upward versus downward entailingness, hence suggesting a competition-based account.

I propose a disambiguation model to derive speakers' reasoning of why they choose an ambiguous expression when an unambiguous one is available. Another two major components of the proposal

are to extend the proposed disambiguation model across and beyond languages, with an attempt to investigate to what extent this model is applicable in general non-linguistic settings as well as in languages that are unrelated to Mandarin Chinese. This extension will ultimately advance the field through examining the building blocks of (linguistic) universals, and their origins.

In particular, Chapter 2 reviews semantic and syntactic accounts of degree expressions, with research gaps identified. Essentially, previous accounts have argued back and forth about what *gao* “tall” means, and it turns out that there are a lot of discrepancies. I propose that the source of the discrepancies is that careful manipulation of linguistic context is missing. Chapter 3 provides a more detailed landscape of the data, in which I made progress with respect to clean up the debates about how to interpret *gao*. Empirical contribution is made in Chapter 3.

I then move on to explore how to explain the facts from the perspective of pragmatic competition. Chapter 4 contextualizes why I make that move, and how (scalar) implicatures accounts set the stage for the current proposal. Chapter 5 propose a competition-based disambiguation model. Basic attempts and revised attempts are elaborated, together with detailed motivations for each step.

However, there is a critical question about the role of *cost* in pragmatic reasoning, which is worth in-depth discussion. The whole proposal starts with an analytical intuition that *cost* breaks the symmetry in non-scalar implicatures. Chapter 1 therefore spells out the notion of *cost*, from the aspects of logical form, primitiveness, and frequency. This Chapter can be viewed as a zoom-in of Chapter 5.

As proof of concept, Chapter 7 simulates the proposal using a probability pragmatic framework the Rational Speech Act model, and Chapter 8 testifies the proposal using a artificial language learning paradigm, and attempts to quantify the extent to which *cost* influences pragmatic reasoning in general cognition.

As a preview, both the simulation outputs and the experiment results show that the effect of cost is as predicted: cost is indeed influencing pragmatic reasoning in a way that the more costly the utterance, the less probable a rational language user would reason about it in an efficient communication. However, the results are not as pronounced as predicted. The possible explanations

and the limitations are given in Chapter 9, including the lack of transparency of various parameters in the RSA models, and the design deficits of the artificial language learning experiments, for example, the software limitations where the experiment is implemented as well as general online experiments drawbacks.

Overall, the proposed approach to the puzzles explains the research questions about how and why language users reason about an utterance in a particular way — strategically, the move they take may not necessarily be a categorically optimal one, yet it needs to be a good enough one to accomplish the communication goal, all things considered (e.g., linguistic context, cost, etc.).

1.5 The structure of the dissertation

This dissertation concerns the symmetry phenomena in the non-scalar implicatures domain. The empirical data are degree expressions related. A pragmatic competition-based disambiguation account is proposed to explain why certain interpretations surface with respect to a degree utterance, and why certain degree utterances are salient with respect to a degree interpretation.

Specifically, the key components of the dissertation include the following:

- Theoretical background: the symmetry problem in the scalar and the non-scalar implicatures domains
- Empirical background: degree expressions and compositions
- Proposal: competition-based disambiguation models
- Proof of concept: artificial language learning experiments and the Rational Speech Act framework

For the background chapters, I review precursors' studies, provide both the theoretical and the empirical settings for the core research gaps, followed by initial attempts of address the research questions. The first attempts face challenges, both theoretically and empirically. This motivates a second attempt. Artificial language learning experiments are conducted as independent evidence

for the revised hypotheses, combined with recent probabilistic models simulating the proposed pragmatic reasoning process.

CHAPTER 2

BACKGROUND ON DEGREE SEMANTICS

In Chapter 1, a general overview of the standard degree-based analysis of gradable adjectives was given. This chapter starts with a close look at cross-linguistic variation in degree semantics in general and follows with a discussion about the semantics of degree expressions in Chinese in particular.

2.1 Degree semantics and cross-linguistic variation

Variation in comparative meaning is concerned with the strategy used to encode the following two parameters:

- (i) The comparison
- (ii) The semantics of the gradable adjectives

Regarding (i), a comparison establishes an ordering relation — greater or less than — between two individuals with respect to some measurement scale. There are two different sub-strategies within (i): the implicit strategy and the explicit strategy (Kennedy, 1999, 2019). Specifically, for the analysis under which the comparative relies on delineations/partitions, the ordering between the comparee and the standard is implicit, whereas for the analysis under which it encodes the *greater-than* relation, the comparison is explicit. Regarding (ii), it has been argued that the setting of the Degree Semantics Parameter (DSP) varies from language to language.

2.1.1 The morphosyntactic realization of the comparative

There are two strategies to realize the comparative: implicit comparison and explicit comparison. The two strategies differ in whether languages grammatically encode the ordering relation between two individuals with respect to certain measurement scale (Kennedy, 2007; Beck et al., 2009; Hohaus

& Bochank, 2020; Deal & Hohaus, 2019; Hohaus, 2015, 2018). For languages using the indirect strategy (i.e., implicit comparison) to make comparison, the ordering relation has to be inferred, whereas the greater/less-than relation is grammaticalized under the explicit strategy. Concretely, consider sentences in example (1) from Motu (Austronesian, Oceanic; Papua New Guinea) and example (2) Samoan (Austronesian, Oceanic; Samoa, American Samoa). They use a biclausal structure that Stassen (1985) calls the conjoined comparative construction.

- (1) *Mary-na lata to Frank-na kwadoḡi.*
 Mary-TOP tall but Frank-TOP short
 Mary is tall but Frank is short. (Motu, Beck et al. 2009:20 (66))
- (2) *E matua Ioane ‘ae la‘itiiti Malia.*
 TAM old John but young Mary
 John is old, but Mary is young. (Samoan, Hohaus 2018:112 (28))

On the other hand, there are situations in which languages mark both the adjective and the comparison standard with dedicated functional morphology, for example *-er* in English. Consider English and Samoan sentences in example (3) and (4) respectively.

- (3) Vera is older than Sandra. (Hohaus & Bochnak 2020:238 (3a))
- (4) *E umī atu Malia nai lō Ioane.*
 TAM tall more Mary from STD John
 Mary is taller than John. (Samoan, Hohaus 2012:335 (1))

Theoretically speaking, it depends on the analysis whether the difference in the morphosyntactic realization of the comparative translates to a difference in the underlying composition. The comparatives in examples (3) and (4) can hypothetically be analyzed in two different ways compatible with either the implicit or the explicit strategy: (a) the delineation-based analysis, namely comparative morphology as encoding *implicit* comparison (Klein, 1982; Van Rooij, 2011); (b) comparative morphology as encoding *explicit* comparison (Cresswell, 1976; von Stechow, 1984).

Regarding (a) the delineation-based analysis of (3) and (4), all that comparative morphology says is that partitioning some set of individuals outputs a comparison class, which contains at least the

comparee (i.e., Vera in (3)) and the comparison standard (Sandra in (3)). The comparee Vera counts as “old” and is consequently in what we call the positive extension of the adjective *old*. Nevertheless, the comparison standard Sandra is not. As in the case of conjoined comparatives in (2) and (1), the resulting truth conditions for (3) generate an inference that Vera’s age exceeds Sandra’s, without *explicitly* asserting this ordering. The basic ingredients of such analysis are sketched in (5).

(5) Lexical entries:

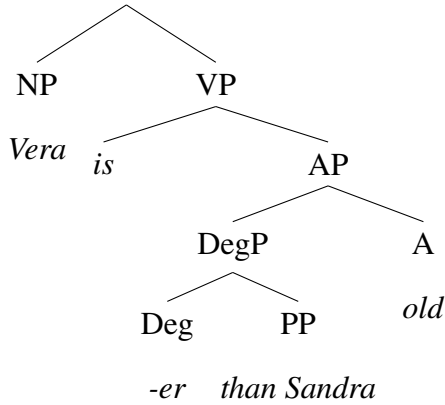
$$\llbracket old \rrbracket = \lambda C_{\langle e, t \rangle} . \lambda z: C(z) = 1. z \text{ is considered old with respect to } C$$

$$\llbracket -er \rrbracket = \lambda y. \lambda A_{\langle \langle e, t \rangle, \langle e, t \rangle \rangle} . \lambda x. \exists C' [C'(y) = 1 \ \& \ A(C')(x) = 1 \ \& \ A(C')(y) = 0]$$

(5) spells out the lexical entries for utterance in example (3) and (4). The logical form is given in (6).

(6) Logical Form (in brackets and in tree):

$$[[NP \text{ Vera}] [VP \text{ is } [AP [DegP [Deg \text{ -er } [PP \text{ than Sandra}]]] [A \text{ old}]]]]$$



(7) spells out the composition. (8) gives the truth conditions of the utterance (3). In plain English, the comparative reading of ‘Vera is older than Sandra’ is true if and only if there is a contextual comparison class such that both Vera and Sandra are in the comparison class C' , and Vera is considered old with respect to C' , and Sandra is not considered old with respect to C' .

(7) Composition:

$$\llbracket DegP \rrbracket$$

$$= \lambda A_{\langle \langle e, t \rangle, \langle e, t \rangle \rangle} . \lambda x. \exists C' [C'(Sandra) = 1 \ \& \ A(C')(x) = 1 \ \& \ A(C')(Sandra) = 0]$$

$\llbracket AP \rrbracket$

$= \lambda x. \exists C' [C'(Sandra) = 1 \ \& \ x \text{ is considered old w.r.t } C' \ \& \ Sandra \text{ is not considered old w.r.t. } C']$

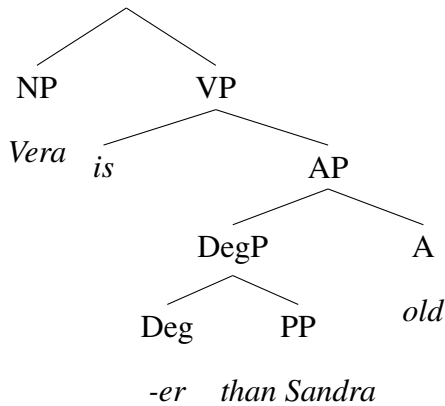
(8) Truth conditions:

$\exists C' [C'(Sandra) = 1 \ \& \ C'(Vera) = 1 \ \& \ Vera \text{ is considered old w.r.t } C' \ \& \ Sandra \text{ is not considered old w.r.t. } C']$

Under such an analysis, the difference between examples (1, 2) and (3, 4) would just be a matter of division of labor between syntax and morphology. The other kind of analysis incorporates morphology (i.e., the comparative morpheme) into the greater-than relation between two measurement degrees d and d' . Cresswell (1976) proposes that when making comparisons, speakers have points on a scale in mind. Under this analysis, degrees can be understood as abstract entities that are elements of scales, which are sets that come with a total ordering relation. Such an analysis is sketched by Hohaus & Bochnak (2020) as follows. The logical form of (3) under this analysis is given in (9).

(9) Logical Form (in brackets and in tree):

$[[NP \textit{Vera}] [VP \textit{is} [AP [DegP [Deg \textit{-er}] [PP \textit{than Sandra}]]] [A \textit{old}]]]]$



The lexical entries of (3) are decomposed in (10). It's worth pointing out that 'old' is interpreted differently in (10), in which $\llbracket old \rrbracket$ takes in a degree argument d and an individual x such that the age of x is greater than or equal to d .

(10) Lexical entries:

$$\begin{aligned}
\llbracket old \rrbracket &= \lambda d. \lambda x. AGE(x) \geq d \\
\llbracket -er \rrbracket & \\
&= \lambda y. \lambda A_{\langle d, \langle e, t \rangle \rangle}. \lambda x. MAX(\lambda d. A(d)(x) = 1) > MAX(\lambda d'. A(d')(y) = 1) \\
&\text{with } \llbracket MAX \rrbracket = \lambda P_{\langle d, t \rangle}. \lambda d \left[\forall d' \left[P(d') = 1 \rightarrow d \geq d' \right] \right]
\end{aligned}$$

The composition and truth conditions of (3) under this analysis are spelled out in (11) and (12) respectively.

(11) Composition:

$$\begin{aligned}
\llbracket DegP \rrbracket &= \lambda A_{\langle d, \langle e, t \rangle \rangle}. \lambda x. MAX(\lambda d. A(d)(x) = 1) > MAX(\lambda d'. A(d')(Sandra) = 1) \\
\llbracket AP \rrbracket &= \lambda x. MAX(\lambda d. A(d)(x) = 1) > MAX(\lambda d'. A(d')(Sandra) = 1) \\
&= \lambda x. AGE(x) > AGE(Sandra)
\end{aligned}$$

(12) Truth conditions:

$$AGE(Vera) > AGE(Sandra)$$

In plain English, under this analysis, (3) ‘Vera is older than Sandra’ is true under the condition that the age of Vera is greater than the age of Sandra.

To tease apart the two strategies, von Stechow (1984) points out that the implicit analysis is not available with English comparatives. A crucial diagnostics on differential measure phrases is proposed by von Stechow (1984) in (13):

(13) Vera is two years older than Sandra.

$$AGE(Vera) \geq SANDRA + 2 \text{ years} \quad (\text{Hohaus \& Bochnak 2020:239, example 4})$$

Under the implicit analysis, the comparee Vera would be classified as old and the comparison standard Sandra as not old. This classification does not allow us to define an addition operation. Therefore it does not allow us to meaningfully talk about age differences between Vera and Sandra.

- The implicit strategy: comparison class (partition sets of entities) and *inference* of the ordering relation

- The explicit strategy: explicitly establishing an ordering relation between two measurement degrees
- Both the implicit and the explicit strategies

Overall, Hohaus & Bochnak (2020) summarizes that languages differ in which strategy they adopt for the composition of comparative meaning, as illustrated in the itemized list above. Consequently, a crucial way to tell implicit from explicit comparison analysis is to examine whether it supports differential expressions.

2.1.2 Degree semantics parameter

Speaking of differential measurement, it always requires a certain semantics of the gradable adjective. As briefly discussed in the first section on canonical analysis of degree expressions, gradable adjectives map an entity to its measurement degree. The standard degree-based analysis has a measure function as the core of gradable adjectives' lexical entry. This measure function relates an entity to its maximal degree of certain dimension, as the function “AGE(x) $\geq d$ ” illustrated in (14).

$$(14) \quad \llbracket \sqrt{old} \rrbracket = \lambda d. \lambda x. \text{AGE}(x) \geq d \quad (\text{type } \langle d, \langle e, t \rangle \rangle)$$

To flesh out the canonical analysis of degree semantics introduced in the first section, under this standard canonical degree-based analysis, the adjective root needs to combine with a covert *pos* operator since it does not denote a *predicate* of individuals. The *pos* operator introduces norm/average-relatedness and vagueness. It also shifts the type to an $\langle e, t \rangle$ -type *predicate*. Hence the composition in (15):

$$(15) \quad \text{Kara is old.} \\ \llbracket \text{Kara} [\text{is} \sqrt{old} + \text{pos}] \rrbracket$$

For languages that *only* have implicit comparison strategies, they have been analyzed as “degreeless”, given that they do not encode an argument slot for a measurement degree into the semantic composition. They are interpreted relative to a contextual comparison class C in the

positive case, which are viewed as *vague* predicates. Under this analysis, consider the denotation of ‘old’ in (16):

$$(16) \quad \llbracket old \rrbracket = \lambda C_{\langle e, t \rangle} . \lambda x . x \text{ is considered old with respect to } C$$

Regarding the semantics of the gradable adjectives, Beck et al. (2009) and Krasikova (2008) suggest that languages vary between (14) and (16), and formally defines it as in (17). Essentially, this variation reflects a systematic parameter in the lexicon of natural languages. It’s worth highlighting that the setting of this degree semantics parameter is not determined by the morphosyntactic realization of the comparative alone.

- (17) Degree Semantics Parameter (DSP): A language {does/does not} have gradable predicates (type $\langle d, \langle e, t \rangle \rangle$ and related), i.e. lexical items that introduce degree arguments. (Beck et al. 2009:19, no. 62)

Against this background, Hohaus (2021) attempts to investigate into functional lexicon from the perspective of pragmatics, which is very reminiscent of the current proposal about Chinese adjectival constructions involving *gao* ‘tall’ and *hen*.

Hohaus believes the parameter proposed by Beck. Chinese, which behaves like Samoan, has been argued to be a $[+DSP]$ language (Gong and Coppock, personal communication). *hen* takes in a degree argument (e.g., *gao* “tall”). Hypothetically, the analysis of Chinese comparatives should be in the same line of the analysis of other $[+DSP]$ language such as Samoan.

Specifically, Hohaus (2021) studies Samoan, which as mentioned earlier is a Austronesian - Oceanic language that lacks a degree-based superlative operator but exploits mechanisms from other domains to express superlative-related interpretations $[+DSP]$. Hohaus (2015) presents that Samoan *atu*-comparative employs the INDIRECT STRATEGY as illustrated in (18), in which comparison is made against some contextually given standard to be provided by an explicitly frame-setter.

- (18) Compared to Alex, Jane is older.
 (= *Compared to the Alex-degree, Jane is older*)
 \leadsto Jane is older than Alex.

This also works with UNMARKED POS scale-adjectives, in Samoan, English, and Chinese:

- (19) *E matua Malia i lō Pita*
 TAM old Mary PREP COMP Peter
 Compared to Peter, Malia is **old**.

According to Hohaus (2015), (19) shows that comparative interpretation is driven by the presence of a single explicit contextual ALTERNATIVE, and the superlative interpretation is driven by the presence of $n > 1$ explicit or implicit contextual ALTERNATIVES. Hohaus also stresses that the number of alternatives has an effect on final interpretation. It seems that similar to Chinese, degree scales in Samoan are crucial in expressing superlative forms that are unmarked. For example:

- (20) CONTEXT: Mount Silisili is 1,858m, Mount Fito is 1,028m, and Mount Afi is 1,563m.

E au-pito maualuga Mauga Silisili.
 TAM **without=next high** mountain NAME

(Lit.) ‘Mountain Silisili is high without next.’

Under the same context illustrated in (20), *gao* ‘high/tall’ would be interpreted as *higher than contextually salient alternatives/mountains*. Consider the Chinese sentence in example (21).

- (21) *Silisili shan gao.*
 Silisili mountain high
 Mountain Silisili is high(-er than contextually salient mountains, including Mount Fito and Mount Afi).

This again speaks to Hohaus’s proposal (2021) that with respect to cross-linguistic comparison, implicit comparison and hence implicit or pragmatic superlative formation should be, theoretically speaking, possible in any language. The Chinese sentence in example (21) appears to show that the comparative/superlative interpretation of bare adjectives in Chinese can be derived *pragmatically*. Later in Chapter 3 and Chapter 5, I will establish this claim with more empirical data and a more elaborate proposal.

That said, most of the previous studies on degree constructions in Chinese propose syntactic-semantic analyses (Liu, 2018; Zhang & Grano, 2019; Xiang, 2005), with only a few making pragmatic proposals (Zhang 2019).

2.2 Degree expressions in Chinese

This section focuses on descriptions as well as previous analyses of degree expressions in Chinese. Reviews are given with respect to how the proposed analyses succeed or fail to explaining the empirical data.

2.2.1 General descriptions about Chinese gradable expressions

According to the earlier works by Li & Thompson (1981), the basic pattern of comparatives in Chinese is (22), in which X is the topic under comparison, Y is the standard degree that X is compared to, and “dimension” is specified by the predicate, along which the comparison is made.

(22) X comparison word Y (adverb) dimension

Li & Thompson (1981) summarize that there are three types of comparison words:

(23) *Ta bi ni gao.*
3SINGULAR than 2SINGULAR tall

(S)he is taller than you.

Superiority: *bi* (“than”)

(24) *Ta mei.you/bu.ru ni gao.*
3SINGULAR NEG.have/NEG.as 2SINGULAR tall

(S)he isn’t as tall as you.

Inferiority: *mei.you/bu.ru...* (literally “not as...as...”)

(25) *Ta gen ni yiyang gao.*
3SINGULAR with 2SINGULAR same tall

(S)he is as tall as you.

Equality: *gen...yiyang* (literally “with...same...”)

Xiang (2005) further proposes that Chinese has two types of superiority comparatives. One is called *bi*-comparative, as in (23). The standard degree can also be introduced without *bi*, as shown in (26). This is called bare comparative.

- (26) *Anna gao wo yi.cun.*
 Anna tall 1SINGULAR one-inch
 Anna is one inch taller than me.

The fact that a bare adjective such as *gao* can participate in comparison constructions (23—26) with no overt comparative morphemes at all suggests that *gao* really does seem to have a comparative reading. This is different from English, which denotes degree addition with a separate comparative morpheme *more*. Put otherwise, bare adjective showing up in superiority comparatives (23) and (26) is an indicator that a bare adjective utterance such as (8) “*Anna gao*” has a *comparative* reading in Chinese.

It is worth restating that in this thesis, by “comparative reading” I mean *more than discourse salient referent(s) / everyone else salient in the context*; and “positive reading” means *more than the norm / average given by the judge, who in most cases is the speaker in the context*. This assumption is not universally agreed upon and it might be an oversimplification. Nevertheless, I will use these paraphrases for simplicity, since the core contribution of the current study is to develop a theory to account for the systematic ambiguity and which meaning surfaces. This thesis is not committed to giving a comprehensive account for the structure of various kinds of comparative and positive constructions. Rather, the thesis focuses on utterances involving bare adjectives such as (8), which can be construed as positive and comparative, despite the lack of an overt comparative morpheme and that of a pronounced standard (degree) argument. Why is it possible for (8) to mean what it means? This dissertation argues that the answer lies in pragmatics: what we don’t say matters.

According to Kennedy’s (2019) recent work, key features of degree expressions in Chinese include (i) the possibility of predicate-marking morphology; (ii) the possibility of comparison without standard morphology; and (iii) clausal comparatives. This thesis focuses on the first two features, since they are directly relevant to the empirical observations to be made in the following chapter.

Regarding (i) the possibility of predicate-marking morphology, prototypical comparatives in Chinese do not include predicate-marking morphology, and even appear not to allow it as shown in (27). The closest form to convey the thought in (27) would be (28), in which the intensifier

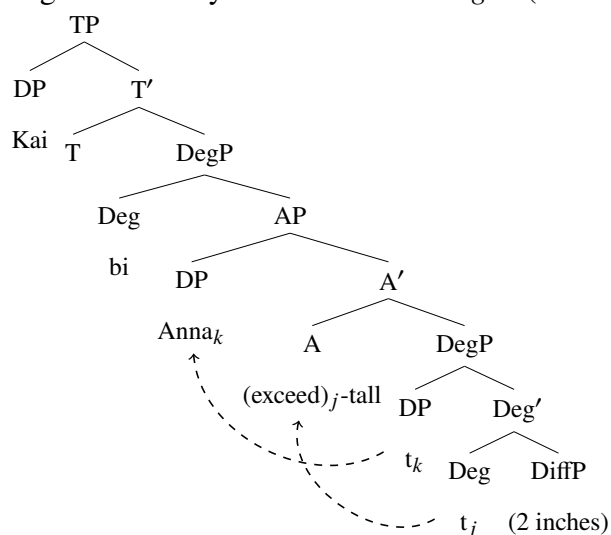
morpheme *geng* is optional.

- (27) *Zhangsan bi Lisi (*bijiao) gao.*
 Zhangsan BI Lisi (*more) tall
 Zhangsan is taller than Lisi.

- (28) *Zhangsan bi Lisi (geng) gao.*
 Zhangsan BI Lisi (even-more) tall
 Zhangsan is even/still taller than Lisi.

Further, regarding the derivation of prototypical comparatives such as *Anna bi Kai gao* “Anna is taller than Kai”, Xiang (2005) proposes (29) and maintains that the Deg-head moves to the higher DegP shell position to introduce its external argument, and at this point of the derivation, the morpheme *bi* can be externally merged with the AP and project the higher DegP-shell. In this way a *bi*-comparative is derived, as shown in (29).

- (29) DegP-Shell analysis for *Kai bi Anna gao* (2 inches) (Xiang 2005:62)



2.2.2 Bare predication *has to be* COMP

In Mandarin, there is a phenomenon that is at odds with the cross-linguistically stable asymmetry between positive and comparative forms:

		<i>Positive form</i>	<i>Comparative form</i>
	English	tall	taller
(30)	Irish	ard	arda
	French	grand	plus grand
	Japanese	takai	takai

- (31) **Descriptive generalization:** Cross-linguistically, the comparative form of a gradable adjective is derived from or identical to its positive form (Grano 2012; Grano and Davis 2018).

If (31) holds universally, then it rules out two of four hypothetically possible derivational relationships that could hold between positive- and comparative-form adjectives (Grano and Davis 2018):

- (32) Grano and Davis' observation (2018)

	<i>Positive form</i>	<i>Comparative form</i>	<i>Examples</i>
Pattern A	Adj	Adj	Japanese, ...
Pattern B	Adj	DERIV(Adj)	English, Irish, French, ...
Pattern C	DERIV(Adj)	Adj	Impossible?
Pattern D	DERIV ₁ (Adj)	DERIV ₂ (Adj)	Impossible?

Superficially, Mandarin instantiates Pattern C (Sybesma 1999; Huang 2006; Gu 2008):

- (33) The Mandarin *hen* puzzle:

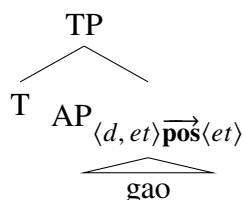
- a. *Anna gao.*
Anna tall
'Anna is taller (than someone known from context).'
- b. *Anna **hen** gao.*
Anna HEN tall
'Anna is tall.'

Grano (2012) argues that despite surface appearances, Mandarin is a Pattern A language. He attempts to address the question of why COMP and *hen* pattern together to the exclusion of POS in the configuration in (34).

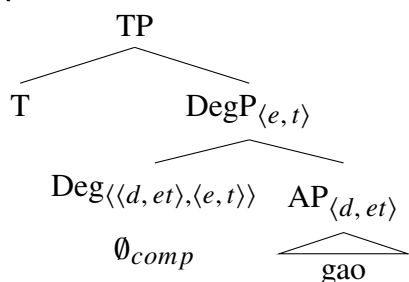
- (34) a. *Anna [POS gao]. *Intended:* ‘Anna is tall.’
 b. Anna [COMP gao]. = ‘Anna is taller.’
 c. Anna [**hen** gao]. = ‘Anna is tall.’

The core proposals have two crucial pieces, one is called ‘Universal Markedness Principle’, which states that Universally, comparative semantics is provided by an explicit morpheme in syntax which is overt in some languages and null in others, whereas positive semantics is provided by a type-shifting rule that does not project in syntax. The other concerns language specific constraint, which is termed ‘The T[+V] constraint’ (see (36)). In Mandarin, the direct complement to T(ense) must either be a verb (35b) or a functional morpheme that can combine with a verb (35c). The consequences of these proposals are demonstrated below. Note that T here stands for whatever head hosts the subject in main clauses.

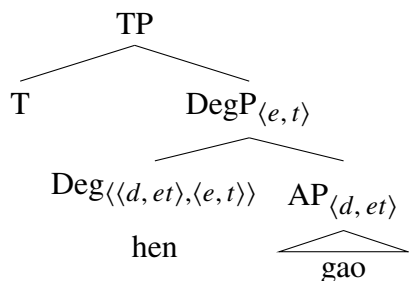
- (35) a. *



- b. ✓



- c. ✓



According to Grano (2012), (35a) is illegal because the status of AP would not be changed by the pos, which would cause violation of the T+[V] constraint, as illustrated in (36), which is a syntactic constraint that forbids bare AP complements to T(ense). On the other hand, both (35b) and (35c) are legal, because COMP and the overt morpheme *hen* would result in a projection called DegP (Degree Phrases; see Kennedy 2002; Liu 2010 for more details), and would not violate the T+[V] constraint.

- (36) The T+[V] constraint: In Mandarin, the direct complement to T(ense) must either be (an extended projection of) a verb or a functional morpheme that can in principle combine with (an extended projection of) a verb.

Note that according to Grano (2012), *hen* is used to achieve positive semantics while satisfying the T+[V] constraint. Its meaning as an “intensifier” is semantically bleached in contexts where it is obligatory, while it still has a mild meaning in contexts where it is not. This is line with Liu (2010). Grano (2012) further suggests that *hen* is very reminiscent of the auxiliary *do* in English, which is semantically vacuous where it is required, but it is contentful in contexts where it is optional.

Grano (2012) predicts that bare gradable adjectives (with positive semantics) should be licit when structure intervenes between T and AP, and they should also be licit when T is not projected. The predictions are supported by sentences like polar questions, all of which can be analyzed as involving a projection of Laka’s (1990) Σ between T and AP.

Following Grano (2012), Zhang (2019b) extends the analysis to property concepts constructions. Under the proposal of Zhang (2019b), we can infer that bare adjectivals *have to* be interpreted as COMP, whereas *hen has to* be interpreted as pos. Table 2.1 is a summary of the generalization.

	POS	COMP
Eval Adj	#	✓
<i>have</i> +N _{quality-wisdom}	#	✓
<i>hen</i> Eval Adj	✓	#
<i>hen have</i> +N _{quality-wisdom}	✓	#

Table 2.1: Y.Zhang & Grano's (2019) observation about how to interpret *hen* and bare adjectivals

In terms of the specifics, Grano (2012) and Zhang (2019b) share the assumption that POS as a type-shifter, different from COMP, does not have syntactic projections. The assumption seems stipulative. Furthermore, Σ in Laka (1990) is high, which is realized in CP. And CP is *above* T. But it is structured in between T and AP according to Grano (2012) and Zhang (2019b).

2.2.3 Bare predication *can* be POS

Zhang (2019a) gives data challenging Grano's (2012) analysis. It has been argued in Zhang (2019a) that bare predication (e.g., bare adjectivals) does not have to be POS, rather it *can* be POS.

- (37) a. *Anna gao*
Anna tall(-er)
Anna is taller (than someone known from context).
NOT: Anna is tall (Sybesma(1999), Grano(2012))
- b. *Anna hen gao*
Anna HEN tall(-er)
Anna is tall (Sybesma(1999), Grano(2012))

According to L.Zhang (2019), (37a) does not sound perfectly unambiguous when uttered out of the blue. Supporting evidence is given in (38) and (39), suggesting that bare adjectivals can have both interpretations available (POS and COMP). In other words, plain form adjectivals can be POS.

- (38) a. *Zhe-xie hai-zi li jiu Anna gao*
these kids among/inside only Anna tall(-er)
Among these kids, only Anna is tall. (POS)

- b. *Anna gao bu gao? Anna (bu) gao.*
 Anna tall(-er) NEG tall(-er) Anna NEG tall(-er)
 Is Anna tall? Anna is (not) tall. (POS)

- (39) *Anna he Bob shui gao? Anna gao*
 Anna and Bob who tall(-er) Anna tall(-er)
 Between Anna and Bob, who is taller? Anna is taller. (COMP)

(40a) with measurement expression is predicted by Grano to be unambiguous COMP, because the combination of ‘long’ and ‘er’ is required to satisfy the T[+V] constraint. But, the fact is that (40a) is ambiguous when uttered out of the blue, and the measurement reading is preferred.

- (40) a. *Zhe gen sheng-zi chang liang mi*
 this CLASSIFIER rope long(-er) two meter
 This rope is 2 meters long/longer. (ambiguous)
- b. *Zhe gen sheng-zi (bi na gen) chang (le) liang mi*
 this CLASSIFIER rope COMP that CLASSIFIER long(-er) ASPECT two meter
 This rope is 2 meters longer (than that one). (COMP)

The preferred addition of aspectual *le* is particularly puzzling for Grano: if a silent morpheme ‘-er’ already satisfies T[+V], why can this *le* be licensed (40b)?

- (41) a. *yi feng chang de xin*
 one CLASSIFIER long(-er) PARTICLE letter
 a long letter (POS)
- b. **(hen) chang de yi feng xin*
 HEN long(-er) PARTICLE one CLASSIFIER letter
 Intended: a long letter (POS (*hen* is obligatory))

(41) contain a relative clause (henceforth RC), because fronted adjectival has to be parsed as RC. When analyzed as RC, *chang de* (‘(NP) that is long/longer’) should still need a silent ‘-er’ to satisfy the T[+V], since it’s at clausal level and it projects T. (41b) is predicted by Grano to be interpreted in a comparative way. But, (41a) has by no means a COMP reading.

Grano’s analysis predicts a comparative reading for (37a) and (40a), where actually, ambiguity arises. Grano’s analysis should predict ambiguous readings for (41a), which has only an unambiguous positive reading.

Zhang (2019a) suggests that (37) and (40a) speak against any morphosyntactic unbalance between comp and non-comp use. Zhang (2019a) analyzes the semantics of gradable adjectives as a relation among three items. The comparison between the measurement of an individual x and a certain **standard** σ results in a **difference** δ . The three uses of gradable adjectives differ in their arguments σ and δ . The use of gradable adjective is inherently ambiguous. The use of other elements (like *hen*) helps to disambiguate. The non-use of these disambiguating elements leads to a seemingly ‘default’ interpretation.

	POS	COMP
Eval Adj	✓	✓
<i>have</i> +N _{quality}	N.A.	N.A.
<i>hen</i> Eval Adj	✓	#
<i>hen have</i> +N _{quality}	N.A.	N.A.

Table 2.2: Implication of Zhang’s analysis (2019) (inspired by Liu (2010b) and Liu (2018))

More recently, Liu (2018) shows that predicate-marking morphology is acceptable as long as the standard is omitted. Consider the sentence in example (42), in which the comparative meaning is available yet the standard is omitted. Overall Liu (2010; 2018) suggests that Chinese has predicate-marking comparative morphology with overt and covert allomorphs, whose form is conditioned by the absence versus presence of an overt standard.

- (42) *Zhangsan bijiao gao.*
 Zhangsan more tall
 Zhangsan is taller.

Regarding feature (ii), namely the possibility of comparison without standard morphology, earlier works by Chao (1968) discuss question (Q) — answer (A) pairs as in (43), in which both the

question in (43a) and the response in (43b) convey a comparison thought using bare adjectival form *gao* ‘tall’.

- (43) a. *Q: Tamen, shei gao (ne)?*
Q: They who tall (SFP)
Which of them is taller?
- b. *A: Lao Er gao.*
A: Lao Er tall
Lao Er is taller.

Moreover, the possibility of comparison without standard morphology concerns transitive comparatives (Xiang 2005; Grano and Kennedy 2012), as shown in (44) and (45).

- (44) *Zhangsan gao Lisi *(san gongfen).*
Zhangsan tall Lisi three centimeters
Zhangsan is *(three centimeters) taller than Lisi.
- (45) *Zhangsan zhong Lisi *(san gongjin).*
Zhangsan heavy Lisi three kilograms
Zhangsan is *(three kilograms) heavier than Lisi.

The asterisk in (44) and (45) indicates that the measurement phrases ‘three centimeters’ and ‘three kilograms’ are obligatory in making transitive comparisons. The absence of measurement phrases makes the sentences in (44) and (45) ungrammatical. Nevertheless, note that measure phrases do not *force* comparative meanings. Consider the sentences in example (46).

- (46) a. *Zhangsan gao liang mi.*
Zhangsan tall two meters
Zhangsan is two meters tall. OR
Zhangsan is two meters taller.
- b. *Zhangsan liang mi gao.*
Zhangsan two meters tall
Zhangsan is two meters tall.
*Zhangsan is two meters taller.

With bare form *gao* ‘tall’, (46a) is ambiguous between positive ‘two meters tall’ and comparative ‘two meters tall-er’. However, the ambiguity disappears when measurement phrase ‘two meters’ precedes *gao* ‘tall’. Further, transitive comparatives show that standards can be composed even without *bi*. Consider the sentence in example (47), in which *gao* ‘tall’ is used to convey a comparative meaning.

- (47) *Zhangsan zhe-ge xingqi gao ta shang-xingqi san gongfen.*
 Zhangsan this-CLASSIFIER week tall he last-week three centimeters
 Zhangsan is three centimeters taller this week than he was last week.

Liu (2010b) and Liu (2018) show that predicate-marking morphology is acceptable as long as the standard is omitted (48). Liu proposes that Chinese has predicate-marking comparative morphology with overt and covert allomorphs, whose form is conditioned by the absence and the presence of an overt standard.

- (48) *Zhangsan bijiao gao/you zhihui.*
 Zhangsan more tall/have wisdom
 Zhangsan is taller/smarter.
- (49) *Anna he Bob, shei bijiao gao?*
 Anna and Bob, who more tall
 For Anna and Bob, who is taller?

Liu (2018) conclude that the Chinese covert comparative marker is the covert allomorph of the comparative morpheme *bijiao* ‘more’. The Chinese positive morpheme has two allomorphs. One is the unstressed *hen* HEN and the other is its covert counterpart. A covert allomorph, regardless of whether it is the comparative or the positive morpheme, is used simply to avoid violating the Constraint on Multiple-Foci.

To explain the Mandarin data, Liu adopts the idea that positive semantics is provided by an explicit functional morpheme *pos*, and proposes that *pos* in Mandarin has two allomorphs: the covert version as found in other languages, and the overt version *hen*. Crucially, the covert version in Mandarin is analyzed as a polarity item subject to licensing conditions. Specifically, *pos* must be in the domain of a predicate accessible operator_[-wh]. In Liu’s words:

- (50) In Chinese, the covert positive morpheme only occurs in a predicate accessible operator_[-wh] domain with a structure like [Op_[-wh] ... X⁰_[-wh-operator] [DegP... Deg⁰ [AP...]]], where the head X⁰, carrying the predicate-accessible operator _[-wh] feature, not only introduces a predicate-accessible operator_[-wh] but also functions to license the occurrence of a degree phrase headed by the covert positive morpheme (i.e., Deg⁰). And this domain must be contained in the smallest clause that contains the adjectival predicate and the operator. (Liu 2010b:1019)

Thus in a simple matrix-level declarative sentence, there is no appropriate operator to license covert POS, and hence *hen* must be used instead. In a variety of other kinds of constructions, however, such as negation, *ma* particle questions, contrastive focus, and embedded epistemic clauses, there is an appropriate operator to license covert POS and so *hen* is not required. In a negated sentence, for example, the negation morpheme *bu* is an appropriate operator to license covert POS:

- (51) Zhangsan [_{NegP} Op [_{Neg} *bu*_{+operator}][_{DegP} POS [_{AP} *gao*]]].
(Liu 2010b:1025)

Similarly, in the case of contrastive focus and embedded epistemic clauses, Liu proposes that they involve a null focus and epistemic operator, respectively:

- (52) Zhangsan [_{FocP} Op [_{Foc}⁰_{+operator}][_{DegP} POS [_{AP} *gao*]]]... (Liu 2010b:1028)
- (53) [_{CP} Zhangsan yaoshi [[_{EpistP} Op [_{Epist} **must**_{+operator}][_{DegP} POS *gao*]]] de-hua]...
(Liu 2010b:1032)

Although it accounts for a wide range of data, Liu's account does not explain why a comparative interpretation arises when covert POS is not licensed. A related point is that nothing in Liu's theory explains why it is in particular POS and not any other morpheme, e.g., a comparative morpheme, that has two allomorphs, with the null version behaving like a polarity item. Ideally, we would want to derive the behavior of POS in Mandarin from more general properties of POS cross-linguistically.

The reason for the _[-wh] specification in Liu's formulation is to capture the fact that in *wh*-questions, covert POS is not licensed and instead a comparative interpretation is found:

- a. $\llbracket \text{pos} \rrbracket = \lambda P_{\langle d, \langle e, t \rangle \rangle} \lambda x. \exists d [P(d)(x) \wedge d > d_s]$, for some contextually valued standard d_s .
- b. $\llbracket \text{John is pos tall} \rrbracket = \exists d [\text{tall}(d)(\text{John}) \wedge d > d_s]$
- (60) a. *Anna you zhihui.*
 Anna have wisdom
 ‘Anna has wisdom.’
- b. *Anna congming.*
 Anna smart
 ‘Anna is smarter (than somebody salient in c).’

The last two conjuncts $|z| \geq d \wedge d > 0_a$ in (57c) are redundant, due to the $\exists z$. This follows that ‘have water’ doesn’t need to project a degree argument. But, $|z| \geq d \wedge d > 0_a$ in (58c) are no longer entailed by the $\exists z$ because of 0_r . It follows that ‘have wisdom’ is gradable and needs a deg argument. This makes another wrong prediction. First, $|z| \geq d$ is still obligatory to make reference to the relation between rice and d. Even if we drop $d > 0_a$, it still makes wrong predictions—the fact is that (61) is infelicitous.

- (61) *Anna bi Bob you dami.*
 Anna COMP Bob have rice
 Intended: Anna has more rice than Bob.

- (62) $\llbracket \text{Anna bi Bob you dami} \rrbracket$
 $= \max\{d: \exists z [\text{rice}(z) \wedge \pi(\text{Anna}, z) \wedge |z| \geq d]\} > \max\{d: \exists z [\text{rice}(z) \wedge \pi(\text{Bob}, z) \wedge |z| \geq d]\}$

Is ‘tall’ associated with 0_a or 0_r ? Li would predict that ‘height’ belongs to the ‘ratio’ scale set, which has an absolute zero. But that would contradict (59).

- (63) Li (2017): Measurement Scale:

	POS	COMP
Eval Adj	#	✓
<i>have</i> +N _{quality-wisdom}	✓	#
<i>have</i> +N _{quality-idea}	✓	#
<i>hen</i> Eval Adj	✓	#
<i>hen have</i> +N _{quality-wisdom}	N.A.	N.A.
<i>hen have</i> +N _{quality-idea}	N.A.	N.A.

2.3 Chapter summary

This chapter attempts to give a thorough review of previous studies on degree expressions in Chinese, which are the core data that this thesis will investigate into. Both standard degree-based analysis and cross-linguistic perspectives are presented in this chapter. Against this general background about how to interpret degree expressions, basic features as well as proposed analyses on Chinese adjectives are discussed. There are three research gaps along the lines of investigation: (i) the interpretation of bare form adjectives such as *gao* ‘tall’; (ii) the analysis of “modified” form adjectivals such as *hen gao* (positive) and *bi gao* (comparative); (iii) the role of pragmatics in the derivations.

As briefly mentioned in the introduction chapter, there are ongoing debates on how to interpret *gao*: is it comparative or positive, or both? Regarding (ii), the complexity of the structures of *hen gao* and *bi gao* varies depending on the analysis. When it comes to (iii), we might see that so far almost all of the analyses are semantics/syntax driven, yet very few attempts concern pragmatics. To address these questions, the following chapters will start with answering (i), and then the next two chapters will take a close look at (ii) and (iii).

CHAPTER 3

A MORE DETAILED LANDSCAPE OF THE DATA

The debate in the degree semantics literature revolves around the interpretations of utterances involving adjectives in plain form. Interpretation is part and parcel of assessing the potential truth conditions of a sentence relative to various contexts. If there is disagreement about whether “*Anna gao*” means POS or COMP or both, one needs to be careful about the context that is being talked about. In other words, what matters is the truth value *relative to the context*. One of the most widely used methodologies have been employed for detecting the possible meaning(s) when the context is known.

3.1 Methodology for detecting ambiguity

Let us suppose a sentence S has two interpretations ϕ and ψ . If S can be judged as true with ϕ as the intended meaning and false with ψ as the intended meaning at the same time in the same context, and vice versa in a different context, then the sentence S is ambiguous between ϕ and ψ . This test of ambiguity is called the test of contradiction in the literature (Gillon, 1990, 2004). It states that a sentence is ambiguous if it can be both truly affirmed and truly denied for a given state of affairs. The clearest cases of ambiguity include lexical ambiguity and structural ambiguity. Structural ambiguity is a type of ambiguity that when an expression can accommodate more than one structural analysis, it is ambiguous. Consider the sentence (1).

(1) Bill saw a man with a telescope. (utterance)

(2) a. Bill saw Fred. Fred possesses no telescope. Bill used a telescope to see Fred.

b. Fred was in possession of a telescope and Bill saw Fred with his naked eyes.

(context)

(3) a. Bill [_{VP} saw [_{DP} a man] [_{PP} with a telescope]]

b. Bill [_{VP} saw [_{DP} a man [_{PP} with a telescope]]]

(simplified structure)

(1) is judged both true and false with respect to the same context specified. In context (2a), the reading of (1) derived by (3a) is judged true, and the other reading of (1) derived by (3b) is judged false in the same context (2a). Relative to context (2a), (1) can be both asserted and denied. When the speaker asserts it truthfully, the speaker is using structure (3a). Whereas when the speaker is denying it, the speaker is denying (3b). Therefore, utterance (1), relative to context (2a), is both true and false. Sentence (1) is genuinely ambiguous. The systematic ambiguities of the degree constructions in Chinese can be illustrated with a similar paradigm.

3.2 Theoretical assumptions

In this section, I extend the methodology to Chinese degree expressions and argue that, based on introspective data, utterances involving plain form adjectives are ambiguous, and their surfaced interpretation is highly sensitive to linguistic environments.

3.2.1 Basic cases

Let us consider whether (4) can be judged true or false under a context where there are two individuals Anna and Kai. They are both below average tall. Anna is taller than Kai. Utterance (4) in the form of a bare adjectival can only be judged true. This suggests that a comparative reading is available for (4), but not a positive reading.

Note that these judgments are my own (i.e., introspective judgment data). This judgment is in line with previous claims that “Anna gao” has only a COMP reading. Nevertheless, it’s worth highlighting that this judgment is preliminary, a kind of “on first glance” judgment – because later I will show that the POS reading of “Anna gao” can surface.

(4) Anna gao.

Anna tall

When a degree modifier such as *hen* is present, (5) is interpreted as “Anna is above average tall”, and it can be only judged false in that same context. This suggests that a positive reading is available for (5), but not a comparative reading.

(5) Anna *hen* gao.

Anna HEN tall

(4) and (5) indicate that in a comparative context in which Anna is taller than contextually salient people and Anna is below average tall, the bare form of a degree construction without *hen* is necessarily judged true. With an overt degree modifier as in (5), the utterance is necessarily judged false. In the scenario below, truth value judgments get reversed.

Consider a context where there are two individuals Anna and Kai. They are both above average tall. Kai is taller than Anna. The salient inference is that (4) is false in this context. This further confirms that the comparative reading is the salient one for (4), and that a positive reading is unavailable or less salient than COMP, because (4) would be judged true if POS is salient. On the other hand, (5), which unambiguously means “Anna is above average tall”, can only be judged true under that context. This confirms that a positive reading is available for (5). In a positive context, the bare form of a degree construction is saliently judged false, yet when the degree modifier is overt, it is necessarily judged true.

So far according to Gillon’s methodology, it appears that “Anna gao” means *unambiguously* COMP, which supports the claims made by Grano (2012).

3.2.2 A more complex picture - Introspective data

Syntax disambiguates *gao* Now suppose there are four individuals, Anna, Bob, Kai, and Lee. Kai and Lee are very short. Anna and Bob are very tall, and Bob is shorter than Anna. (6) is both felicitous and true. It is explicitly disambiguating through the presence of the morpheme *ye* “also/and”. Bare adjective is semantically ambiguous. My claim is that it’s not always possible to detect the ambiguity, because it’s always being disambiguated. (6) is one way to (explicitly) disambiguate *gao* in a sense that the two occurrences of *gao* can’t both have a COMP interpretation, which would otherwise be contradictory. *Anna gao* entails that Anna is taller than all the rest, including Bob. Thus a continuation of *Bob ye gao_{comp}* would contradict the initial proposition.

- (6) *(Zhexie haizi li) Anna gao Bob ye gao.*
 These kid among/inside Anna tall Bob also tall

Intended: (Among these kids,) Anna is tall and Bob is tall too.

To sum up and to highlight the tension, on the one hand, out of the blue, “Anna gao” unambiguously (*or at least saliently*) means COMP and “Anna hen gao”, unambiguously POS. Both seem to be totally unambiguous. On the other hand, the actual facts are more complex than that. (6) shows that bare adjectives such as “gao” really are ambiguous, because they can be explicitly disambiguated.

Later on data show that under negation, both readings are accessible. This dissertation will argue that Gillon’s methodology is actually incomplete. An expression can be genuinely ambiguous and have multiple interpretations accessible, but in a way that does not reveal itself in terms of truth-value judgments in out of the blue context such as the baseline cases in section 2.2.1, as well as those reported in Grano (2012).

Downward entailing environments Consider a context where there are four individuals Anna, Kai, Lee, and Al. All of them are below average tall. Anna is taller than Kai, Lee, and Al. Under that context, it *can* be **both true and false** to utter (7). In that context, no salient individual exceeds the average degree of height, but at the same time, Anna does exceed every other salient individual’s degree of height. Since (7) can be judged true relative to the context, it follows that the bare adjective may be interpreted in a positive, rather than comparative way: no one is such that they are taller than average.

- (7) *Zhexie haizi li mei.you yi.ge xiaohai gao.*
 These kid among/inside NEGATION.have one-CLASSIFIER kid tall

Intended: Among these kids, no one is above average tall.

Conversely, (7) may also be judged false relative to that context. A dialogue (8) helps us to see why (7) can be judged false. In (8), the second speaker corrects the first speaker by using *gao* to convey COMP. The fact that (8) sounds natural indicates that (7) can be false. Since (7) can be judged false relative to that context, it follows that a bare adjective may be interpreted in a comparative,

rather than a positive way: it is false that no one is taller than every other salient person, because Anna is taller than every other salient person.

- (8) • *Zhexie haizi li mei.you yi.ge xiaohai gao.*
 These kid among/inside NEGATION.have one-CLASSIFIER kid tall
 Intended: Among these kids, no one is taller than everyone else salient in the context.
- *Bu, ni cuo le. Anna gao.*
 No, you wrong PARTICLE Anna tall
 Intended: No, you are wrong. Anna is taller than contextually salient people.

Under the same context, it *has to be true* to utter (9). Thus, *hen* together with the adjective has only a positive reading available. Unlike (7), (9) can only be judged true and it is thus not ambiguous.

- (9) *Zhexie haizi li mei.you yi.ge xiaohai hen gao.*
 These kid among/inside NEGATION.have one-CLASSIFIER kid HEN tall
 Intended: Among these kids, no one is above average tall.

To see why (9) cannot be judged false, consider the dialogue in (10). The fact that a response in (8) sounds natural whereas the same response sounds odd in (10) indicates that under the same context, degree constructions without *hen* have both the positive and the comparative reading available. But degree constructions with *hen* only has positive reading. The reply in (10) is infelicitous with or without *hen*.

- (10) • *Zhexie haizi li mei.you yi.ge xiaohai hen gao.*
 These kid among/inside NEGATION.have one-CLASSIFIER kid HEN tall
 Intended: Among these kids, no one is above average tall.
- *# Bu, ni cuo le. Anna (hen) gao.*
 No, you wrong PARTICLE Anna (HEN) tall

A similar paradigm is discovered in negation. Suppose the intended meaning is Anna is not above average tall, (11) *can be both true and false* under the same context where all the individuals are below average tall, and Anna, as one of the individuals, is taller than other individuals in the context.

- (11) *Anna bu gao*
 Anna NEGATION tall

A natural dialogue is given in (12), which shows that the comparative interpretation *can* be judged true, and the positive reading can be judged false. In other words, (11) can be judged true **and** false relative to the same context.

- (12) • *Anna bu gao.*
 Anna NEGATION tall
- *Bu, ni cuo le. Zhexie haizi li, Anna gao.*
 No, you wrong PARTICLE these kid among/inside Anna tall
 Intended: No, you are wrong. Among these kids, Anna is taller than everyone else
 salient in the context.

Yet again, when *hen* is present, ambiguities disappear. Under the same context, it *has to* be **true** to utter (13).

- (13) *Anna bu shi hen gao*
 Anna NEGATION COP HEN tall
 Anna is not above average tall.

Now consider another operator that licenses set to subset entailment: *only*. Consider a context where there are four individuals Anna, Bob, Al, and Lee, they are all below average tall; yet Anna is taller than Bob, Al and Lee. In this comparative downward entailing context, the utterance with the intended meaning COMP sounds odd (14a).

- (14) a. *?/# Zhexie haizi li, jiu Anna gao.*
 These kids among/inside, only Anna tall
 Intended: Among these kids, only Anna is taller (than everyone salient in the context).
- b. *?/# Zhexie haizi li, jiu Anna zui gao.*
 These kids among/inside, only Anna SUPERLATIVE tall
 Intended: Among these kids, only Anna is taller (than everyone salient in the context).

The oddness of (14a) may, on first glance, suggest that “gao” does not have a COMP reading here. However, oddity may be attributable to the vacuousness of *zui* “only”: it’s necessarily the case that

only one person can be taller than everyone else, therefore saying “only” is superfluous. This can be confirmed in (14b), with a superlative *zui*, which shows the exact same behavior: only one person can ever be the tallest, thus uttering “only” is superfluous.

In downward entailing environments, besides the comparative reading, the bare form of degree constructions can really have a positive reading, but a degree construction with an overt *hen* morpheme only has the positive but not the comparative reading.

A note on Superlatives in Chinese As shown in the previous discussion, the existence of the superlative morpheme *zui* indicates that Chinese does not lack a way to express the meaning of superlative. Similar to English, the usage of *zui* presupposes that the size of the comparison group is equal to or greater than three. The Chinese data also shows that existential comparatives differ from superlatives, whereas universal comparatives overlap with superlatives. Bare adjectivals in Chinese have both the existential and the universal comparative meanings available. On the surface, it appears to be the case that using bare adjectivals, without composing the extra superlative morpheme *zui*, is the most efficient strategy in a given context. I will give full-fledged discussions about this intuition in chapters 6–8.

Question under discussion Last, but not the least, it’s worth highlighting that whatever surfaces is lining up with the question under discussion (QUD). In a context, two speakers are talking about two individuals Anna and Bob, who are not standing right in front of them. And they are both above average tall. Bob is taller than Anna. The conversation below is *possible*. The fact that (15) is a natural dialogue shows that the negation of a COMP reading is possible, when the QUD is specified as COMP.

- (15)
- *Speaker1: Shui gao?*
who tall
who is taller
 - *Speaker2: Anna gao.*
Anna tall
Anna is taller.

- *Speaker1*: Bu, Anna bu gao. Bob gao.
No, Anna NEGATION tall. Bob tall
No, Anna is not taller. Bob is taller.

On the other hand, when the QUD is specified as *pos*, the *pos* interpretation will be available for a bare adjective utterance such as *gao*. Consider a context where two speakers are talking about Anna and Bob. One speaker knows that Bob is a basketball player. Bob is above average tall, but Anna is not.

- (16)
- *Anna gao ma?*
Anna tall PARTICLE
Is Anna tall?
 - *Bu, Anna bu gao. Bob gao.*
No, Anna NEGATION tall. Bob tall
No, Anna is not tall. Bob is tall.

The fact that a dialogue such as (16) sounds natural confirms that when the QUD is specified as POS, *gao* can have a (negated) POS reading under negation. Note that different from “Anna *gao*”, “Anna *hen gao*” is not sensitive to the QUD in the same way. “Anna *gao*” can be pushed to either POS or COMP, depending on the QUD. But “Anna *hen gao*” cannot be pushed to COMP, no matter what the QUD is. For instance, when the QUD is COMP, as in (15), the same response from Speaker2 with an extra morpheme *hen* (i.e., “Anna *hen gao*”) would implicate that Anna is necessarily tall but she is possibly *not taller* than Bob. Such inference is not available for bare adjectives in (15) Speaker2’s response. Overall, the data point to a QUD sensitive analysis.

3.3 Empirical assumptions - Quantitative data

Spelling out the potential truth conditions of a predicate is a delicate task. On top of that, judgments on an utterance with multiple interpretations can be really slippery. In order to make the theoretical claim more convincing, besides introspective data, this study provides experimental data. Specifically, participants judge whether or not a sentence can truly describe a picture taken to represent a context.

3.3.1 Method

Participants Thirty-three Chinese speakers were recruited on Prolific, an online recruitment platform for web-based experiments. They were compensated \$3 each for their participation. Prescreening was conducted. Participants were filtered for the following features: first language, nationality, country of residence, and previous participation (those who participated in the old pilot studies were excluded). All participants self-declared that their first language is Mandarin Chinese, their nationality is China, and they did not participate in the previous pilots. It is precommitted that participants who finished the whole experiment in fewer than four minutes, or those who didn't pass the attention check would get excluded (the accuracy threshold is preset as 80%). This applied to three participants. Thirty participants' average age is 25.931 and the female to male ratio is 16:14.¹

Procedure Participants were tested online. PsychoPy v3 online (Peirce et al., 2019) was used throughout the whole experiment to display one set of stimuli each time. At the beginning of the experiment, there is a brief Graphic User Interface (GUI) box logging their demographic information. In the instruction session, they were instructed that their task is to judge whether or not the utterance sentence can truly describe a picture given a particular context. In the practice session, participants were given a training trial. This helps them familiarize themselves with the concept of “ambiguity”. After a 500 ms blank screen, the experiment starts once the participants press space-bar on their own keyboard. Desktop or laptop is required for this experiment.

During the experiment, a context sentence is displayed, right below it is a picture, and right below the picture is the utterance sentence. The options are binary. Participants were asked to click on *keyi* “can” or *bu.keyi* “cannot”, where 1 is coded as positive (the utterance can truly describe the picture) and 0 as negative (the utterance cannot truly describe the picture). By pressing the space-bar, they move on to the next trial. Participants did not receive any feedback during the experiment. Each trial is preceded by a 500 ms blank screen. Participants' answers as well as their response times were recorded on each trial. The whole experiment takes around 7 to 10 minutes.

¹ All materials and results can be found at the GitHub repo <https://github.com/yancong222/scripts/tree/main/PsychoPy3%20experiments>

Design The experimental design involved two fully crossed factors: LINGUISTIC ENVIRONMENTS and SCENARIOS. The two levels of linguistic environments factors were: upward and downward entailing environments. For upward entailing environments, the adjective is in plain form. The same adjective shows up in downward entailing environments: negation “not” *bu*, “only” *zhi.you* (literally “only.have”), and “nobody/nothing” *mei.you-yi-CLASSIFIER* (literally “NEGATION.have-one-classifier”). The scenarios factor refers to the state of affairs, which consists of a sentence called “context” and a picture called “referent”. The context sentence introduces the relevant information under discussion, and the referent picture illustrates a possible meaning of the “utterance” sentence. The combination of a context sentence and a picture gives rise to four levels of scenarios: (i) two for target (true on POS false on COMP; and the reversed); (ii) two for control (true on both POS and COMP; and false on both POS and COMP).

The control items are predicted to be necessarily true or necessarily false, thus there are both “can” and “cannot” in participants’ answers; whereas the target items are predicted to be possibly true, therefore “can” is predicted to be the (significant) majority in participants’ answers.

The study chooses four evaluative adjectives from a list of monosyllabic gradable adjectives in Chinese (*gui* “expensive”, *gao* “tall”, *da* “big/spacious”, *kuai* “fast”). Each of them pairs with the two factors (scenarios and linguistics environments). In order to control the lexical item variable, thirty participants were divided into three groups. Each group consists of ten participants. The first group saw the combination of plain form adjectives stimuli and those with the same adjectives but in a negation linguistic environment. The second group saw the combination of stimuli involving plain form adjectives and those involving the same adjectives in an “only” linguistic environment. The third group saw the combination of stimuli involving plain adjectives and those involving the same adjectives in a “nobody/nothing” linguistic environment. Each group of participants saw 4 (adjectives) x 4 (scenarios) x 2 (linguistic environments) = 32 experimental items in total.

Stimuli The full set of 32 items was randomized for each participant. This randomized set was preceded by 4 practice items (the same ones, in the same order, for each participant). This is to

ensure that target items did not show up at the very beginning. As to the experimental stimuli, (17a) and (17d) pair with (3.1); while (17b) and (17c) are associated with (3.2).

For instance, given a scenario (17a + 3.1), an utterance such as (18a) in an upward entailing environment would be judged “can truly describe the picture” as long as participants access the positive interpretation, since the red car is above the 7.unit² threshold (POS-true). On the other hand, the prediction would be “cannot truly describe the picture” if participants only access the comparative interpretation, because the red car’s price is the lowest among the three cars in the picture (COMP-false).

(17) Context:

- a. According to the woman in the picture, 7.unit is expensive for a car. She wants to buy a car. [POS-true-COMP-false]
- b. The woman in the picture is comparing the prices of the three cars for sale. She thinks that 50.unit is expensive for a car. [POS-false-COMP-true]
- c. According to the woman in the picture, 15.unit is expensive for a car. She is comparing the prices. [POS-true-COMP-true]
- d. The woman in the picture is comparing the prices. She thinks that 40.unit is expensive for a car. [POS-false-COMP-false]

(18) Utterance:

- a. (*Zhe san.liang.che li,*) *Hong.che gui.* (**plain form**)
These 3-CLASSIFIER-car inside, Red.car expensive
- b. (*Zhe san.liang.che li,*) *Hong.che bu gui.* (**negation**)
These 3-CLASSIFIER-car inside, Red.car NEGATION expensive
- c. (*Zhe san.liang.che li,*) *Mei.you yi.liang.che gui.* (**nothing**)
These 3-CLASSIFIER-car inside, NEG.have one-CLASSIFIER-car expensive
- d. (*Zhe san.liang.che li,*) *Zhi.you hong.che gui.* (**only**)
These 3-CLASSIFIER-car inside, Only.have red.car expensive

²The unit morpheme used in this experiment is *wan*. Literally, 1 *wan* is equivalent to 10 thousand.

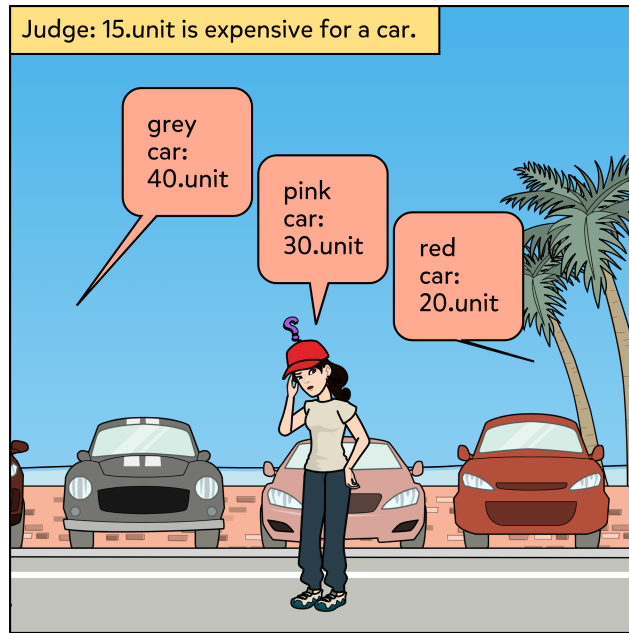


Figure 3.1: *gui*("expensive")_{pos}

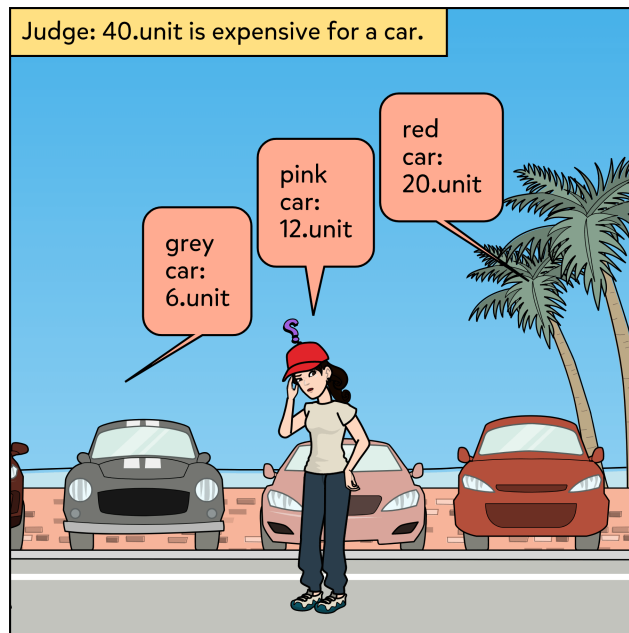


Figure 3.2: *gui*("expensive")_{comp}

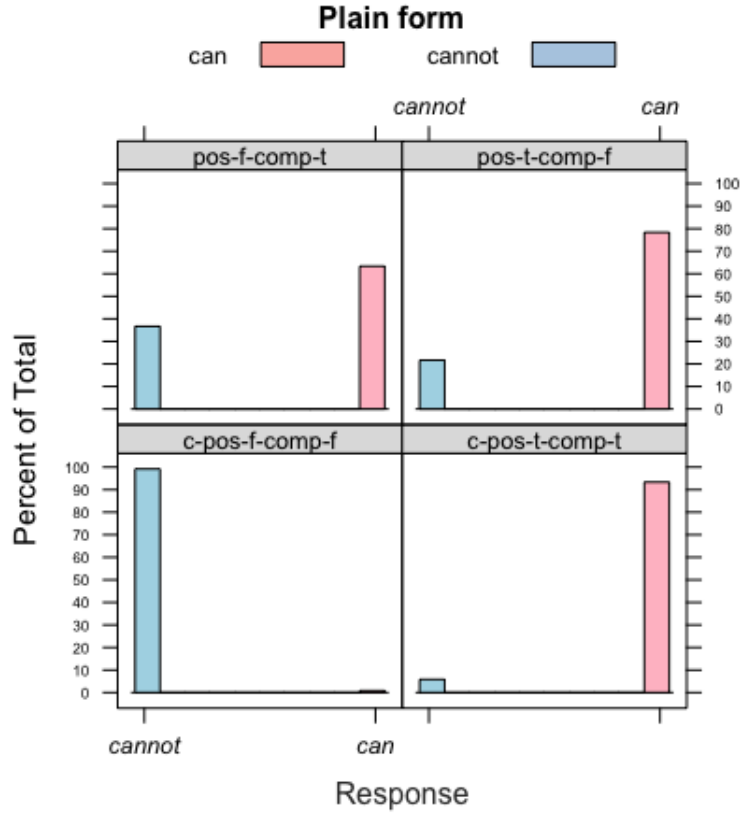


Figure 3.3: Truth value judgment survey on plain form

3.3.2 Results

Figure 3.3 reports participants' responses about the upward entailing environment, where the utterance sentence contains an adjective in plain form. The top two cells represent the target items' responses, while the bottom two summarize the responses to the controls, where "c-" means control.

Compared with the controls, judgments on targets suggest salient ambiguity. Specifically, the upper right cell shows that 78.33% of the thirty participants chose "the utterance *red.car expensive* can truly describe the picture" given a scenario where the red car is above the threshold (POS true) and it's cheaper than the other cars in the picture (COMP false). The upper left cell shows that 63.33% people chose "the utterance *red.car expensive* can truly describe the picture" given a scenario where the red car is below the threshold (POS false) and the red car is more expensive than the other two

(COMP true).

Participants' responses to control items are much more binary. The bottom right cell shows that 94.12% people chose "the utterance *red.car expensive* can truly describe the picture" given a scenario where the red car is above threshold (POS true) and the other two cars are not, namely the red car is the most expensive among all three cars in the picture (COMP true). The bottom left cell shows that 0.83% people chose "the utterance *red.car expensive* can truly describe the picture" given a scenario where the red car is below the threshold (POS false) and it's the cheapest among the three cars displayed in the picture (COMP false).

It's worth mentioning the difference in participants' responses to a positive scenario (upper right) and those to a comparative scenario (upper left). On the one hand, the top two cells do indicate that both the positive and the comparative interpretations *are* available in upward entailing environments. On the other hand, it seems that the positive reading is easier to access, considering that responses to the comparative scenario are close to a split.

Figure 3.4 summarizes participant's responses to downward entailing environments. The top two rows represent responses to the targets and the other two represent responses to the controls. The first (leftmost) column, where the adjective is modified by negation *bu*, suggests that there is *some* amount of ambiguity, though not as salient as the upward entailing environment.

Specifically, the upper left cell shows that 47.5% chose "the utterance *red.car not expensive* can truly describe the picture" given the scenario where the red car is above the threshold and it's cheaper than the other two cars in the picture. The second cell in the first column shows that 82.5% chose "the utterance *red.car not expensive* can truly describe the picture" given the scenario where the red car is below the threshold and it's more expensive than the other two cars in the picture. Comparison of those two cells leads to the conclusion that the positive interpretation is more salient than the comparative. In the meantime, there is a clear-cut distribution with respect to participants' responses to the controls: 5% said *can* to a false control and 100% said *cannot* to a true control.

Regarding the other two columns, where quantifiers *mei.you* "nothing" and *zhi.you* "only" are involved, there appears to be no ambiguity on the surface. For the "nothing" column, 2.5% people

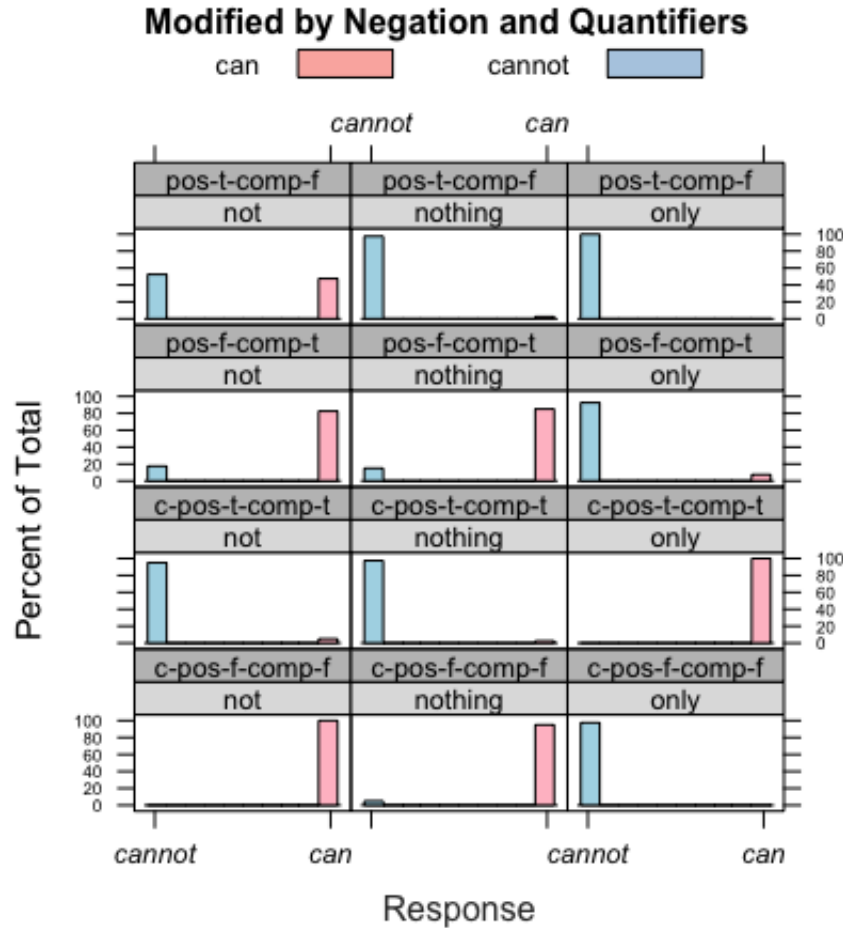


Figure 3.4: Truth value judgment survey on modified form

chose “the utterance *no.car expensive* can truly describe the picture” given the scenario where all three cars are above average expensive (pos true) and they are of equal prices (comp false). If participants interpreted it as comparative, they should choose “can”. This is because in that context, it’s not the case that there is a car that is more expensive than the others.

Since the significant majority didn’t, one possible explanation is that they access the positive reading. This also supports the theoretical claim that pos is more salient than comp in downward entailing environments. A similar pattern is found in the second cell, where 95% people chose “the utterance *no.car expensive* can truly describe the picture” given the scenario where all cars are below the threshold (pos false) and they are of different prices, meaning there exists one car such

	<i>gao</i>	<i>hen gao</i>
upward entailing environment	POS; COMP	POS
downward entailing environment	<u>POS</u> ; COMP	POS

Table 3.1: A more detailed landscape of the data: summary

that its price is more expensive than the other two (COMP true). If participants interpreted it as COMP, then the majority should go for “cannot”. Since they chose “can”, it indicates that the salient reading is POS but not COMP.

As to the “only” column (rightmost), in the first cell, 0% chose “the utterance *only red.car expensive* can truly describe the picture” given the scenario where all three cars are above the threshold (POS true) yet the red car is cheaper than the other two (COMP false). In the second cell, 7.5% chose “the utterance *only red.car expensive* can truly describe the scenario” given the scenario where all cars are below the threshold (POS false) and the red car is more expensive than the other two (COMP true). This indicates that the majority did not interpret it as comparative, but interpreted it as positive. Again, this provides evidence that the positive interpretation is salient in downward entailing environments.

3.3.3 Discussion

There is strong evidence that “gao” is ambiguous in upward-entailing sentences, some weaker evidence that it’s ambiguous in downward-entailing sentences, and that there’s a preference for the positive reading in downward-entailing sentences, as illustrated by the wave line in table (3.1). When the QUD is specified, the ambiguous utterance “gao” gets disambiguated relative to the QUD. The POS interpretation of the utterance “gao” is available when the QUD is POS; and the COMP is available when the QUD is COMP.

One theoretical claim that does not get quantitative support is COMP being salient in upward entailing environments. Plausibly, the reason why COMP interpretation does not surface has to do

with the experiment presentation mode. Participants first read a context sentence where a “standard” is explicitly given, then observe the follow-up picture. The sequence of stimuli presentation may bias them toward the positive interpretation, which involves a contextual standard.

Why is the positive reading so salient in downward entailing environments, especially when quantifiers “nothing” and “only” are involved? One likely explanation is the competition of other frequently used alternatives, including equatives and superlatives. They are more informative and specific than the comparative. Importantly, in the meantime, participants have no problems accepting the true controls or rejecting the false controls, as shown in the lower two rows. This dissertation acknowledges that *hen* expressions have not been included in the stimuli for now, neither have the equatives or the superlatives. This dissertation leaves competition of various alternatives for a separate experiment.

The other possible explanation for why the comparative reading doesn’t surface in downward entailing environments is that negating a comparative is logically implausible. Suppose the logic space of a comparison relation can be divided into three cells: *greater than*, *smaller than*, and *equal to*, negating *greater than* or negating *smaller than* does not output an *unique* cell. Concretely, negating the comparison relation *greater than* gives rise to two possible cells: *smaller than* and *equal to*, which is not informative. Similarly, negating *smaller than* leads to more than one logically plausible cell: *greater than* and *equal to*, which results in inefficient communication.

To sum up, this section starts off introspective judgments, which seems to suggest that “gao” means COMP and “hen gao” POS, but with some evidence (ellipsis) that a positive reading of “gao” can be accessed. The experimental data shows a more complex picture. The overall take-home message is that: (i) “gao” definitely can have the POS interpretation, namely the ambiguity is real; (ii) linguistic environment matters.

3.4 Chapter summary

This chapter provides a more detailed landscape of the data, given that Chapter 2 shows that there is a lot of debate going on in terms of how to interpret Chinese degree expressions. In addition

to the ongoing debate about the Chinese *hen* puzzle, Chapter 2 suggests that most of the previous studies take the semantics-syntax approach, whereas relatively few investigations are focused on the role of language users and contexts. Chapter 3 therefore designed a truth-value judgment survey. This chapter has clarified what interpretations are available for *gao* “tall”, and has provided a way to show that linguistic environment does play a role in natural language understanding.

Overall, this chapter provides quantitative empirical evidence that plain form adjectivals in Chinese have more than one interpretation available, and the surfaced reading is sensitive to linguistic context. These are hallmarks of competition-based pragmatic analysis. Now let’s turn to Chapter 4 and see how to implement such an analysis.

CHAPTER 4

COMPETITION: PREVIOUS STUDIES AND IMPLEMENTATION

By carefully reviewing and implementing previous accounts of competition, this chapter shows that previous studies on competition cannot be directly transferred to the degree expressions puzzle here. Moreover, a close scrutiny of the implementations helps reveal the distinct features of non-scalar implicatures.

4.1 Competition of utterances

The fact that “Anna gao” is systematically ambiguous between two readings (pos and comp), and in some cases only one reading not the other surfaces, calls to mind a pragmatic explanation involving competition. Available degree interpretations are conditioned by various contexts. That seems very suggestive of what we have seen in other areas of pragmatics. Competition between possible utterances (so-called alternatives) leads to pragmatic enrichment of interpretation. One interpretation surfaces in some cases, and in another cases, another interpretation surfaces. Thus, it follows naturally that this phenomenon calls for a pragmatic story. However, that line of analysis regarding pragmatic competitions does not come without challenges. This section extends competition models to the Chinese degree puzzle, and shows that competition models are not directly applicable, since they either over-generate inferences that are not true, or they assume a logical relation that does not exist.

4.1.1 Gricean pragmatics and (Scalar) Implicature

Competition between utterances is called implicature in the literature. Grice (1975) attributes to language users the recognition of particular rational, cooperative efforts in their attempts to achieve “a common purpose or set of purposes, or at least a mutually accepted direction” (Grice 1975:45). This assumption is encoded in the general principle of conversation, the Cooperative Principle (Grice 1975:46):

(1) Cooperative Principle

Make your conversational contribution such as it is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.

Grice developed maxims to explain what he took to be a tendency of conversational participants to strive for “a maximally effective exchange of information” (Grice 1975:47). Gricean maxims can be loosely arranged into four rubrics:

- Quality: Try to make your contribution one that is true:
 1. Do not say what you believe to be false.
 2. Do not say that for which you lack adequate evidence.
- Quantity:
 1. Make your contribution as informative as is required (for the current purposes of the exchange).
 2. Do not make your contribution more informative than is required.
- Relation: be relevant.
- Manner: Be perspicuous:
 1. Avoid obscurity of expression.
 2. Avoid ambiguity.
 3. Be brief (avoid unnecessary prolixity).
 4. Be orderly.

The maxims are generally followed but are occasionally flouted. Flouting a maxim allows a speaker to exploit these principles of communication and thereby conversationally implicate something. And that’s called conversational implicature in the literature: it’s “what has to be supposed in order to preserve the supposition that the Cooperative Principle is being observed”

(Grice 1975/1989: 39-40). ϕ is a conversational implicature licensed by the speaker's utterance of ψ if ψ is not conventionally associated with ϕ , and the hearer has good reasons for believing that, assuming the speaker intends to be cooperative, he wouldn't have uttered ϕ unless ψ was the case.

This paper deals with two expressions, “gao” and “hen gao”, that appear to overlap in meaning, but where one of them (“gao”) routinely gets enriched¹. This is reminiscent of the classic (Scalar) Implicature cases like the “or” “and” alternation, where “or” is routinely enriched. Let's see if this works.

When it comes to scalar implicature, an utterance “S” implicates not-S' for some alternative utterance S'. For instance, the literal meaning of “some” is “at least some”, it competes with the utterance “all”. A language speaker can understand the interpretation of “some” as a combination of the basic meaning “at least some”, plus the pragmatic inference “not all”. In the case of “gao”, however, there is no such negative inference: either we get COMP or we get POS; we never get COMP + not-POS, or POS + not-COMP. Concretely, *Anna gao* “Anna tall” never means “Anna is taller than everyone else salient in the context” AND “she is *not* above average tall”. That's how the classic Quantity implicature studies fail conceptually. Now let's do a step-wise belief-based implicature calculation, so that it further informs us on where it fails formally.

Belief-based Standard Recipe is mostly used for deriving Quantity implicatures in the literature (Geurts, 2010). Anna says (2), which may implicate that Bob didn't eat all of them. This implicature is explained by assuming that the hearer reasons, and is entitled to reason as in (3).

(2) Bob ate some of the cookies.

- (3)
1. Rather than uttering (2), Anna could have said (2*) Bob ate all the cookies, which is logically stronger/more informative. Why didn't she do so?
 2. The most likely explanation is that Anna doesn't *believe* that (2*) is true: $\neg \text{BEL}_A(2^*)$.
 3. Anna is likely to have an opinion as to whether (2*) is true: $\text{BEL}_A(2^*) \vee \text{BEL}_A(\neg 2^*)$

¹For consistency purpose, this dissertation takes *hen gao* as two words. Similarly for other alternatives including *bi gao*, *geng gao*, *bijiao gao*, this dissertation analyzes each alternative as two words. This could be important for structural accounts of alternatives.

4. Between them, (2) and (3) entail $BEL_A((\neg 2^*))$: Anna believes that Bob didn't eat all the cookies.

Now suppose a speaker Kai uttered: “*Anna gao*”. The semantic assumption is that bare adjectives such as *gao* is two-way ambiguous: above average tall (POS) and taller than contextually salient people (COMP). With degree morphemes, utterance such as *hen gao* only means POS.

(4) Belief-based Gricean Reasoning in a plain upward entailing context

1. Rather than uttering “*Anna gao*”, Kai could have used the alternative utterance “*Anna hen gao*”, which is not ambiguous. But Kai did not. Why not?
2. Presumably, Kai doesn't *believe* that “*Anna hen gao*” is true: $\neg BEL_{POS}(\text{“Anna hen gao”})$.
3. Kai is likely to have an opinion as to whether (“*Anna hen gao*”) is true: $BEL_{POS}(\text{“Anna hen gao”}) \vee BEL_{POS}(\neg(\text{“Anna hen gao”}))$
4. Between them, (2) and (3) entail $BEL_A((\neg(\text{“Anna hen gao”})))$: Kai believes that Anna is not above average tall.

Step (4) spits out a wrong inference that Anna believes that the POS interpretation is false, which is exactly what's *not* supported by the empirical observation. We saw that treating “*gao*” versus “*hen gao*” as a case of Scalar Implicature doesn't work, but maybe “*gao*” is nevertheless enriched by competition of a different kind, for example competition that leads to an “atypicality” inference (rather than negative scalar inference), which has been extensively studied in Rett (2014a) and Rett (2019). Let's see if this works.

4.1.2 Non-scalar Implicature

Conceptually what Rett (2014;2019) is trying to do is to explain why adjectives such as “tall” has a particular interpretation, namely the *evaluative*². Rett attempts to do that on the basis of the Maxim

²Note that here “evaluative” is not completely equivalent to what we called POS in this dissertation. According to Rett (2014), who assumes Cresswell (1976), POS is a null operator that both *contributes*

of Manner, as illustrated in (5).

(5) Manner

1. Avoid obscurity of expression.
2. Avoid ambiguity.
3. Be brief (avoid unnecessary prolixity).
4. Be orderly.

(6) is an example given by Grice to illustrate how the Manner maxim may be exploited for inducing conversational implicatures. Here, the speaker goes out of their way to avoid the verb “sing”, using a considerable more prolix expression in its stead. The speaker would not do so unless he believed, and WANTED to convey to their audience that the the predicate “sing” was not entirely appropriate, presumably because Miss X’s singing wasn’t very good.

- (6) Miss X produced a series of sounds that corresponded closely with the score of “Home sweet home”.

Manner maxim in (5) would have compelled the speakers to use the simpler expression. This reasoning algorithm is very reminiscent of Quantity reasoning: cooperative speakers say as much as they believe. Except that it’s not about speaker’s belief state. The Quality and Quantity maxim have a lot to do with speaker’s belief. By contrast, Manner maxim concerns reasoning about speaker’s intention. That’s another reason why Manner implicature studies can better explain the degree puzzles here, compared with Scalar implicature studies.

Rett’s proposal in a nutshell Rett (2008a) and Rett (2014a) defines evaluative as the following: a sentence is *evaluative* with respect to a gradable adjective \mathcal{G} iff it entails that some individual instantiate \mathcal{G} to a significant degree. She argues that evaluativity arises where it does as the result of Manner implicature. To account for this relatively wide distribution of evaluativity, Rett (2007, “*evaluativity*” and values the degree argument.

2008) formalizes *eval* in (7), which can contribute evaluativity to an adjectival construction without manipulating a degree argument. Its domain is a degree property. It's applied after the individual argument of a gradable adjective is saturated. And its range is an evaluative degree property. *eval* is a modifier version of *pos*, according to Rett.

- (7) *eval* is a degree modifier; it denotes a relation between sets of degrees *D*, type $\langle d, t \rangle$.

$$\llbracket \text{eval} \rrbracket = \lambda D_{\langle d, t \rangle} \lambda d. D(d) \wedge d > s, \text{ for some contextual standard } s.$$

According to Rett, (7) means that for “Bob is *eval* tall”, we get back a set of degrees (*D*), rather than a truth value. To get a truth value, we existentially quantify over that set of degrees: *there is* a degree *d* such that Bob is (at least) *d*-tall and *d* > *s* for some contextual standard *s*.

A crucial ingredient in Rett's semantics of *eval* concerns *M*-alternatives. Inspired by Horn (1972) and Katzir (2007), Rett defines *M*(anner)-alternatives and the *M* Principle as below:

- (8) *M*-ALTERNATIVES: Let ϕ be a parse tree. The set of **M-alternatives** for ϕ , written as $A_M(\phi)$, is defined as $A_M(\phi) := \{\phi' : \llbracket \phi' \rrbracket \longleftrightarrow \llbracket \phi \rrbracket\}$.

The arrows in (8) mean semantically equivalent. In plain English: the set of *M*-alternatives for a parse tree (regardless of complexity), is the set of other parse trees with the same meaning. For example, “Alice likes Bob.” denotes a proposition. It also has a parse tree, where Alice is the subject and “likes Bob” is VP. One of the *M*-alternatives of “Alice likes Bob” is the passives “Bob is liked by Alice”. The two expressions have the same semantics but different parse trees, i.e. different syntax. All parse trees that have the same semantics are in that set.

- (9) *THE M PRINCIPLE*: Don't use ϕ if there is another $\phi' \in A_Q(\phi)$ st. ϕ' is assertable and $\phi' < \phi$.

The *M*-principle says: do not use ϕ if there is another semantic equivalent sentence that is assertable, shorter, and less complex. Rett further formalizes Horn's Principle of Least Effort (“Marked forms are associated with marked meaning”) as below:

- (10) *The Marked Meaning Principle*: For parse trees ϕ, ϕ' such that $\phi' \in A_M(\phi)$ and $\phi' < \phi$: ϕ carries the Manner implicature: “ $\llbracket \phi \rrbracket$ is atypical”.

Put otherwise, the Marked Meaning principle states that if the speaker uses a more complex sentence with equivalent semantics, then the meaning is atypical. The above accounts of Manner implicatures do not seem to get readily applied to the Chinese data. Rett proposes that the diagnostic for evaluativity is that an evaluative sentence entails the negation of its antonymic counterpart. *Jane is tall* is evaluative because it entails the sentence *Jane is not short*. Interestingly, (11a) does not *always* entail its negative antonym (11b). In every possible world where (11a) is true, it doesn't necessarily follow that (11b) is true. Under a comparative context where three individuals Anna, Bob and Al are all below average tall, and Anna is taller than Bob and Al, the utterance (11a) is judged true, whereas (11b) is judged false. On the other hand, (12a) does *always* entail (12b). (12b) is true in every possible world where (12a) is true. It appears that the addition of *hen* gives rise to the evaluativity.

(11) a. *Anna gao.*
Anna tall

b. *Anna bu ai.*
Anna NEGATION short

(12) a. *Anna hen gao.*
Anna HEN tall

b. *Anna bu ai.*
Anna NEGATION short

Moreover, Rett assumes that positive constructions qualify as tautologies (“degree tautologies”). She argues that positive constructions are constructions whose scalar or degree semantics is trivially true, and evaluativity is the (only) natural consequence of this meaning being strengthened in context via a Quantity implicature. If positive constructions (11a) and (12a) are degree tautology, then the negation of tautology is contradiction, namely “has no tallness”. But negating (11a) is negating “taller than contextually salient people” or negating “above average tall”. And similarly, the consequence of negating (12a) is the negation of “above average tall”. Neither of those two negated degree constructions leads to contradiction. So far it appears that, in general, Rett’s analysis about positive constructions cannot get transferred to Chinese directly. Specifically, when Rett’s markedness based evaluativity implicature gets extended to the Chinese data, there are two possible hypotheses, both of which run into problems.

Hypothesis I Hypothesis I assumes that *gao* and *hen gao* are synonymous, and neither of which is ambiguous, which is not empirically supported. Why consider this hypothesis, if it's not empirically supported in the first place? The reason is that I'm giving a (potential) *theoretical* analysis, which may involve intuitively dubious assumptions, but is nevertheless worth exploring *if* it can output the right predictions in the end. Essentially, it's still worth to be explicit about these directions.

Hypothesis I claims that *gao* and its synonym *hen gao* start with one interpretation - comparative, which is presumably the typical reading. They are markedness scale mates: they are of the same semantics, i.e. comparatives, and the parse tree of *gao* is less marked than that of *hen gao* due to the lack of the degree morpheme *hen*. Therefore when the M-alternative *hen gao* is uttered, it gives rise to atypicality implicature, namely the atypical meaning - positives. A step-wise illustration of Hypothesis I is given below.

Step-wise:

1. $\llbracket gao \rrbracket = \llbracket hengao \rrbracket = \text{COMP}$
2. Markedness scale: $\langle gao, \text{hen gao} \rangle$
3. “hen gao” \leadsto POS (M-implies - atypicality)

Hypothesis I assumes *hen gao* means COMP (step 1), which is already going against the empirical claim established in section 2 that *hen* never means COMP. Furthermore, one may wonder that given that the notion of “atypicality” is widely distributed, why is it lining up with POS in Hypothesis I step 3? According to Rett, the property “atypical” in (10) is one way of characterizing Horn’s description of marked forms as “conveying a marked message” (Horn 1989:22), or as being associated with a marked situation. The Marked Meaning Principle in (10) defines marked meaning relative to the markedness alternatives. The property ATYPICAL is thus essentially shorthand for “*atypical* given what the speaker *could have* said”. Rett concludes that when an adjective or degree construction is associated with an implicature that strengthens its meaning, it is strengthened from a typical extension to an atypical one. Suppose the typical reading of *hen gao* is COMP, the *atypical* meaning doesn’t have to be POS. Given what the speaker could have said, an *atypical* message conveyed

by a more marked alternative *hen gao* could be any one of the following: the absence of *direct* knowledge, uncertainty, or mild intensification, etc.. Some of these have been independently argued in Fang (2016). Why does “*hen gao*” have to be associated with an implicature that strengthens its meaning to an atypical extension pos, but not say, directness? Independent evidence is needed in order to get other logical possibilities excluded.

Besides, Hypothesis I wrongly predicts COMP to be the surfaced interpretation in upward and downward entailing environments, without making any distinctions. Throughout the derivations, the speech act of uttering “*gao*” is categorically interpreted as COMP, just as started (step 1). This is not in line with the empirical observation made in the previous section (section 2).

Hypothesis II Similarly, Hypothesis II assumes the two alternatives are synonymous. It claims that the two expressions start with positive, and *hen gao* is less marked than *gao* on the markedness scale. Thus when the more marked M-alternative *gao* is used, it M-implies an atypical meaning, namely the comparatives. A step-wise illustration of Hypothesis II is given below.

Step-wise:

1. $\llbracket gao \rrbracket = \llbracket hen\ gao \rrbracket = \text{POS}$
2. Markedness scale: $\langle hen\ gao, gao \rangle$
3. “*gao*” \rightsquigarrow COMP (M-implies - atypicality)

For Hypothesis II, “*gao*” starts with pos (step 1) and gets strengthened into COMP with the atypicality implicature. Yet the problem is that it’s relying on structural complexity and the theory of M-alternatives. A counterintuitive assumption has to be made: *hen gao* is considered to be less *complex* than *gao*, despite the presence of an additional degree morpheme *hen*. What’s worse is that Rett doesn’t really talk about ambiguity. It’s unclear exactly why COMP is *the* “atypical” interpretation we get and not something else, for example information source and information force. Rett’s proposal focuses on two structures, *tall* and *not short*, which have only one meaning, and they happen to have that exact same meaning. Then the more complex one gives rise to an atypicality

implicature. Rett (2015) devotes a big amount of discussion to antonyms but very few to negations and downward entailment environments in general. She has a sophisticated story about *tall* and its antonym *short*, but that doesn't seem to get readily transferred to *tall* versus *not tall*. It's thus hard to draw a consistent conclusion from any kinds of extensions illustrated in Hypotheses I and II.

4.2 Competition of interpretations

So far this dissertation has talked about two theories about pragmatic enrichment of utterances: the classic scalar implicature and markedness based Manner implicature. They run into various problems when applied to the degree semantics puzzles in Chinese, mainly because those two models are not designed to tackle the issues about competition of interpretations. Therefore another relevant direction of analysis concerns pragmatic enrichment of interpretations. This section is devoted to such a competition model called strongest meaning hypothesis (SMH) (Dalrymple et al., 1994, 1998).

Instead of assuming competition between *utterances* (“gao” versus “hen gao”), maybe we can capitalize on competition between *interpretations*, given that “gao” is (taken to be) ambiguous (POS vs COMP), which is reminiscent of the SMH. Let's see if this works.

4.2.1 True on the strongest meaning

Studies about the strongest meaning hypothesis deal with ambiguities of reciprocal expressions. The basic paradigm is illustrated below.

- (13) a. The girls know each other.
 - b. ...# but Mary doesn't know Sue
 - c. Every girl knows every other girl
- (14) a. The girls are hugging each other
 - b. ... but Mary is not hugging Sue
 - c. # Every girl is hugging every other girl

(13a) is uttered in a context where there are Anna, Sue, Alice and Mary all four girls, and they know each other. A salient inference is that Anna knows ALL the other three girls, namely the universal reciprocal interpretation. That is why (13c) is strongly preferred to (13b). The point is that (13b) is an odd follow-up, which indicates that the meaning of (13a) can be paraphrased as in (13c).

But for *hug* in (14a), the preferred reading is the pairwise one (14b) but not universal (14c). The point is that, now, (14b) is *not* odd, which indicates that the meaning of (14a) is *not* paraphrasable as in (14c) – it’s something weaker.

How come in (13a) you have this set of universal relation, whereas “hugging” in (14a) only has the pairwise interpretation? The proposal of Dalrymple et al. (1994) and Dalrymple et al. (1998) is an account of such contrasts. They argue that there is a principle in semantics and pragmatics interpretation: as a language speaker, we pick the logically strongest interpretation among all the alternatives, which is still consistent with what we know about the world and the context. Thus in the case of knowing, the strongest possible reading is the universal one. But for hugging, it is difficult to have universal hugging at the same time (you can only hug one, maybe two people at a time). In upward entailing environments, the strongest reading is the universal one. By contrast, in a negated context, the strongest reading ends up the pairwise reading due to the negation. As shown in (15), the meanings are that there is no such a pair of girls who know each other, or are hugging each other. They arise because in downward entailing environments, the pairwise reading is the strongest one, according to the SMH.

- (15) a. The girls do not know each other.
b. The girls are not hugging each other.

The SMH model can be implemented to explain the Chinese data. It appears to be able to capture a phenomenon that you have a sentence that is ambiguous between reading 1 and reading 2, then the preferred interpretation is different under different context or with the addition of an entailment-reversing expression such as negation. SMH model argues that the reason why the preferred reading changes is due to the change of the environment. This seems to be applicable to the Mandarin degree data. An utterance like *Anna gao* “Anna tall” is ambiguous between positive

and comparative reading. SMH would predict that the language speaker always try to assign the strongest reading that one can. Therefore in upward entailing environments, that might be the comparative reading, and in downward entailing environments, the positive one is the strongest.

It turns out that SMH can't be readily transferred to Chinese. The two readings of bare adjectival expressions in Chinese are logically independent of each other, as shown in the empirical chapter (Chapter 3). To restate it here: the fact that X is taller than average is independent from whether X is taller than everyone else in context (those two things can both be true, both be false, or have opposite truth values). There is no way to actually go further and pick out the so called "strongest" meaning for disambiguation.

4.2.2 True on all meanings

Speaking of the Strongest Meaning Hypothesis, there is a more general and recent version of SMH: Spector (2017)'s disambiguation model. Conceptually, the main idea is that, if it's true on all interpretations then it's true on the strongest one. The SMH essentially states that if something is ambiguous between two readings, then the one that surfaces is the strongest one. Spector argues that if an utterance is ambiguous then it's interpreted as a CONJUNCTION OF all of its readings. In other words, this ambiguous utterance is interpreted as BEING TRUE ON all of its readings. Spector's theory of plurals says that sentences with plurals have multiple different readings, and the sentence is true on all interpretations.

Without going through the technical details, one may notice that this idea is not going to solve the degree semantics puzzle in Chinese. If we assume that *Anna gao* is ambiguous, then the prediction would be that Anna is both above average tall and she is taller than everyone salient in the context at the same time. It's true on all readings, meaning that it's having both a true POS interpretation and a true COMP. That is not supported by the empirical data. Uttering "Anna gao" out of the blue means COMP, and it doesn't simultaneously license a negated or an affirmed POS interpretation.

However, Spector (2017) is worth a careful review for the following reasons. The proposal intends to address the puzzle that if I were a rational speaker, why I would choose to use the

expression that is ambiguous if I knew that there are some uncertainty regarding how to interpret it. The answer is that I use whichever is true no matter which interpretation is picked. That's why we get true on *all* readings. To address the puzzle about the Chinese degree semantics phenomenon, some sort of uncertainty concerning how the listener is going to interpret the utterance is needed.

Furthermore, the way that Spector (2017) formalizes QUD seems insightful. The proposal is that a sentence S is judged true in the context of a QUD Q only if all its relevant candidate meanings relative to Q are true. Recall the QUD data points in section 2 (15, 16). Once the QUD gets established, the utterance *Anna gao* should be able to address it, with the relevant candidate meaning. When the QUD concerns pos, then the same response *Anna gao* is predicted to mean pos. This hypothesis is justified in section 2.

4.3 More about the intention based view

Geurts (2010) provides an extensive discussion about an intention based Recipe, and how it can explain free choice inference, as in sentence (16) (c.f. Geurts:2010:105).

(16) A: What's for dessert?

B: You can have fruit or chocolate cake.

After speaker B's utterance in (16), speaker A comes to the conclusion that they can have chocolate cake, namely the so-called "free choice inference". The proposition that x *has fruit or chocolate cake* is entailed by, and is weaker than, the proposition that x *has chocolate cake*.

It appears that speaker A arrives at the conclusion from a premise that is (logically) weaker. As a consequence, their reasoning cannot be valid. Roughly this is the problem of "free choice permission" (von Wright 1968, Kamp 1973). This part first gives a recap of Geurts' proposal about "free choice", followed by an extension of this theory to the Chinese data.

Ingredients of the intention-based method Here are the ingredients: first, a sentence ψ , as uttered by a speaker S ; second, a partitioning i_1, \dots, i_n of S 's possible intentional states. Intentional states can have different flavors, among which some concern belief states, while some don't. For

instance free choice permission is not about S's beliefs. The third ingredient is about alternatives: sentences S could have used instead of ψ , had he wanted to express any of i_1, \dots, i_n . According to Geurts, each of these alternatives should entail ψ , given that presumably S intended to signal their commitment to ψ ,

We now turn to the hearer H, who proceeds to weed out those states in i_1, \dots, i_n that fail to meet certain conditions. According to Geurts, anything H assumes about S's beliefs, hopes, desires, etc. may help to narrow down the space of possible intentional states should attain. Moreover, H can try to weed out candidate intentional states by considering the alternatives associated with them, using the following ALTERNATIVE HEURISTICS:

- (17)
1. If S could have expressed that he is in state i by using a sentence that is no more complex than ψ , then probably S is not in state i .
 2. For any pair of intentional states, i and j , if it had been easier for S to express that he is in i than that he is in j , then S is more likely to be in j than in i .

Epistemic free choice Certain assumptions are presupposed by Geurts (2010) before he turns to free choice examples. First, utterance of ψ is sincere; hence, any alternative must be at least as strong as ψ . Further, there is an inverse correlation of alternative's complexity and the likelihood that the speaker is in the corresponding intentional state. Now, suppose Mildred uttered (18):

- (18) George may be British or Canadian.

Mildred must be in one of the following belief states. Note that " $\text{POSS}_S(\psi)$ " means that S considers it possible that ψ .

- (19)
- a. $i_1: \text{POSS}_M(\text{George is British}) \wedge \text{POSS}_M(\text{George is Canadian})$
 - b. $i_2: \text{POSS}_M(\text{George is British}) \wedge \neg \text{POSS}_M(\text{George is Canadian})$
 - c. $i_3: \neg \text{POSS}_M(\text{George is British}) \wedge \text{POSS}_M(\text{George is Canadian})$
 - d. $i_4: \neg \text{POSS}_M(\text{George is British}) \wedge \neg \text{POSS}_M(\text{George is Canadian})$

The recipe for epistemic free choice is given as follows. Of these states, i_4 can be dismissed straightaway, because it is inconsistent with Mildred's believing (18). For the remaining states, alternatives in (20) would have been available for Mildred to utter. Assuming that Mildred believes what she says, she could have uttered (20a) if she was in belief state i_2 , (20b) if in belief state i_3 , and something along the lines of (20c) if she was in belief state i_1 . Now, the second alternative heuristic in (17) tells us that, since (20a) and (20b) are considerably simpler than (20c), Mildred's intentional state probably is i_1 rather than i_2 or i_3 . Namely, as far as Mildred knows, George may be British and George may be Canadian. Thus, it appears that free choice gets explained.

- (20) a. George may be British.
- b. George may be Canadian.
- c. George may be British and he may be Canadian.

However, a close scrutiny indicates that there is a problem. It's not clear how exactly (20a) conveys (19i2), given that (20a) literally means only $\text{POSS}_M(\text{George is British})$. Presumably, one may need to consider the *enriched* meaning of (20a), but then this starts to become circular: why is (20a) enriched to (19i2)?

To sum up, in the process of weeding out possible intentional states, the hearer may have to take into account the availability of alternatives associated with those states, according to Geurts (2010). Geurts' proposal also implies that availability is a matter of economy, previous use, frequency of use, familiarity, general or particular expectations of specificity.

Why this recipe cannot explain the adjectives data Close scrutiny indicates that this intention-based view can not readily get extended to the Chinese puzzle, for reasons that are related to both Geurts' proposal itself and the empirical observations presupposed in Chapter 3. Specifically, (20a) does not entail i_2 , and (20b) does not entail i_3 . Further, why Mildred has to be in the particular POSS listed above in (19)? All (20a) conveys is that George may be British, and it does not say anything about whether or not he may be Canadian. Third, for utterance *Anna gao*, one does not get any kind of negative inference. The last thing this dissertation would like to argue is that the belief

state is the combination of POS & \neg COMP. If this dissertation were to apply Geurts' proposal, an assumption has to be made: the speaker is in one of those four states in (19), for example, consider (21).

(21) Kai gao.

Kai tall

Presumably the speaker can be in belief state (19i1) POS & COMP (i.e., the conjunction of POS and COMP), suppose we were implementing the “free choice inference” analysis to (21). However, this speaks against the empirical observation provided in Chapter 3. When using (21) (intended POS or intended COMP), the speaker has one lexical intention in mind, yet this intention has no logical relation with the other intention. The two are mutually independent. By contrast, for the conjunction of POS and COMP as in (19i1), if one of the coordinates is false, then the conjunction is false.

i_4 can be dismissed easily. Even if i_1 seems to be close to the intended meaning, it does not disambiguate *gao*. In the meantime, there is no felicitous utterances to express i_2 and i_3 , since *hen gao* just means POS, and it does not mean POS & \neg COMP.

The inference POS & \neg COMP is not desirable, thus it's necessary that Geurts' proposal gets updated and overlapping possible belief states are allowed. Moreover, the assumption that *Anna gao* is ambiguous needs to be modified into one that *Anna gao* is UNDERSPECIFIED. Underspecification here relates to a contextually driven specification of a grammatically well-formed, yet underspecified structure (cf. Hemworth; Bierwisch 1982, 1983; Egg 2005). Speaker uses it when it's in one of these two possible belief states: i_1 POS and i_2 COMP. Crucially, these are overlapping belief states. And they can both be true, or one can be true and the other can be false. Whenever speaker uses *Anna gao*, these are two alternative belief states that language users think about when speaker uses *gao*.

The reasoning process is illustrated as follows: if speaker were in belief state i_1 , then they could have also used *Anna hen gao*. And there were no questions about which belief state they were in, thus they must be in i_2 . It's underspecified with regards to these two belief states. Speaker has a choice to make between these two belief states. In the spirit of Geurts, this proposal also raises the

same issues, which is that why just associate alternative *Anna hen gao* with i_1 , but not alternative (22) with i_2 .

- (22) *Anna bi Bob gao.*
Anna than Bob tall
Anna is taller than Bob. (COMP)

4.4 Chapter summary

Given that Chapter 3 has illustrated an empirical picture that the Chinese adjectival puzzle involves both competition of utterance and competition of interpretation, Chapter 4 thus reviews and implements both aspects. In this chapter, I explored several standard competition-based accounts of interpretation, some of which may have seemed promising, but none of which apply successfully to the Chinese adjective puzzle.

To sum up, here are the findings of this chapter: (i) the Chinese degree expressions puzzle is better off explained as a non-scalar implicature puzzle, meaning that the competing utterances do not have logical-strength relation. Moreover, the competing readings are logically independent of each other. This results in the failure of scalar-implicatures-related accounts; (ii) Rett's Manner implicatures analysis cannot be easily transferred essentially because it goes against the empirical facts collected in Chapter 3, and the implementation of markedness-scalemates requires independent motivation; (iii) non-scalar implicatures concern language users' intention, whereas scalar implicatures concern their belief. Thus, Geurts intention-based recipe is revised and implemented. It turns out that this approach has ran into some challenges. A major challenge is that it generates inference that's not actually observed.

CHAPTER 5

A COMPETITION-BASED DISAMBIGUATION MODEL

As discussed in Chapter 4, non-scalar implicatures concern intention. Further, non-scalar implicature is not based on entailingness. In this chapter, the core proposal of the thesis is laid out. Specifically, the discussion is about how to account for non-scalar implicatures, and the answer to which is in a competition-based disambiguation model. In this dissertation, the focus is on alternatives in non-scalar implicatures, which leads to the symmetry problem observed in the domain of scalar implicatures. Possible solutions and directions are given. Moreover, this chapter proposes that alternatives in non-scalar implicatures are associated with cost difference, including interpretation, prevalence, and conceptual primitiveness. The following chapter (Chapter 6) expands the discussion on cost.

5.1 Motivation for a competition-based proposal for NSIs

The basic competition-based proposal attempts to answer the question why listeners reason about the utterance “Anna hen gao”, which has a pos interpretation. As a consequence, they infer COMP from hearing “gao”. The initial attempt of the competition proposal starts with the classic Standard recipe. Prior to introducing the proposals, a context will be provided for them and the motivations behind them will be detailed.

This section hence introduces some of the well-studied theories of scalar implicatures, and examines how alternatives are determined and *constrained* for a given sentence. The goal is to collect inspirations from Katzir (2007) and Fox & Katzir (2011) in terms of how to analyze alternatives, so that we can find a way to break the symmetry in non-scalar implicatures. A fully fledged proposal is elaborated accordingly in the next two sections.

5.1.1 The symmetry problem in SIs

The scalar implicatures of utterance u are the negation of all alternatives of u that are logically stronger than u and are relevant. Consider an example below.

- (1) a. Alex has been to Chicago or Milwaukee. (utterance)
- b. \neg Alex has not been to both Chicago and Milwaukee. (scalar implicature)
- c. ALEX HAS BEEN TO (BOTH) CHICAGO AND MILWAUKEE. (alternative)

By using (1a), the speaker implies (1b), which is derived by negating the alternative (1c). In other words, (1b) is the scalar implicature of (1a), and the alternative of (1a) is (1c).

The so-called *symmetry problem* has to do with the constraints on what counts as alternative. Without such constraints, every potential alternative ψ has a symmetric partner, *not* ψ (Fox, 2007; Katzir, 2007; Chierchia et al., 2012; Asherov et al., 2021). Concretely, a theory is needed here to make sure that (2) must not be an alternative.

- (2) Alex has been to Chicago or Milwaukee but not both.

If (2) were considered as an alternative, then it would generate its negation as a scalar implicature: Alex has been to Chicago and Milwaukee. As a consequence, the speech act of using (1) would give rise to an inference that Alex has been to Chicago AND Milwaukee, which is not the implicature that's observed here, as it contradicts (1b). Therefore, to better explain why (1a) has (1b) as a scalar implicature, it needs to be explained first why (2) does not enter the computation of scalar implicatures yet (1c) does.

Conceptually, the idea is that u with a scalar implicature that $\neg\psi$ for some alternative ψ must not have such an utterance as an alternative: $u \wedge \neg\psi$. If $u \wedge \neg\psi$ were considered as an alternative, it would generate a scalar implicature that $\neg(u \wedge \neg\psi)$. This is a contradiction to the literal meaning of u and its scalar implicature $\neg\psi$. The literature calls such alternatives as ψ and $u \wedge \neg\psi$ *symmetric alternatives*.

Essentially, the symmetry problem is how to make sure that ψ is an alternative of u , but its symmetric alternative $u \wedge \neg\psi$ is NOT an alternative of u . Horn (1972) proposes a method for avoiding

the symmetry problem, which is to assume that alternatives are restricted lexically. For example, *A and B* is a lexically specified alternative to *A or B*, whereas *A or B but not both A and B* is not. As a result, *A or B but not both A and B* can be successfully prevented from entering the competition. Built up on that, a more principled explanation has been proposed — the *structural approach*.

5.1.2 Structural alternatives in SIs

Katzir (2007) and Fox and Katzir (2011) intended to adopt the structural approach to solve the symmetry problem with a general algorithm for constructing alternatives. The intuition is that alternatives of an utterance cannot be of greater structural complexity than the utterance itself. as illustrated below. What (5.1) means in plain English is that the structural alternatives of an utterance ϕ are those structures that are at most as complex as ϕ , where complexity is defined in terms of structural operations: basically, the alternatives of ϕ are those obtained by replacing lexical items in ϕ with other lexical items from the language, and/or deleting structure from ϕ .

$$\text{Katzir 2007 : } A_{\text{str}}(\phi) = \{\psi \mid \psi \lesssim \phi\} \quad (5.1)$$

For example, *some* and *some but not all* are not structural alternatives because the latter is more complex; by contrast, *some* and *all*, which are of equal complexity, are legitimate alternatives. In particular, Katzir (2007) defines the set of *formal alternatives* in (3).

- (3) The set of *formal alternatives* $F(S)$ of sentence S in context c is the set of sentences derivable by successive replacement of constituents of S with items in the *substitution source* of S in c .

The definition of *substitution source* is further illustrated in (4) by Katzir (2007).

- (4) An item α is in the *substitution source* of S in c if
- a. α is a constituent that is salient in c (e.g. by virtue of having been mentioned); or
 - b. α is a sub-constituent of S ; or
 - c. α is in the lexicon

As a consequence, one could replace a word (e.g., a noun) with any other noun in the lexicon. Accordingly, the theory would overgenerate if all formal alternatives were always employed to derive implicatures. Fox and Katzir (2011) therefore assume that the context specifies a subset C of $F(S)$, so that contextually *relevant* alternatives can be represented.

The scalar implicatures of S are thus computed by negating the members of C that are not weaker than S . Stepwise calculation of scalar implicatures is given in (5) by Sauerland (2004) and Fox (2007).

- (5) *Scalar implicatures*: The scalar implicatures of S are the negations of all alternatives A of S such that:
- a. A is an element of $F(S)$ and A is RELEVANT and NOT WEAKER than S , and
 - b. $\neg A$ does not contradict the negation of any relevant A' element of $F(S)$ given S .

For instance, consider the following example:

- (6) Context: Alex got drunk. What did Joe do?
- a. He smoked pot.
 - b. $\neg \rightarrow$ *Joe didn't get drunk*

The scalar implicature in (6b) is derived by negating the sentence *He got drunk*. This is a formal alternative to (6a), because the constituent *got drunk* is contextually salient in this example.

There are generally two problems for the structural approach, one of which is undergeneration. The needed formal alternative cannot be generated under the structural approach assumptions. Consider the following examples (7) in Japanese:

- (7) *Alex-wa ki-te yoi.*
 Alex-TOP COME-GERUND good
 'Alex is allowed to come.'

Utterance (7) expresses deontic possibility with the predicate *yoi*. *yoi* is morphologically an adjective. On the other hand, deontic necessity in (8) do not involve adjectives. In particular, (8)

involves *nar-* or *ike-*, which is a verbal stem, with the negative suffix *-(a)nai*. As a results, the main predicate turns out to be more complex than in (7) (Breheny et al. 2018).

- (8) *Alex-wa ko-naku-te-wa nar-anai/ike-nai.*
 Alex-TOP COME-NEGATION-GERUND-TOP become-NEGATION/go-NEGATION
 ‘Alex must come.’

The structural approach assumes that addition (*adding negation*) means increasing complexity, thus (8) cannot be derived from (7) by substitution and deletion alone.

To sum up and to circle back to the current puzzle about non-scalar implicatures, the structural approach’s problems get carried on when we adopt it to address the non-scalar implicatures issue. Essentially, the assumption that formal alternatives cannot be derived from addition¹ since it’s *increasing* complexity can be hard to satisfy cross-linguistically. This results in undergeneration. That said, the structural approach is insightful to the current puzzle for two aspects: (i) it takes into consideration *contextual* salience, which speaks to the empirical observation in Chapter 3; (ii) the operations (addition, deletion, and replacement) are still “revisable” to constrain alternatives in non-scalar implicatures.

5.2 Basic proposal

Now we are ready for the proposal. This section illustrates an initial attempt to address the puzzle, using the classic Standard Recipe, both the belief-based recipe and the intention-based recipe. A disambiguation model is proposed in a nutshell. This raises fruitful new questions about the symmetry problems in non-scalar implicatures, which ultimately necessitates a refined model.

5.2.1 The disambiguation model in a nutshell

This subsection discusses the details of the proposed disambiguation proposal, which aims at integrating intention and non-scalar inferences. It turns out that this first attempt fails to fully

¹Or other Katizirian operations, for example, replacement — replace the existing unit with a larger and more complex unit.

explain the puzzle, yet it leads to new questions, which challenges well-established theories of scalar implicatures. Overall this subsection provides critical motivations for the *refined* proposal in the following sections.

I argue that non-scalar implicature (i.e., NSI) differs from scalar implicature (i.e., SI) in that NSI concerns the *intentions* of the speaker, not (just) the speaker's *beliefs*. A speaker who utters (9) below uses an ambiguous expression not because they intended to convey an ambiguity, nor that they meant one reading is true and the other false, nor that they are uncertain about which reading is true. Instead, they have a particular *intention* in mind, and they trust the listener to arrive at the target interpretation. In the case of uttering (9), the target interpretation is usually *COMP*.

(9) *Kai gao.*
Kai tall
“Kai is tall/ taller”

(10) *Kai hen gao.*
Kai HEN tall
“Kai is tall”

(11) *Kai bi Anna gao.*
Kai than Anna tall
“Kai is taller than Anna”

To derive the inference “speaker intended *COMP*” for (9), one can try to adapt the Standard (neo-Gricean) Recipe for Quantity implicature (Geurts, 2010), but for Manner implicature, as illustrated in the steps below:

1. Speaker *S* uttered (9), which is ambiguous between *POS* and *COMP*. On the surface, this speech act violates the Maxim of Manner.
2. Nevertheless, we can still assume that *S* is cooperative and rational, and that *S* is following Manner.
3. We can assume, then, that either *S* intends to convey *POS*, or *S* intends to convey *COMP* (*Intention-based Competence Assumption*).

4. If *S*'s intended meaning were POS, then *S* could have used the unambiguous alternative (10).
But *S* did not.
5. Why not? The most likely explanation is that POS is not what *S* intended.
6. Therefore, *S* intended COMP, and (9) is disambiguated to its usual interpretation.

5.2.2 Alternatives in non-scalar implicature

The reasoning algorithm above relies crucially on the presence of (10) as an alternative of (9), and on the *absence* of (11). Both points raise questions.

First, from a structural perspective, (10) is strictly more complex than (9) (at least, *prima facie*), hence should not be a formal alternative, on a strict Katzirian view. Second, if (11) or its variants (12a,b,c) but not (10) were the alternative considered, then (9) would disambiguate to POS, not COMP.

- (12) a. *Kai bi suoyou.ren gao.*
Kai than everyone tall
Kai is taller than everyone (in the context).
- b. *Kai bi.jiao gao.*
Kai more tall
Kai is taller (than contextually salient individuals).
- c. *Kai geng gao.*
Kai much/more gao.
Kai is taller (than contextually salient individuals who are taller than average).

Moreover, if both were considered, there would be no disambiguation at all. These puzzles about alternatives are reminiscent of the symmetry problem that arises for SIs, which is what my discussion is going to turn to.

5.2.3 The symmetry problem in non-scalar implicatures

To put it formally and to further narrow down the puzzles, why is “Anna hen gao” the salient alternative to “Anna gao”? Why “Anna hen gao” (*but not* (13)) leads to the desired implicature? If

we were to consider (13), we will get a symmetry, which is the exact same problem that arises in Scalar Implicature. The Symmetry Problem is replicated in the domain of *Non-scalar Implicature*.

- (13) *Anna bi Bob gao.*
 Anna than Bob tall
 Anna is taller than Bob. (COMP)

There can be two hypothesized solutions. Specifically, Hypothesis I says that suppose “Anna gao” is ambiguous between intended COMP and intended POS, corresponding to the two separate unambiguous alternatives, then if the speaker intended POS, they could have used the unambiguous alternative “Anna hen gao”; similarly if the intended message is COMP, the unambiguous COMP alternative “Anna bi Bob gao” would have been used. As a result, the question of how to reconcile the ambiguity leaves unanswered.

Essentially, note that under Hypothesis I, we would derive “ S intended POS or S intended COMP, but S did not intend POS (because S could have uttered (10)), and S did not intend COMP (because S could have uttered (11))”. This is a contradiction.

Hypothesis I:

1. $\llbracket \text{Anna gao} \rrbracket$: Intended COMP OR POS
2. alternative $\llbracket \text{Anna hen.gao} \rrbracket$: Intended POS
3. alternative $\llbracket \text{Anna bi Bob gao} \rrbracket$: Intended COMP
4. Intended POS \leadsto “Anna hen gao”
5. Intended COMP \leadsto “Anna bi Bob gao”
6. How to reconcile the ambiguity?

Hypothesis II also starts with the assumption that “Anna gao” has two interpretations, corresponding to two separate alternatives. However, only the comparative alternative “Anna bi Bob gao” enters the computation of disambiguation-based manner implicatures. Therefore, using “Anna gao” yields the negative inference that the speaker does not intend COMP. Consequently, we get the

disambiguation-based manner implicature that the speaker intended POS, which does not always align with the implicature that's usually observed². Recall that when uttering “Anna gao”, the speaker intends the COMP message.

Hypothesis II:

1. $\llbracket \text{Anna gao} \rrbracket$: Intended COMP OR POS
2. alternative $\llbracket \text{Anna bi Bob gao} \rrbracket$: Intended COMP
3. Intended COMP \leadsto “Anna bi Bob gao”
4. “Anna gao” $\leadsto \neg(\text{Intended COMP})$
5. “Anna gao” \leadsto Intended POS

While one could propose that COMP is somehow irrelevant (hence, not considered, despite (11) being an alternative), such a proposal lacks independent motivation. Absent such an account, a theory of alternatives is needed that allows (10) but not (11) to be an alternative of (9).

5.3 Refined proposal

As discussed in the previous sections, the symmetry problem exists in non-scalar implicatures as well, yet the well-cited theories for scalar implicatures cannot account for the symmetry puzzle in the non-scalar implicatures domain. This section thus devotes discussion to possible directions that can be adopted to break symmetry observed in the non-scalar implicatures domain.

This first attempt and the basic proposal do not stand up to close scrutiny, as it's overly simplistic – any utterance and its symmetric partner can be an alternative. We won't be able to make good predictions with the basic proposal.

In other words, it does not seem that one can capture symmetry problem in Manner implicature in (classic) Gricean way. Because the standard theory of alternatives involve Horn scale (Horn, 1972), which is irrelevant to the Mandarin *hen* puzzle that's observed here. Furthermore, Katzir

²It does sometimes. I found in my experiment that people readily interpret “Anna gao” a POS when the standard is salient. See Chapter 3

(2007) argues that alternatives are either more complex or not, meaning it's a discrete notion. Under the analysis of Katzir (2007), *prima facie*, *bi gao* and *hen gao* are more complex than *gao*. Operation-wise, both expressions are *adding* lexical items. Thus, hypothetically speaking, neither of them should enter into the competition with *gao* in the first place.

Essentially, This dissertation argues that gradient cost breaks symmetry. Katzirian operations, for example *addition*, indeed add cost. But as long as it's not adding *significant* amount of cost to the alternative, this alternative is allowed to enter into the competition. Once in the competition process, alternatives' *costs* get compared against each other. Consequently, the cost difference will be able to break the symmetry as the reasoning iteration proceeds.

5.3.1 Assumptions

A likely solution to break the symmetry is to update Katzir by incorporating costs and therefore turning a discrete view into a continuous one³. This dissertation still adopts Katzir's theory of complexity. Crucially, I assume that alternatives get associated with costs, which are calculated in a Katzirian way. Any alternative that costs at most *one* unit is involved in the competition. Thus, *gao* (COST = 0) and *hen gao* (COST = -1) are in but *bi NP gao* (COST = -2) gets excluded. Note that the values here are given for illustration purpose only. They are *relative* cost, instead of precise cost. What matters is that "hen gao" is a relatively cheap alternative of "gao", while "bi..." (as exemplified in (14) is a relatively costly alternative.

- (14) *Anna bi Bob gao.*
 Anna than Bob tall
 Anna is taller than Bob. (COMP)

Put otherwise, this dissertation updates Katzirian account by making it gradient and continuous. I argue that utterance *hen gao* is not a VERY costly alternative of "gao". The consequence of that claim is that even though it goes against Katzir at surface, involving *hen gao* in the competition is

³See the explanation of Katzir in section 5.1.

not adding *significant* amount of complexity. By contrast, the comparative alternative *bi gao* will add significant amount of complexity.

Conceptually, the main idea is that a speaker used an ambiguous expression not because they wanted to convey an ambiguity, not that they wanted to convey that one interpretation is true while the other is false. Instead, they have a specific intended meaning in mind, and they lead the listener to figure it out. In the following discussion I will use the probability pragmatics framework as a tool to formalize this intuition, and to implement various versions of pragmatic theories (e.g., Gricean pragmatics, probability pragmatics, etc.).

Concretely, I'm following Bergen et al. (2016) and Buccola et al. (2021) in that I assume alternatives come with different costs. For instance, in terms of Katzir's replacement operation, presumably the larger the unit being replaced, the cheaper the cost, the more probable the unit is going to enter the computation of (non-)scalar implicatures. Conversely, the more complex the replacement unit, the more costly, the less probable it will get activated.

5.3.2 Derivations

This dissertation argues that (11) is *substantially* more costly than (10) because it has extra layers of complexity, for example it requires a second argument which is a content word.

The cost difference between utterance (11) and utterance (10) breaks the symmetry. Further, Zhang & Ling (2020) argue that the English comparative morpheme “more” actually marks the discourse salience of the comparison standard, rather doing the comparison itself. Chinese is among the group of languages in which the comparative constructions never have a true comparative morpheme like “more” to mark the comparison standard; instead the *bi* morpheme as in (11) introduces the standard. One likely explanation of why it's hard to access the comparative alternative (11) is that marking the comparison standard is costly. Accessing words that are (distantly) available in the discourse but are beyond one sentence unit can be computationally expensive. The moment speakers consider the costs of (9)'s alternatives, they are likely to favor the “taller” over the “tall” meaning.

- (15) *Kai gao.*
 Kai tall
 “Kai is tall/ taller”
- (16) *Kai hen gao.*
 Kai HEN tall
 “Kai is tall”
- (17) *Kai bi Anna gao.*
 Kai than Anna tall
 “Kai is taller than Anna”

Consider sentences (16,17) as alternatives of the utterance (15), the step-wise iteration for the disambiguation-based competition model is illustrated below.

1. In scenario (a) where the speaker intends to communicate comparative “taller”, (11) *bi gao* is more informative than (9) *gao*, but the cost of (11) is greater than the cost of (9).
2. In scenario (b), speaker intends positive “tall”, (10) *hen gao* is more informative than (9) *gao* and the cost of (10) is greater than the cost of (9), but the *difference* is small, compared to that in scenario (a).
3. Presumably, both the speaker and the listener observe Grice maxims. Speaker has a particular intention in mind. Listener assumes that the speaker is being cooperative.
4. Because speaker was more probable to use (9) *gao* in the comparative “taller”-situation than in the positive “tall”-situation, and listener thus tends to interpret (9) *gao* as comparative “taller”.
5. As a consequence, listener infers, from hearing (9) *gao*, that the speaker most likely intends to communicate the comparative meaning “taller”.

In the next iteration, the efficiency of using (9) *gao* to communicate the comparative meaning “taller” has increased, so the effect gets amplified. One may wonder why the comparative alternatives are more costly than the positive ones in the first place.

So far I have illustrate the derivations about the cost difference using Gricean and probabilistic terms. How to further formalize these steps? What are the independent motivations for different costs? That's what we will turn to – the syntax-semantics structure of the degree expressions alternatives, and the independent stipulation for the notion of “cost”.

5.4 Chapter summary

This Chapter gives the core proposal. I start with the classic standard recipe proposed for scalar implicatures, implement it into the analysis of non-scalar implicatures, and I find that it cannot be readily transferred to address the core puzzle: how to constrain alternatives and break the symmetry in non-scalar implicatures. Given that this puzzle has been extensively studied in the scalar implicatures domain, I therefore revisit the classic Katzirian views. Revised proposal is given afterwards, together with the key assumptions and detailed implementations of various versions of pragmatics.

CHAPTER 6

UNDERSTANDING *COST*

This chapter proposes that *POS* has simpler semantic interpretation than *COMP* does. (Zhang 2020; Zhang & Ling 2020), it is more primitive than *COMP* cross-linguistically (Grano, 2012; Grano & Davis, 2018), and it has higher distribution frequency compared to *COMP*. Thus, one can assign a bigger value to *cost(comparative alternatives)* than to *cost(positive alternatives)* when using probabilistic pragmatics (e.g., RSA) to model *disambiguation* inference.

The key question this part addresses concerns the cost parameter in the RSA framework. Namely, what exactly makes for independent evidence for the value assignment in (1):

- (1) a. $COST(gao) = 0$
- b. $COST(hen\ gao) = -1$
- c. $COST(bijiao\ gao, geng\ gao, bi\ NP\ gao) = -2$

Before diving into the question of independent evidence of cost assignments, it is worth highlighting that the values here (0, -1, -2) are not categorical in the sense that “precise” costs were assigned to each alternative. Instead, the cost values in (1) are meant to be *relative*. For instance, “Anna hen gao” does appear to be costly with respect to “Anna gao”, however, “Anna hen gao” is *not* costly relative to “Anna bi Bob gao”. Thus the positive alternative “hen gao” is activated and enters the computation of manner implicatures yet “bi NP gao” is not.

6.1 “Costly” with respect to semantic interpretation

This part analyzes the role of logical form and semantic compositions as the source of degree expressions’ “costs”. I first reconstruct the argumentation of similar proposals by Moracchini (2018), followed by a close scrutiny of Zhang (2019a) and Zhang & Ling (2021). Accordingly, I gave the semantic interpretations for ‘gao’ and other alternatives, which are associated with different costs

and complexity. The proposed interpretations imply that (null) operators and structural complexity are related.

6.1.1 Structural competition among alternatives

It turns out that the view that compositional complexity in lexical items determine lexical items' costs has been sketched by Moracchini (2018). Moracchini (2018) seems to pursue the same idea of structural competition among alternatives in a similar domain (non-scalar implicatures).

Moracchini's (2018) proposal in a nutshell Rett (2008) argues evaluativity inferences result from a competition between "marked" versus "unmarked" degree constructions that are semantically equivalent. Moracchini (2018) examines the source of markedness in these constructions by proposing that structural complexity is the right metrics for a competition-based account of evaluativity.

Concretely, (3a) and (3b) are related to (2) through a negative operator. They argue that *short* is the negation of *tall* (3a), and the comparative operator is changed from *-er* to *less* (3b). The core data is illustrated below in (2).

- | | | |
|-----|--|----------------------|
| (2) | Athos is taller than Porthos is. | <i>utterance</i> |
| (3) | a. Porthos is shorter than Athos is. | <i>alternative 1</i> |
| | b. Porthos is less tall than Athos is. | <i>alternative 2</i> |

Therefore, the three comparatives entail each other. The core data is given in (4). What's puzzling is that (4) is not a good paraphrase for (2). (4) introduce an *additional* entailment that Athos and Porthos have to count as 'short' in the context (4a,b); and Athos and Porthos count as 'tall' in the context (4c). However (2) does not have such entailment.

- | | | |
|-----|---|---|
| (4) | a. Athos is less short than Porthos is. | <i>(the comparative of inferiority)</i> |
| | b. Porthos is more short than Athos is. | <i>(the analytic comparative)</i> |
| | c. Athos is more tall than Porthos is. | <i>(the analytic comparative)</i> |

The above paradigm raises the following questions: how can we derive suitable meanings for the evaluative sentences in (4) while preserving our assumptions about the semantics of degree morphemes? Is there a way of predicting which degree constructions are evaluative? Moracchini (2018) provides a theory of what counts as alternatives for semantic competition in the aP domain.

The EVAL theory of evaluativity (Rett 2007, 2008) proposes that in some degree constructions, only the non-EVAL reading is available, while in others only the EVAL reading is available, suggesting the non-EVAL construal is blocked. This blocking effect is due to a markedness competition. Rett suggests that there are three types of markedness triggers, illustrated as below:

1. The negative adjective is marked whereas the positive adjective is unmarked
2. *less* is marked, whereas the degree operator *-er* is unmarked
3. Analytic comparatives (*more smart*) are marked whereas synthetic (*smarter*) comparatives are unmarked

Against this background, Moracchini (2018) argues that analytic comparatives do not have EVAL parse (6a) because (5a) and (6a) have the same semantics but (5a) is synthetic (less marked). Thus (6a) is blocked by (5a). When the EVAL operator is included, (5b) and (6b) no longer have the same semantics. As a consequence, (6b) is not blocked by (5b). So far Rett's theory works well.

(5) Athos is taller than Porthos is.

- a. *Non-evaluative parse*: [Athos is tall] [-er [than Porthos is tall]]
 $\text{MAX}(\lambda d.\text{tall}(\text{athos},d)) > \text{MAX}(\lambda d'.\text{tall}(\text{porthos},d'))$
- b. *Evaluative parse*: [Athos is tall] [-er [than Porthos is EVAL tall]]
 $\text{MAX}(\lambda d.\text{tall}(\text{athos},d)) > \text{MAX}(\lambda d'.\text{tall}(\text{porthos},d')) \wedge d' > \text{Standard}_{\text{tall}}$

(6) Athos is less short than Porthos is.

- a. *Non-evaluative parse*: [Athos is short] [less [than Porthos is short]]
 $*\text{MAX}(\lambda d.\text{short}(\text{athos},d)) < \text{MAX}(\lambda d'.\text{short}(\text{porthos},d'))$

- b. *Evaluative parse*: [Athos is short] [less [than Porthos is EVAL short]]
 $\text{MAX}(\lambda d.\text{short}(\text{athos},d)) < \text{MAX}(\lambda d'.\text{short}(\text{porthos},d')) \wedge d' > S_{\text{tall}}$

Nevertheless, (7b) and (5b) have the same semantics — they compete and there is a symmetry. (7b) is more marked than (5b), thus such markedness difference breaks the symmetry. Consequently, (7b) should be blocked by (5b). However, this is unwanted, because (7b) is an available parse. How to address this issue? Why is (7) licensed under (7b)? Moracchini (2018) has a story.

(7) Athos is more tall than Porthos is.

- a. *Non-evaluative parse*: [Athos is tall] [more [than Porthos is tall]
 $*\text{MAX}(\lambda d.\text{tall}(\text{athos},d)) > \text{MAX}(\lambda d'.\text{tall}(\text{porthos},d'))$
- b. *Evaluative parse*: [Athos is tall] [more [than Porthos is EVAL tall]]
 $\text{MAX}(\lambda d.\text{tall}(\text{athos},d)) > \text{MAX}(\lambda d'.\text{tall}(\text{porthos},d')) \wedge d' > S_{\text{tall}}$

The markedness competition account from Rett correctly explains the missing non-EVAL readings of analytic comparatives, but it only captures half of the puzzle: there are no explanations of why analytic comparatives are licensed under their EVAL parse. According to Moracchini (2018), ‘Markedness’ is structural complexity. The key assumptions include the Syntactic Negation Theory of Antonymy (Heim 2008), and the decomposition of the degree operator *more* suggested in Solt (2009). Specifically, There are no (semantics) entries for *less*, *short* and *more*. Their meaning is derived via two covert operators **little** and **much**.

- (8) a. [-ER+LITTLE] > *less*
 b. [LITTLE+TALL] > *short*
 c. [-ER+MUCH] > *more*

Inspirations for the current disambiguation-based competition proposal A crucial consequence of Moracchini’s (2018) analysis is that the (hidden) operator adds up extra layers of syntactic structure when they are included in the derivation. In particular, *short* is structurally more complex than *tall*. The difference between degree operators can be stated in terms of structural complexity. A degree

expression that contains *less* or *more* is more complex than a degree expression that only involves the comparative head *-er*. The constructions considered as ‘marked’ on Rett’s approach involve structurally complex aPs (namely, *less short*, *more tall* and *more short*. Moracchini (2018) follows the LF-Economy principle, by which structurally complex aPs are precluded by simpler structural alternatives whenever these alternatives express the same meaning.

(9) *Minimize aPs! (Final version)*

For any LF ϕ , any aP α in ϕ , α is deviant in ϕ if α can be replaced in ϕ with an **expressible structural alternative**, β , such that

- a. β is **semantically equivalent** to α , and
- b. β is **structurally simpler** than α

One of Moracchini’s (2018) success is that evaluative synthetic aPs no longer qualifies as potential competitors. EVAL analytic forms satisfy both interfaces: EVAL occurs at word boundaries without intervening in PF processes, and in absence of relevant structural alternatives, their EVAL parses is licensed by *Minimize aPs!*, as illustrated above.

To sum up and to circle back to the current puzzle about *cost*, Moracchini (2018) provides a way to show that *markedness* can be further explained through structural complexity. The main take-home message is that having covert operators in the structure can lead to complex and “marked” interpretations, making the utterance/interpretation heavy in terms of weight, as it has more components than its primitive counterparts does, which ultimately increases its cost. Overall, Moracchini (2018) is a crucial inspiration for the proposal of connecting *cost* with semantic interpretation.

6.1.2 Interpreting alternatives *geng* and *hen*

This part of the discussion takes a close look at Zhang (2019) and Zhang & Ling (2020). Their studies provide extensive investigations into the semantics of comparatives and positives, on which the current proposal is based. Moreover, similar to Moracchini (2018), Zhang (2019a) and Zhang &

	σ (standard)	δ (difference)
POS	a typical or relevant average (often overt) (<i>hen</i>)	an unspecified value (always covert)
COMP	a salient standard (<i>bi</i> -phrase) (covert or overt)	a measurement phrase (covert or overt)

Table 6.1: Zhang’s (2019) analysis

Ling (2021) take a compositional approach to interpret comparatives and positives. This will offer another semantics-related piece of evidence revealing the cost difference of the alternatives.

Semantic approach of degree expressions in Chinese A denotation of *gao* is given in (10) by Zhang (2019). It takes a degree argument, a second degree argument, and an individual. Zhang (2019) analyzes the semantics of gradable adjectives as a relation among three items: σ , δ , and x ; namely the comparison between the measurement of an individual x and a certain standard σ leads to a difference δ . Zhang further argues that positive and comparative differ in their arguments σ (standard) and δ (difference). The essence of the analysis is summarized in table (6.1). Essentially, when δ is unspecified, the positive¹ reading is available; otherwise the difference of measurements is specified, and/or there is an overt measurement phase, we get the comparative interpretation.

(10) L.Zhang (2019:5 ex 9)

$$\llbracket gao \rrbracket_{\langle d, \langle d, et \rangle \rangle} \stackrel{def}{=} \lambda \sigma_d. \lambda \delta_d. \lambda x_e. HEIGHT(x) - \sigma = \delta$$

Table (6.1) suggests that the three uses of gradable adjectives differ in their arguments σ and δ : the use of gradable adjective is inherently ambiguous; the use of other elements (like *hen*) helps to disambiguate; and the non-use of these disambiguating elements leads to a seemingly ‘default’ interpretation.

¹Note that the term “evaluative” or EVAL is never used in Zhang & Ling (2021) or in Zhang (2019a). Their work stick with the term *positive*.

Discourse salience in degree expressions Recently, Klein (1980)’s puzzle and the core contribution of English *-er/more* are extensively discussed in Zhang & Ling (2020). They use intervals, namely ranges of values, to represent scalar values in a generalized way. Measure function is given in (11). It maps a single entity to an interval, representing the position corresponding to the measurement of the entity along a relevant scale.

$$(11) \quad \textbf{Measure function: } \text{HEIGHT}_{\langle e, dt \rangle} \stackrel{\text{def}}{=} \lambda x. \text{HEIGHT}(x)$$

Here HEIGHT means a measure function, mapping an individual to a degree on a relevant scale (here *height*). In other words, HEIGHT takes in an individual x and returns the interval of degrees. Zhang & Ling highlight that measurements are subject to uncertainty, and vagueness is involved in the informative interpretation of a measure function. Specifically, the exact position range on a scale of height corresponds to the given entity’s height measurement. Such a range is dependent on contextual factors, for example acceptable criteria of precision. With respect to “comparison class”, i.e., ‘objects deemed somehow similar to the target of prediction’ (Kennedy 2011), Zhang and Ling (2020) argues that this notion is relevant to contextually informative precision level of measurement.

Built up on the measurement analysis, Zhang and Ling take the semantics of a gradable adjective as a relation between an individual x and interval I (12). Under this view, *tall* and *short* are mapped to the same dimension of height, but the scales of orderings are the opposite to each other. Similarly for *early* and *late*, which are both associated with time, but with scales of opposite orderings.

$$(12) \quad \text{Zhang \& Ling (2020:25 ex 49)}$$

$$\llbracket \textit{tall} \rrbracket_{\langle dt, et \rangle} \stackrel{\text{def}}{=} \lambda I_{\langle dt \rangle}. \lambda x_e. \text{HEIGHT}_{\langle e, dt \rangle}(x) \subseteq I$$

i.e., the measurement of x falls at the position I on the scale of height.

Denotation (12) indicates that the measurement of x falls at the position I on a scale associated with the dimension of the adjective. This analysis is in line with the canonical view of gradable adjectives (Kennedy, 1999).

The relevant part of their discussion to the current puzzle about cost is the following. Zhang & Ling argues that, first, *-er/more* contributes to the semantics of comparatives by playing the role of

σ (standard) [2019]/ I_{STDD} [2020]	
Comparative	<i>than-</i> (with discourse salience)
Measurement construction	absolute zero point (no discourse salience)
Positive use	a typical or relevant average (no discourse salience)

Table 6.2: The standard and differential involved in comparison (Only the marker of discourse salience and numerals are pronounced) summarized in Zhang (2019) and Zhang & Ling (2020)

the default differential. Second, The default positive value $(0, +\infty)$ aside, the differential status of *-er/more* is due to its additivity, a kind of anaphoricity. In this sense, what *-er/more* marks is actually the discourse salience of the value serving as the standard of comparison. Last but not the least, compared to other uses of gradable adjectives, comparatives are special in involving standards that have discourse salience. As illustrated below, (13a) implies that the heights of John and Bill are compared with the same context-relevant standard; while (13b) implies that the height of John is compared with a context-relevant standard, while the height of Bill is compared with the height of John². Crucially, here the height of John has discourse salience.

(13) a. If John is tall, then Bill is tall.

b. If John is tall, then Bill is taller. (Zhang&Ling 2020:49, example 102)

(14) a. $\llbracket \text{-er/more} \rrbracket^{\text{def}} = (0, +\infty)$ (i.e., the most general positive interval)

Requirement: there is a discourse salient scalar value serving as comparison standard
(i.e., the base for increase)

b. $\llbracket \text{POSITIVE-VALUE} \rrbracket^{\text{def}} = (0, +\infty)$ No additional requirement

The standard and differential involved in comparison is illustrated in (table 6.2), in which only the marker of discourse salience and numerals are pronounced. Note that Zhang (2019) used σ to represent *standard*, while Zhang and Ling (2020) used I_{STDD} instead.

²However, one may find (13b) does not sound natural, since it implies that whether or not Bill is taller than John depends on whether John is considered tall, which is not the case.

Overall, Zhang (2019a) and Zhang & Ling (2021) provide extensive discussion for their hypothesis that comparatives are special in a way that it involves marking discourse salience. I argue, following their investigation, that this distinct feature of comparatives makes it *costly*, relative to the positives. Furthermore, from the perspectives of compositional semantics, their analysis implies that the comparatives have more a costly semantic interpretation than the positives do. Let's implement their analysis and spell out the semantics of *gao* and the alternatives.

6.1.3 Spelling out the semantics

This part of the discussion spells out the semantic analysis of *gao*, and its alternatives *hen gao*, *bi NP gao*, and *geng gao*.

Initial attempt My initial attempt is that *pos* is an operator, which takes in 'gao' as an argument. This aligns with Kennedy & Levin (2008):

$$(15) \quad \llbracket \text{Anna } hen.gao \rrbracket = \text{pos}(gao)(Anna) = \text{HEIGHT}(Anna) > \text{STANDARD}$$

In plain English: the proposition is true iff the height of Anna is greater than the standard of tallness; in other words, *hen* is analyzed as *pos* (*hen* == *pos*)

But this hypothesis has very limited predictive power. Chinese seems to be the only language that supports the claim that *hen* is the *pos* operator (i.e., *hen* == *pos*). There seems to be more evidence showing that *hen* is not *pos* (i.e., *hen* != *pos*). For example, Rett (2014) provides evidence and arguments against the claim that *hen* == *pos*. The equative below is evaluative and it's like the English counterpart. However, (16) is ungrammatical with *hen*.

- (16) *John he Mary yiyang (*hen) ai.*
 John and Mary same (*hen) short
 John is as short as Mary.

For some other accounts of the data in (16), *hen* is glossed as 'very' (Sybesma, 1999). Grano (2008) also provides an analysis in which *hen* is argued to be a construction-specific assertive marker

(rather a marker of POS). These analyses all stand against Kennedy (1999b)³. They suggest that evaluativity is *present* in Chinese degree constructions in the *absence* of *hen*.

Moreover, *hen* cannot be analyzed as a pragmatic EVAL either. As discussed in the competition literature review chapter (Chapter 4), EVAL, in Rett's (2015) sense, is *NOT* in the semantics; instead it arises in the process of pragmatics reasoning.

Revised analysis In line with Zhang's (2019) proposal as elaborated earlier, I maintain that particularly for Mandarin Chinese, the use of gradable adjectives is intrinsically underspecified, and its semantics is a relation among three items: the comparison between the measurement of an individual x and a standard σ results in a difference δ . The three uses of gradable adjectives in Mandarin Chinese differ in their arguments σ and δ . Zhang's (2019) analysis succeeds in that (i) it explains why comparative use in Chinese involves comparisons and does not need a (silent) marker – comparisons are encoded by gradable adjectives themselves; (ii) it predicts that a competition based account can disambiguate the inherently underspecified gradable adjectives such as *gao* – different alternatives entering into the competition results in the surfaced reading of *gao*.

Specifically, I would argue that, following Zhang (2019); Zhang & Ling (2020) yet different from Rett (2015), POS is a degree argument but not an operator. The denotation is given below.

(17) Lexical entries (Zhang 2019):

$$\llbracket gao \rrbracket_{\langle d, \langle d, et \rangle \rangle} \stackrel{def}{=} \lambda \sigma_d. \lambda \delta_d. \lambda x_e. HEIGHT(x) - \sigma = \delta$$

³Kennedy (1999b) argues that POS has an *overt* counterpart in Chinese based on (1):

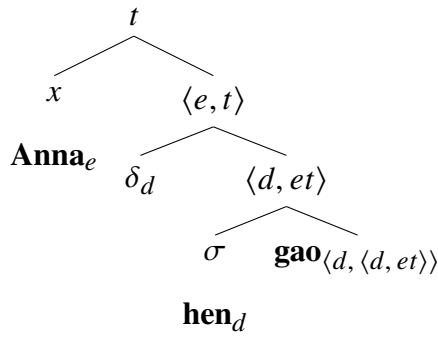
- (1) a. *Zhangsan hen gao.*
 Zhangsan hen tall
 Zhangsan is tall.
 b. *Zhangsan bi ni (*hen) gao.*
 Zhangsan than you (*hen) tall
 Zhangsan is taller than you.

The positive construction in (1a) is evaluative, by contrast, a corresponding sentence without *hen*, (“*Zhangsan gao*”), is infelicitous unless the context of utterance allows for a comparative interpretation. In that case the sentence is not evaluative. Kennedy likens *hen* to POS by suggesting that *hen* is in complementary distribution with the (null) comparative morpheme (1b).

Crucially, for the positive use, σ is a contextually relevant average and often overtly expressed with the presence of *hen*, and δ is a covert unspecified positive value. Under this analysis, *hen* marks σ . The semantics of *hen* refers to an unspecified high value serving as the standard on a relevant scale. Thus, the presence of *hen* in ‘Anna *hen* *gao*’ is analyzed as the standard on the scale associated with gradable adjective *gao* “tall”, which naturally gives rise to the positive reading of *Anna hen gao* ‘Anna is tall’.

In other words, different from Rett’s (2015) view of EVAL being realized in the process of pragmatic reasoning, I propose that EVAL is part of the semantics, and *hen* behaves *like* the standard ($hen \approx \text{STANDARD}$).

(18) Logical Form:



(19) Derivation:

$$\begin{aligned}
 & \llbracket \text{Anna hen.gao} \rrbracket \\
 &= \text{GAO}(\text{hen})(\delta)(\text{Anna}) \\
 &= \lambda \sigma_d. \lambda \delta_d. \lambda x_e. \text{HEIGHT}(x) - \sigma = \delta (\text{hen})(\delta)(\text{Anna}) \\
 &= \lambda \delta_d. \lambda x_e. \text{HEIGHT}(x) - \sigma = \delta (\delta)(\text{Anna}) \\
 &= \lambda x_e. \text{HEIGHT}(x) - \sigma = \delta (\text{Anna}) \\
 &= \text{HEIGHT}(\text{Anna}) - \sigma = \delta^4
 \end{aligned}$$

(20) Truth conditions:

$$\text{HEIGHT}(\text{Anna}) > \text{STANDARD}_{\text{tall}}$$

⁴ δ must be greater than 0.

As to the comparative alternative such as *geng*, according to Zhang (2019), σ is a contextually salient standard, a discourse salient referent, or introduced by a *bi*-phrase. δ can be a covert or overt positive value (i.e., a numerical measurement phrase). For the measurement use, σ refers to the absolute zero point, which is always covert, and δ is always overtly expressed as it is a numerical measurement phrase.

The empirical motivation is that morphemes such as *geng* and *haiyao* appear to often co-occur with the *comparative* reading. As extensively discussed in the previous chapters, without *geng*, (21) is ambiguous between POS and COMP, whereas with the presence of *geng* (21) is unambiguously comparative. Further, Zhang (2019) follows Liu (2010a), arguing that *geng* is never a marker of COMP, instead it brings a presupposition requirement.

(21) Anna *geng* gao.

Anna GENG tall(-er)

Anna is (even) taller.

\leadsto comparative

Specifically, Zhang (2019) observes that with *geng*, the sentence (22) presupposes that Kai (i.e., *the relevant person that Anna is being compared to*) is tall (i.e., taller than the typical tallness or relevant average of their comparison class); without *geng*, there is no presupposition.

(22) Anna *bi* Kai (*geng*) gao (wu li-mi).

Anna COMPARE Kai GENG tall(-er) five centimeter

Anna is (5cm) taller than Kai.

As a consequence, *geng* is a modifier for gradable adjectives, marking the existence of a presupposed comparison and indicating the discourse salience of σ , namely the standard used in the asserted comparison. I therefore give the following denotation for Zhang's (2019) lexical entries in (23):

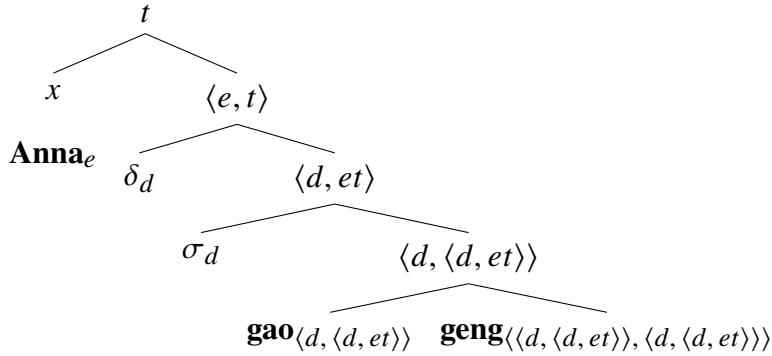
(23) Lexical entries (Zhang 2019):

- a. $\llbracket geng \rrbracket \langle \langle d, \langle d, et \rangle \rangle, \langle d, \langle d, et \rangle \rangle \rangle \stackrel{def}{=} \lambda G \langle d, \langle d, et \rangle \rangle \cdot \lambda \sigma_d \cdot \lambda \delta_d \cdot \lambda x_e \cdot \underline{\exists \sigma' . \sigma - \sigma' = \delta' . G's\ scale(x) - \sigma = \delta}$

$$\text{b. } \llbracket gao \rrbracket_{\langle d, \langle d, et \rangle \rangle} \stackrel{def}{=} \lambda \sigma_d. \lambda \delta_d. \lambda x_e. \text{HEIGHT}(x) - \sigma = \delta$$

The assertion of (23) is that the measurement of x exceeds the standard σ by a difference value δ . The underlined part of (23) is the presupposition: the standard σ exceeds another standard σ' . Zhang mentions that σ' does not have discourse salience, yielding a positive reading for this presupposed comparison). In other words, (23) presupposes EVAL. As to the standard σ , it has to be a salient discourse referent, which can neither the absolute zero point nor a typical “average”. This rules out the co-occurrence of *geng* and measurement phrase. Moreover it explains why with the presence of *geng*, (21) unambiguously means comparative: *geng* marks the discourse salience of the referent standard in the asserted comparison. Here the “standard” σ does not mean the typical tallness or relevant average of the comparison class. Instead, σ' yields the positive meaning, i.e., σ' refers to the average standard tallness that does not often have discourse salience.

(24) Logical form:



(25) Derivation:

$$\begin{aligned}
& \llbracket Anna \text{ geng } gao \rrbracket \\
&= \lambda G_{\langle d, \langle d, et \rangle \rangle}. \lambda \sigma_d. \lambda \delta_d. \lambda x_e. \underline{\exists \sigma'. \sigma - \sigma' = \delta'}. G's \text{ scale}(x) - \sigma = \delta \text{ (gao)}(\sigma)(\delta)(Anna) \\
&= \lambda \sigma_d. \lambda \delta_d. \lambda x_e. \underline{\exists \sigma'. \sigma - \sigma' = \delta'}. \text{HEIGHT}(x) - \sigma = \delta \text{ (}\sigma\text{)}(\delta)(Anna) \\
&= \lambda \delta_d. \lambda x_e. \underline{\exists \sigma'. \sigma - \sigma' = \delta'}. \text{HEIGHT}(x) - \sigma = \delta \text{ (}\delta\text{)}(Anna) \\
&= \lambda x_e. \underline{\exists \sigma'. \sigma - \sigma' = \delta'}. \text{HEIGHT}(x) - \sigma = \delta \text{ (Anna)} \\
&= \underline{\exists \sigma'. \sigma - \sigma' = \delta'}. \text{HEIGHT}(Anna) - \sigma = \delta
\end{aligned}$$

(26) Truth conditions:

HEIGHT(*Anna*) > STANDARD_{tall}

Presupposition: the STANDARD exceeds another STANDARD_{tall}

In plain English, *Anna geng gao* means that Anna is taller than some salient degree (assertion), and the discourse salient referent whom Anna is comparing against *evaluatively* tall (presupposition). In other words, *geng* ‘more’ serves as a (higher order) function that takes *gao* as an argument. As a consequence, relative to the positive forms such as *hen*, the comparative form *geng* is complex and *costly* in terms of its semantic interpretation. As related back to Moracchini (2018), a higher order function adds up extra layers of semantic structure whenever they are included in the derivation⁵.

As to the semantics of another comparative alternative *Anna bi Kai gao* ‘Anna is taller than Kai’, the derivation is spelled out below:

(27) Lexical entries (Zhang 2019):

- a. $\llbracket bi \rrbracket_{\langle \langle d, \langle d, et \rangle \rangle, \langle e, d \rangle \rangle} \stackrel{def}{=} \lambda G_{\langle d, \langle d, et \rangle \rangle} . \lambda x_e . G's \text{ scale}(x)$
- b. $\llbracket gao \rrbracket_{\langle d, \langle d, et \rangle \rangle} \stackrel{def}{=} \lambda \sigma_d . \lambda \delta_d . \lambda x_e . \text{HEIGHT}(x) - \sigma = \delta$

According to Zhang (2019), *bi* is a measure function that generates the standard for comparison. It's worth highlighting the discourse salience of the standard σ for the comparative use in Zhang's (2019) analysis, which is not required for the positive and the measurement use. Therefore, either a *bi*-construction or context will be required to introduce a salient referent/standard. This additional requirement of marking a (discourse) salient referent makes comparative alternatives more costly than the positive ones in pragmatic reasoning.⁶

More about *geng*: does it correspond exactly to *comp*? The fact that (28 — 31)⁷ do not sound natural further supports Zhang's (2019) observation that the morpheme *geng* is not a simple

⁵That said, in the end, “*geng*” is still analyzed as a single lexical item, not decomposed into multiple morphemes. This is a bit different from Moracchini's (2018) analysis, which decomposes “*short*” as [*less* + *tall*].

⁶Note that discussions about comparative semantics can also be referred back to Hohaus' degree semantics analysis in Chapter 2 analysis (9 — 12).

⁷These are introspective data, from my own judgment.

comparative alternative. It does not appear to correspond exactly to COMP. The co-occurrence of *geng* and the negated positive gives rise to contradiction.

(28) #*Anna geng gao, dan ta bu shi hen gao.*

Anna GENG tall, but she not COPULA HEN tall

Anna is (even) taller, but she is not tall.

(29) #*Anna geng gao, dan ta hen ai.*

Anna GENG tall, but she HEN short

Anna is (even) taller, but she is short.

(30) #*Anna geng gao, dan ta he Kai dou hen ai.*

Anna GENG tall, but she and Kai both HEN short

Anna is (even) taller, but she and Kai are both short.

(31) #*Anna he Kai (hen) ai, dan Anna geng gao.*

Anna and kai HEN short, but Anna GENG tall

Both Anna and Kai are short, but Anna is even tall(er).

By contrast, the co-occurrence of *hen* and the *bi* comparative does not give rise to contradiction, as illustrated in (32).

(32) *Anna he Kai (hen) ai, dan Anna bi Kai gao.*

Anna and kai HEN short, but Anna BI Kai tall

Both Anna and Kai are short, but Anna is taller than Kai.

Regarding the comparative alternative *bi* construction, Zhang (2019) provides another crucial piece of evidence showing that the distinction between a positive and a comparative interpretation is driven by context as opposed to syntax. Consider Zhang's (2019) example in (33).

(33) a. *A: Anna bu shi hen gao.*

A: Anna not COPULA HEN tall

b. *B: Na gen Kai xiang-bi Anna gao bu gao?*

B: then with Kai compare-to Anna tall not tall

A: Anna is not tall. B: Then compared with Kai, is she taller? (COMP)

(adapted from Zhang (2019:623, example 19)

	Positive form	Comparative form	Example
Pattern A	A	A	Japanese, Swahili
Pattern B	A	DERIVE(A)	English, Irish, French, Spanish

Table 6.3: Morphosyntactic relationship between positive and comparative forms cross-linguistically (adapted from Grano (2012) and Grano & Davis (2018))

B’s response in (33b) shows that, if a salient standard (i.e. Kai’s height) is supplied in the context, then *gao* in (33b) is interpreted as comparative. Its comparative reading surfaces.

The main takeaway of this section is that, given the sentence “Anna *gao*”, the alternative “Anna *hen gao*” is a relatively cheap alternative (“*hen*” is simply the overt realization of the standard required by “*gao*”), while the various comparative alternatives are all more complex, either structurally or semantically, or both.

6.2 “Costly” with respect to primitiveness

This section captures the notion of “costly” from the perspective of primitiveness, which can be sub-divided into two aspects: cross-linguistic variation (i.e., lexicalization) and mental representations (i.e., conceptual alternatives).

6.2.1 Cross-linguistic lexicalization

Grano (2012) and Grano & Davis (2018) offer an observation about the cross-linguistic picture of POS and COMP:

- (34) “*Universally, the comparative form of a gradable adjective is derived from or identical to its positive form.*” (Grano 2012:515).

The investigation in Grano (2012) and Grano & Davis (2018) indicate that the POS forms tend to be simpler than COMP forms, which may further implicate that the POS meanings are more primitive than COMP meanings.

As a consequence, if a language, for example Mandarin Chinese, does not distinguish the POS from the COMP form, then one would expect the following competition mechanism: when people hear “Anna gao”, which is ambiguous between POS/COMP, they reason about POS because it’s more primitive, thus yielding COMP.

This hypothesized mechanism is borne out — at least the final step reasoning about COMP does speak to the empirical observation summarized in Chapter 3. Overall, this cross-linguistic view about the lexicalization of comparatives and positives provides another piece of evidence suggesting that comparatives are more costly than positives.

6.2.2 Conceptual alternatives

Buccola et al. (2021) illustrates graded stance on Symmetry, with multi-factor come into play on costs. In terms of the basis for choosing an Alternative, Buccola et al. (2021) argues that some alternatives seem to be “language independent”. The core proposal is that contrasts based on conceptual primitivity, instead of linguistic complexity, break symmetry.

The motivation behind their proposal is that the existing algorithms (e.g, structural approach) for the symmetry problem seem to be less explanatory when it comes to alternatives that are inexpressible, ungrammatical, or uninterpretable (Carcassi et al., 2021; Carcassi & Szymanik, 2021; Smith, 2020; Charlow, 2016). Buccola et al. (2021) therefore propose a new algorithm for *conceptual* alternatives. They update Katzir’s algorithm by proposing (35):

- (35) “...alternatives are located in the language of thought, and replacements occur only on the basis of single lexical items, albeit lexical items of the language of thought.”

(Buccola et al 2020:13)

Essentially, the idea is to take the conceptual representation of the sentence, and replace one primitive element with another primitive element. In particular, such a lexical item of the language of thought that replaces part of the structure can be as follows (36) (Buccola et al 2020:13):

- (36) a. a special empty element;

- b. a special pronoun capable of pointing at structures; and
- c. other elements that may or may not be lexicalized in the actual language

Buccola et al. (2021) further revises the above formulation of (36c) by associating replacements with costs that vary on the basis of, among other things, lexicalization in the given language, frequency of the given element / lexical item, or contextual salience (i.e., QUD).

Following Buccola et al. (2021), I argue that “conceptual” alternatives can address the problems that the structural approach (Fox, 2007; Katzir, 2007) runs into (see Chapter 5), in a sense that Buccola et al. (2021) allow the replacing lexical items to be elements that may not necessarily be lexicalized in the actual language. Recall in Chapter 6 section 6.1.3 on *geng*, we have seen that *geng* does not correspond exactly to COMP. Under the analysis of Buccola et al. (2021), we naturally reach the conclusion that in terms of mental representation, there is asymmetry between POS and COMP. POS gets lexicalized as EVAL or *hen* in Chinese, English, and other languages. Yet the COMP alternative is a bit complicated: (i) it does not appear to be readily lexicalized across languages (Grano, 2012); (ii) for languages that does seem to have the lexical items for COMP, a close scrutiny indicates that they do not correspond exactly to the thought COMP (see section 6.1.3).

6.3 “Costly” with respect to frequency

Additionally, word frequency plays a role in determining the degree of “primitiveness”, and thus in influencing the extent to which an alternative is costly. Illustrated below are frequency statistics describing the usage distribution of the positive forms and the comparative forms in Chinese speakers’ daily communication.

Moreover, the Beijing Language and Culture University (henceforth BLCU) created a balanced corpus of 15 billion characters. It’s based on news (Renmin ribao 1946-2018, Renmin ribao (oversea) 2000-2018), literature (books by 472 authors, including a significant portion of non-Chinese writers), non-fiction books, blog and *weibo* entries as well as classical Chinese. The frequency list in table 6.5 is derived from the BLCU corpus. It contains a global frequency list based on the whole corpus and the frequency lists based on specific categories (e.g. news, literature, etc.) of the corpus. It has

rank/serial number	word and its semantics	individual raw frequency	cumulative frequency %
138	<i>hen</i> POS	284252	48.051303513501
199	<i>bi</i> COMP	200645	55.674762267727
221	<i>geng</i> COMP	187378	57.875054046681
491	<i>jiao</i> COMP	82870	75.421247841996

Table 6.4: Word frequency comparison (Source: Da, Jun 2005)

word	frequency	semantics gloss
<i>hen</i>	29046232	POS
<i>bi</i>	14464951	COMP
<i>geng</i>	14182556	COMP
<i>bi.jiao</i>	6575380	COMP
<i>jiao</i>	3685900	COMP

Table 6.5: Global (*blog, literature, news, tech, weibo*) word frequency from BLCU corpus

word	frequency per million	frequency raw count
<i>hen</i>	4202.51	566918
<i>bi</i>	1521.78	205288
<i>geng</i>	1108.57	149546
<i>bijiao</i>	143.46	19353

Table 6.6: Word frequency and count from Chinese Lexical Database (CLD)

been claimed that the frequency lists derived from this corpus might be the most reliable frequency lists currently available (Xun et al., 2016).

Further, Sun et al. (2018) present a large-scale lexical database for simplified Chinese, which is called the Chinese Lexical Database (CLD). The CLD provides rich lexical information for 3913 one-character words, 34,233 two-character words, 7143 three-character words, and 3355 four-character words, and is publicly available through <http://www.chineselexicaldatabase.com>. Specifically, For each of the 48,644 words in the CLD, a comprehensive set of frequency measures is provided. I extracted word frequency per million as well as frequency raw count for the relevant words and the results are given in (6.6).

Overall, by eyeballing the description tables above, one can come to the conclusion that the positive form *hen* is more frequently used than the comparative form such as *bi* / *geng*. This frequency asymmetry between POS and COMP further suggests that speakers *tend* to activate the positive alternative more often than they activate the comparative one.

This observation is also in line with the general principle that speakers tend to use less linguistic encoding for more *frequent* or predictable information (Jurafsky et al., 2001; Levshina, 2019). For example, more reduced expressions like pronouns are more frequently produced for more predictable referents (Arnold 2001; Rosa and Arnold 2017; Zerkle et al., 2017).

6.4 Chapter summary

This chapter provides detailed discussion about the cost assumptions about alternatives in (non-scalar) implicatures. The goal is to further contextualize the proposal in Chapter 5 with a more in-depth and comprehensive landscape.

In particular, I adopt the analyses in Zhang (2019) and Zhang & Ling (2020) to describe the structure of *gao* (and its alternatives). Additionally, I discuss the inspirations of connecting (hidden) operators with complexity, which I get from Moracchini's (2018) work.

Following Zhang (2019) and Zhang & Ling (2020), who consider additivity a phenomenon of QUD-based anaphoricity, I maintain that comparative forms, such as English *-er*; *more* and Chinese *bijiao*; *bi*; *geng*, is an anaphora to a QUD and requires that there is a discourse-salient, positive, non-overlap partial answer to the Current Question. This *additional* requirement can be satisfied by accommodation, antecedents, or *than*-expressions, which presumably makes the comparatives costly. Further, I give the semantics of *gao* and its alternatives, based on Zhang's (2019) and Zhang & Ling's (2020) analyses. I analyze POS as a degree argument such that the notion of "cost" gets associated with various factors influencing degree expressions' interpretations.

As to the primitiveness perspective, I find independent evidence from Grano (2012) and Buccola et al. (2021), suggesting that COMP is conceptually less primitive, and it's less likely to get lexicalized cross-linguistically, compared to POS. Moreover, independent evidence from word frequency offers

additional evidence for the cost asymmetry of the comparative alternative and the positive alternative.

So far the core proposal as well as a zoom-in discussion about its components have been given. The next two chapters serve the purpose of providing proof of concept. In particular, Chapter 7 simulates the competition-based disambiguation model using the probability pragmatics framework RSA. This chapter also serves the purpose of formalizing and visualizing the derivation steps illustrated in Chapter 5. Chapter 8 provides a way to show that artificial language learning experiments can be used to quantify cost when examining *conceptual* alternatives (see Chapter 6).

CHAPTER 7

SIMULATING THE PROPOSAL

This part details the more recent Rational Speech Act (hence forth RSA) models and their implementations (Frank and Goodman 2012; Bergen et al. 2016). The main goal is to use the RSA framework as a proof of concept as well as a disambiguation tool to formally describe the proposal illustrated in Chapter 5 and Chapter 6.

7.1 Cost in RSA models

Thus far, the pragmatic reasoning process in prose was described with *some* LFs and Bayes formulations. Recent progress in simulation-based probabilistic programs have paved the way for advances in formal, implementable models of pragmatics. Among these models, RSA has the most sophisticated way to capture the proposed disambiguation-based competition model, in particular, the costs in a *gradient* way. Plus, recent studies on RSA models by Franke & Bergen (2020) stress that RSA is a particularly useful tool to model *disambiguation*. It quantifies alternatives through a cost function to measure “proximity”. Intuitively, that could work for the non-scalar “gao” versus “hen gao” puzzle too.

Specifically, probabilistic pragmatics tools (Frank & Goodman 2012) are used to implement this proposal, in which *U* means utility, *P* means Bayesian probability, symbol $>$ reads “greater than”, and pragmatic speaker is represented by subscript *s*. Presumably speakers maximise a utility that is increasing with the informativity of an utterance but is decreasing with its cost.

- (1) *Kai gao.*
Kai tall
“Kai is tall/ taller”

- (2) *Kai hen gao.*
Kai HEN tall
“Kai is tall”

- (3) *Kai bi Anna gao.*
 Kai than Anna tall
 “Kai is taller than Anna”

Consider sentences (2,3) as alternatives of the utterance (1), the step-wise iteration for the disambiguation-based competition model is modeled below under the RSA framework:

1. In scenario (a) where the speaker intends to communicate comparative “taller”, (11) *bi gao* is more informative than (9) *gao*, but $\text{cost}(11) > \text{cost}(9)$.
2. In scenario (b), speaker intends positive “tall”, (10) *hen gao* is more informative than (9) *gao* and $\text{cost}(10) > \text{cost}(9)$, but the *difference* is small, compared to that in scenario (a).
3. With flat priors, $P_S((9) \mid \text{“taller”}) > P_S((9) \mid \text{“tall”})$. Speaker has a particular lexical intention in mind.
4. Because speaker was more likely to use (9) *gao* in the comparative “taller”-situation than in the positive “tall”-situation, listener increases the probability of interpreting (9) *gao* as comparative “taller”.
5. With flat priors, $U((9) \mid \text{“taller”}) > U((9) \mid \text{“tall”})$. Thus listener infers, from hearing (9) *gao*, that the speaker most likely intends to communicate the comparative meaning “taller”.

7.2 Stepwise implementation – vanilla RSA

This section uses the RSA framework to formalize and implement the disambiguation-based proposal. The RSA framework is considered in this dissertation as a tool to simulate the proposed hypotheses. Through manipulating and visualizing the parameters and function variables, one can easily identify the role of “cost” in pragmatic reasoning, and we will be capable of making further predictions in a general communicative setting.

Moreover, this part focuses on the vanilla RSA model, which relies on a general Gricean based picture of a pragmatic speaker trying to maximize the amount of information conveyed by a particular

utterance, and a pragmatic listener interpret this utterance based on such information-optimizing behavior. Consequently, the model predicts a probability, indicating the degree of belief a pragmatic listener assigns to each world state after hearing any given utterances.

Here is a condensed overview of how this model works conceptually. Suppose there is a speaker and a listener in a shared context. R is a set of states (worlds, referents, propositions, etc.). M is a set of messages. The speaker has a target referent $r^* \in R$, and the speaker *intends* to choose a message from M that will lead the listener to pick r^* as the target. Presumably the RSA connections to Grice include the following:

- (i) Quality - all agents assign 0 probability to false utterances
- (ii) Quantity - the speaker favors informative utterances
- (iii) Manner - the cost function C is sensitive to Katzirian complexity, while the lexical intention parameter concerns semantic complexity (i.e., ambiguity). It amplifies the pragmatics with caution.
- (iv) Relevance - the referent prior is supposed to help.

Concretely, let's start with a simple situation in which there are two utterances (i.e., messages):

(4) Messages:

- | | |
|--|--|
| <p>a. <i>Hong.che gui.</i>
 red.car expensive</p> | <p>b. <i>Hong.che hen gui.</i>
 red.car HEN expensive</p> |
|--|--|

The truth conditions of the messages are illustrated in the pictures (7.1) and (7.2). In scenario r_1 , the red car is more expensive than the other two contextually salient cars, namely the COMP reading is true. By contrast, in scenario r_2 , the red car is both more expensive than the *standard* given by the Judge AND more expensive than the contextually salient cars. In other words, the POS reading and the COMP interpretation are both true.

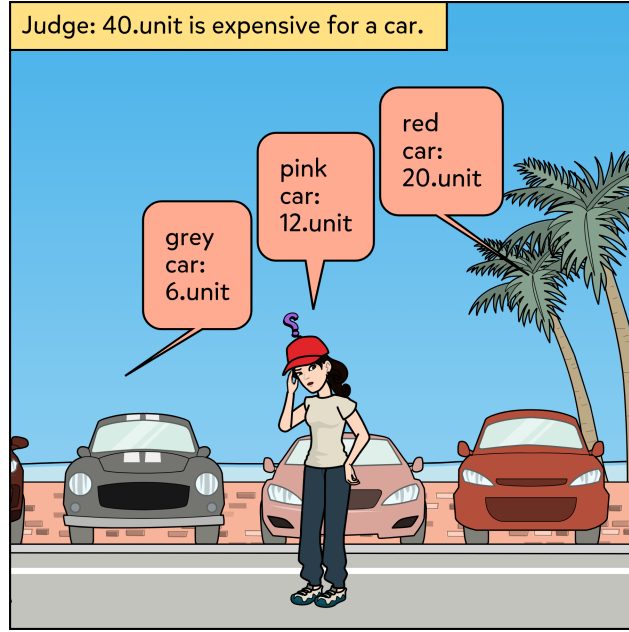


Figure 7.1: scenario r_1 : COMP true POS false

Similar to the gradable adjective *gui* “expensive”, *gao* “tall” can be illustrated in the same way. For this dissertation, I assume that distributionally, *gui* and *gao* share the syntactic and semantic features.

Inspired by and Franke & Bergen (2020), I use RSA framework to simulate the pragmatic reasoning process. The outputs are illustrated below. (5a) is a truth conditions table. Since “*hen gao*” never means COMP, it’s assigned 0 to a COMP world and 1 to POS world. Similarly, “*bi NP gao*” is assigned 1 for a COMP world state and 0 for POS. Given that “*gao*” is ambiguous, 1 is assigned for both world states. (5b) illustrates the priors. Since we assume flat prior, 0.5 is assigned to each world state.

- (5) a. $\llbracket \cdot \rrbracket: M \mapsto R \mapsto \{0,1\}$ is a semantic interpretation function

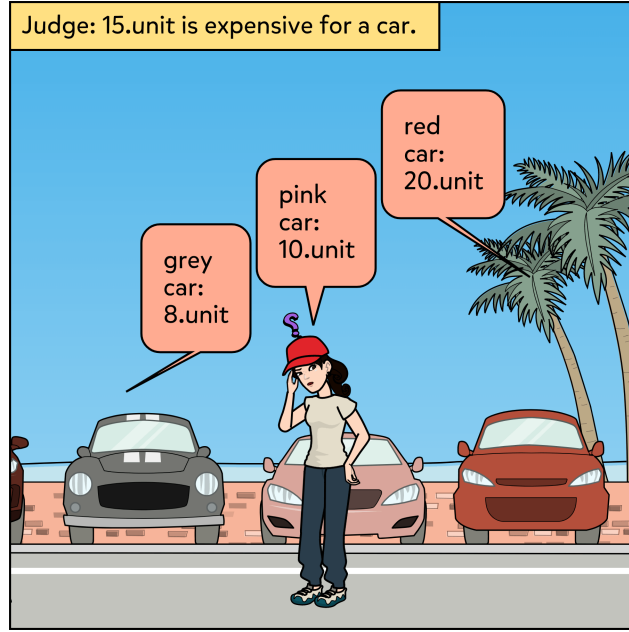


Figure 7.2: scenario r_2 : COMP true POS true

	COMP	POS
“ <i>hen gao</i> ”	0	1
“ <i>bi NP gao</i> ”	1	0
“ <i>gao</i> ”	1	1

- b. $P: R \mapsto [0,1]$ is a prior probability distribution over states

COMP	0.5
POS	0.5

- c. $C: M \mapsto \mathbb{R}_{\leq 0}$ is a cost function on messages.

“ <i>hen gao</i> ”	-1
“ <i>bi NP gao</i> ”	-2
“ <i>gao</i> ”	0

(5c) illustrates the cost function assignment, in which bigger negative values indicate greater cost. Given that the positive alternative “*hen gao*” is less costly than the comparative “*bi NP gao*”, -1 is assigned to “*hen gao*”, whereas -2 to “*bi NP gao*”. “*gao*” is assigned 0 cost.

Before we dive into the derivations, the intuition behind Bayesian update/reasoning is as follows. First, *speakers* choose what to utter based on how costly those utterances are and how likely a listener hearing that utterance would be to deduce what situation the speaker is trying to describe. In other words, cooperative speakers avoid unnecessarily complicated messages, and gravitate towards messages that *precisely* describe the situation at hand. Obviously in most cases, these pressures are in conflict, so speakers are generally forced to consider the relative utilities of various utterances, essentially an information-to-cost ratio.

Second, *listeners* interpret utterances based partially on their prior expectations about the world and in part on how likely a (presumably cooperative) speaker would have been to choose that utterance while trying to describe a given situation. Put otherwise, listeners perform Bayesian inference to update their beliefs about the world, given a model of how the facts condition speakers’ decisions.

The above two ideas are essentially for simulating optimal agents. However, sometimes in real time communication, there may be a number of additional sources of uncertainty. To name just a few, the speaker might not have complete knowledge of the situation they are describing. The listener might not completely know what question the speaker takes themselves to be answering. As a result, they may not know how to partition a set of possible worlds into the relevant hypothesis space. Even the message itself may contain implicit content.

In general, the simulation starts with a literal listener, who draw inference from the speaker’s utterance. Thus, the given is the utterance/message. The next step concerns the pragmatic speaker, who has a target meaning in mind. The speaker hopes to use a message to guide the listener to that target inference. The final agent is the *pragmatic* listener, who is just like the literal listener – making inference from the utterance. But this pragmatic listener is derived from the pragmatic speaker. Thus, hypothetically, the pragmatic listener should demonstrate “rational speech act”.

In particular, (6a) is the literal listener. It defines a conditional probability distribution of referents given messages. The numerator is the product of the semantics and the priors. The denominator is summing over those P_{Lit} values, and the prior is multiplied in.

$$(6) \quad a. \quad P_{Lit}(r \mid m) = \frac{\llbracket m \rrbracket(r) \cdot P(r)}{\sum_{r' \in R} \llbracket m \rrbracket(r') \cdot P(r')}$$

P_{Lit}	COMP	POS
b. “ <i>hen gao</i> ”	0	1
“ <i>bi NP gao</i> ”	1	0
“ <i>gao</i> ”	0.5	0.5

(7) is the full speaker matrix with its corresponding definition. To get the pragmatic speaker, transpose the literal listener matrix: swapping rows and columns in (6b) and we get (7b).

$$(7) \quad a. \quad P_S(m \mid r) = \frac{\exp(\alpha \cdot (\log P_{Lit}(r \mid m) + C(m)))}{\sum_{m' \in M} \exp(\alpha \cdot (\log P_{Lit}(r \mid m') + C(m')))}$$

P_S	“ <i>hen gao</i> ”	“ <i>bi NP gao</i> ”	“ <i>gao</i> ”
b. COMP	0.0	0.068	0.932
POS	0.351	0.0	0.649
$\alpha_{\text{lexical intention}}$	2		

To better understand the simulation outputs, the relationship of the relevant parameters and Bayesian probability are demonstrated in Figure 7.3 and Figure 7.4. Specifically, Figure 7.3 illustrates that as the cost of the alternative ‘*bi gao*’ goes up, it more and more probable that the pragmatic speaker uses the ambiguous ‘*gao*’ given the meaning COMP. Figure 7.4 demonstrates that as the lexical intention increases, the probability of pragmatic speaker using GAO given COMP goes up too. Note that the x axis in Figure 7.3 is the absolute value. Negative number means larger cost, and the raw negative values are transformed into absolute values. This serves the purpose of easing visual inspection, and showing that Figure 7.3 and Figure 7.4 demonstrate the same trend.

(8a) is the final agent: the pragmatic listener. To derive it, transpose is needed again, so that the messages are back on the rows and the referents on the columns. This agent is parallel to the

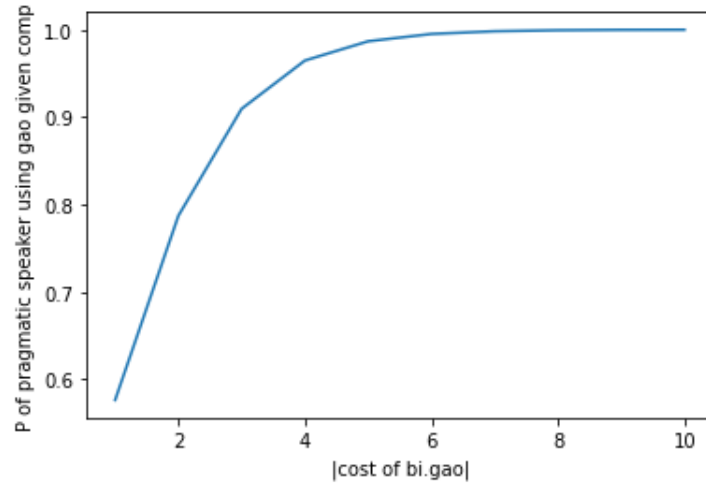


Figure 7.3: $C(\text{hen gao})=-1$; $C(\text{gao})=0$; $\alpha(\text{LI})=1$

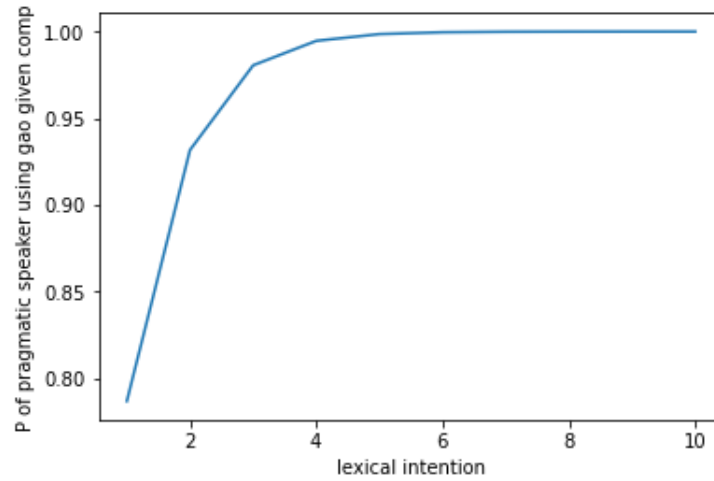


Figure 7.4: $C(\text{hen gao})=-1$; $C(\text{bi gao})=-2$; $C(\text{gao})=0$

literal listener. Crucially, (8a) uses the values from the pragmatic speaker (7a), rather than the truth conditions (6a). The effect is that this *recursion* is three levels deep.

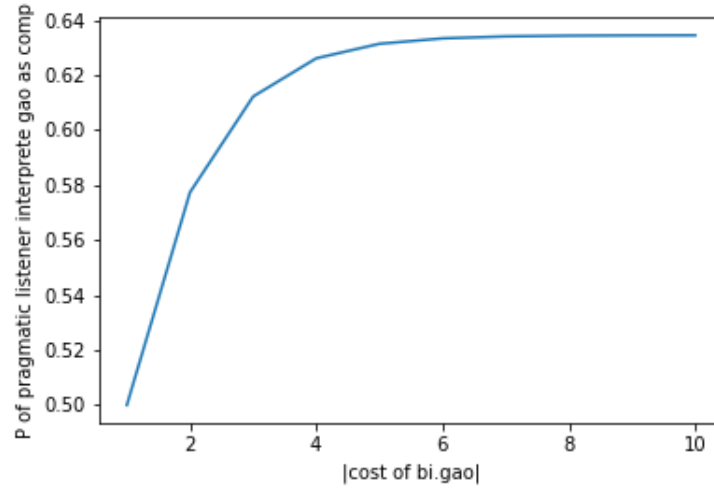


Figure 7.5: C(hen gao)=-1; C(gao)=0; α (LI)=1

$$(8) \quad a. \quad P_L(r \mid m) = \frac{P_S(m \mid r) \cdot P(r)}{\sum_{r' \in R} P_S(m \mid r') \cdot P(r')}$$

	P_L	COMP	POS
b.	“hen gao”	0	1
	“bi NP gao”	1	0
	“gao”	0.6	0.4

(8) reflects the desired implicature, which is a bias for COMP given “gao”. the salient meaning of “gao” (in a plain upward entailing context) is comparative. Figure 7.5 illustrates the relationship of the cost of the alternative ‘bi gao’ and the probability of pragmatic listener interpreting ‘gao’ as COMP. The Bayesian probability increases as the cost goes up.

LI can amplify the pragmatics in a way that we can get a .6 versus .4 contrast. However once LI passes the threshold (LI=2), PL(comp|gao) starts going down, as shown in Figure 7.6. The listener still remembers that “gao” is ambiguous, which should have been avoided. Given “gao”, the degree of favoring the disambiguity inference of COMP over POS is getting smaller and smaller, since ambiguous utterances are less likely to influence the listener’s reasoning process.

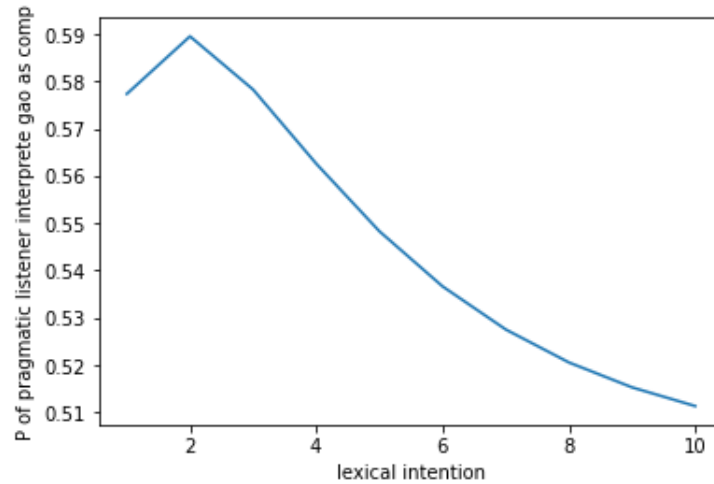


Figure 7.6: $C(\text{hen gao})=-1$; $C(\text{bi gao})=-2$; $C(\text{gao})=0$

7.3 RSA models as disambiguation simulation tools

Recently, various versions of RSA models are developed. This part of the discussion provides a condensed yet comprehensive overview of novel RSA models. The goal is to show that these general developments are on the right direction in terms of accounting for disambiguation inferences.

Franke & Bergen (2020) studies grammatically generated scalar-implicature readings. Examples are given in (9). When hearing utterance (9a), listeners may take (9b) as the intended meaning, since speakers would have but did not use the more informative alternative in (9c). The mainstream explanation is that the implicature reading in (9b) arises through combining the literal meaning of (9a) and the negation of (9c), and the consequence is (9d), which is equivalent to (9b).

- (9)
- a. I own some of Johnny Cash's albums.
 - b. I own some but not all of Johnny Cash's albums.
 - c. I own all of Johnny Cash's albums.
 - d. I own some of Johnny Cash's albums, and it's not true that I own all.

However, the traditional Gricean accounts could hardly explain that (10a) intuitively communicates (10b). (10d) is the result of conjoining the negation of alternative in (10c) with the literal

meaning of (10a). In other words, the problem is that (10d) expresses a different meaning than (10a). It's stronger. (10d) says that some soldier showed *some* (but not all) signs, and no soldier showed all, whereas (10a) says merely that some soldier showed *some* but not all signs. (10a), but not (10d), is compatible with some other soldier showing all signs.

- (10) a. A soldier showed some signs of malaria.
- b. A soldier showed some but not all signs of malaria.
- c. a soldier showed all signs of malaria.
- d. A soldier showed some signs of malaria, and it's not true that a soldier showed all signs.

Grammaticalism, e.g., Chierchia, Fox, & Spector (2012), argues that pragmatic inferences are generated through the exhaustification operator *Exh* which, roughly speaking, negates the alternatives. Consider the nested Aristotelians, exemplified in (11). *Exh* can occur in matrix position (*M*, applying to the whole sentence), and it can apply to the outer (*O*) or inner (*I*) quantifier, as illustrated in (12).

- (11) {None | Some | All} of the aliens drank {none | some | all} of their water.

- (12) $Exh_M [Exh_O(Q_O) \text{ of the aliens drank } Exh_I(Q_I) \text{ of their water}]$

Franke & Bergen (2020) attempts to explore how far a combination of Gricean ideas about efficient communication and a grammatical approach that generates potential pragmatic readings could complement each approach. They therefore consider four RSA models, which differ in terms of how grammatically supplied ambiguity is integrated, as well as how semantic ambiguity affects speakers' and listeners' decisions. Figure 7.7 illustrates conceptually how speakers' reasoning proceeds under different models.

Generally, first, the vanilla RSA of Frank & Goodman (2012) associates each utterance with its literal semantics only. Second, lexical uncertainty model of Potts et al. (2016) assumes that speakers have a fixed lexicon, yet this model takes a larger subsets of readings. Third, lexical intentions model and global intentions model of Franke & Bergen (2020) assume that speakers are aware of the full semantic ambiguity and they *flexibly* use it to maximize information flow in communication.

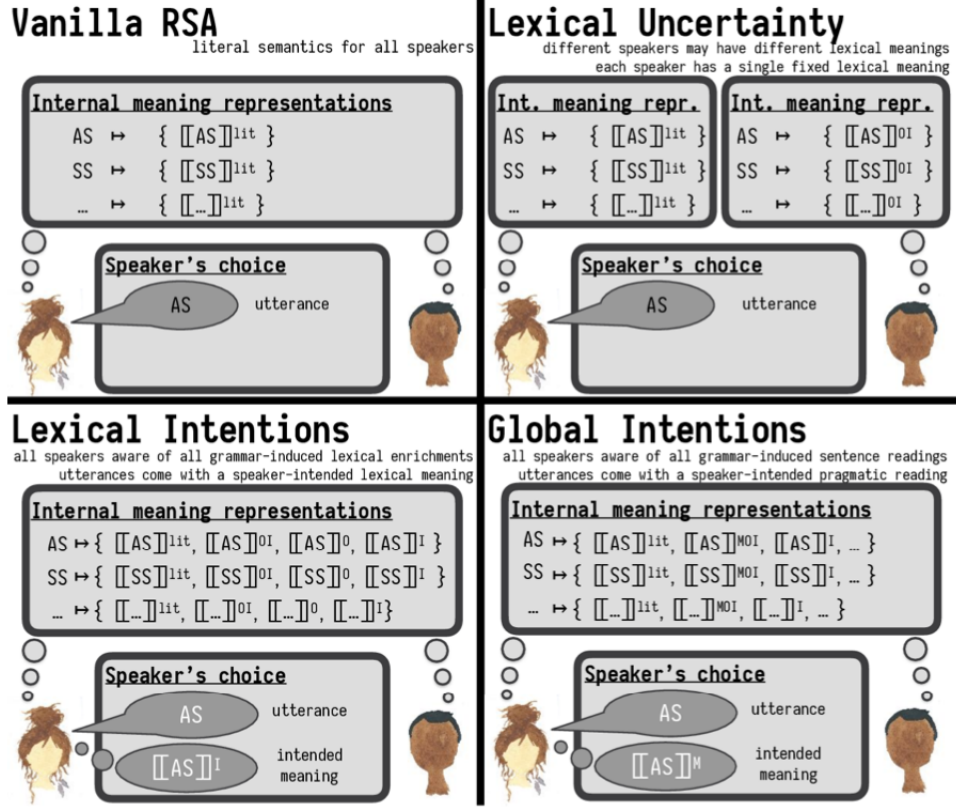


Figure 7.7: Franke and Bergen (2020:e81): Schematic representation of main conceptual differences in the speaker production part of the four models

Specifically, Franke and Bergen states that the Bergen et al's (2016) LEXICAL UNCERTAINTY (LU) model extends the vanilla RSA model through incorporating the listener's potential uncertainty about the lexical meaning that the speaker assigns to certain expressions. The listener's reason that given that the speaker said u and u has two interpretations, which pair $\langle t, l \rangle$ of a state t and mental lexicon l is most likely to have caused THIS speaker to have produced u .

$$(13) \quad P_{S_I}(m \mid t, l; \alpha) \propto [P(t \mid \llbracket m \rrbracket^l)]^\alpha$$

(13) says that the speaker selects utterances u based on a semantic interpretation of utterances influenced by the speaker's lexicon l . As illustrated in equation (13), the listener does not know the speaker's l , and they infer which state-lexicon $\langle t, l \rangle$ pairs are likely to drive the speaker to produce the utterance.

$$(14) \quad P_{L_I}(t \mid \mathbf{l}, t; \alpha) \propto P(t) \cdot P(\mathbf{l}) \cdot P_{S_I}(m \mid t, \mathbf{l}; \alpha)$$

Equation 15 defines the listener's beliefs for each $\langle t, \mathbf{l} \rangle$ pair. It's summing over all those P_{L_I} values, starting from \mathbf{l} .

$$(15) \quad P_{L_I}(t \mid m; \alpha) = \sum_{\mathbf{l}} P_{L_I}(t, \mathbf{l} \mid m; \alpha)$$

Potts et al. (2016) use (15) to account for truth-value judgment task data. Crucially the production rule P_{S_I} defined in (13) makes predictions for a given lexicon \mathbf{l} . By contrast, Lassiter & Goodman (2017) and Franke & Bergen (2020) define a speaker S_2 who reasons about pragmatically adequate utterance choice based on the state-interpretation of L_I . As a consequence, the final derivations of the LU production P_{S_2} and comprehension P_{S_2} are given in equations 16.

$$(16) \quad \begin{aligned} \text{a. } P_{S_2}(m \mid t; \alpha) &\propto [P_{L_I}(t \mid m; \alpha)]^\alpha \\ \text{b. } P_{L_2}(\mathbf{l} \mid m; \alpha) &\propto P(t) \cdot P_{S_2}(m \mid t; \alpha) \end{aligned}$$

When carrying an ambiguous lexicon, speakers have different LEXICAL INTENTIONS (LI). In other words, the intended meaning for each utterance of a word may vary. In this LI model, listeners can then try to recover the speaker's reasoning based on speaker's lexical intentions. Equations 17 formalizes these intuitions.

$$(17) \quad \begin{aligned} \text{a. } P_S(m, \mathbf{l} \mid t; \alpha) &\propto [P(t \mid \llbracket m \rrbracket^{\mathbf{l}})]^\alpha \\ \text{b. } P_L(t, \mathbf{l} \mid m; \alpha) &\propto P(t) \cdot P_S(m, \mathbf{l} \mid t; \alpha) \end{aligned}$$

The GLOBAL INTENTION models (GI) resembles the LI model, and it integrates all parses p . Equation 18 illustrates that the GI model gives a lot power to speakers. They use the rich ambiguities provided by their language's grammar to flexibly *intend* to express this or that reading, depending on the communication goal.

$$(18) \quad \begin{aligned} \text{a. } P_S(m, \mathbf{p} \mid t; \alpha) &\propto [P(t \mid \llbracket m \rrbracket^{\mathbf{p}})]^\alpha \\ \text{b. } P_L(t, \mathbf{p} \mid m; \alpha) &\propto P(t) \cdot P_S(m, \mathbf{p} \mid t; \alpha) \end{aligned}$$

	$\llbracket \text{SS} \rrbracket^M (\text{cost} = 0)$	$\llbracket \text{SS} \rrbracket^M (\text{cost} = 100)$
vanilla RSA	0.22	0.40
lexical uncertainty	0.25	0.45
global	0.28	0.46

Table 7.1: How cost influences pragmatic reasoning under different RSA models

Franke & Bergen (2020) concludes that the GI model has so far the best predictive performance. The main reason for the GI model’s predictive success is the availability of a reading $\llbracket \text{SS} \rrbracket^M = \{210\}$ of the sentence ‘SS’ obtained from inserting an *Exh*- operator at matrix position – “Some drank Some (henceforth “SS”) and **it’s not true that some drank all, or that all drank some**”. Essentially, if this reading is available for the speaker, it yields a very strong reading that uniquely singles out the world state m . Consequently, the listener’s interpretation of ‘SS’ also puts substantial probability on the interpretation M . The full set of grammatically induced readings is needed.

With respect to how cost influences pragmatic reasoning, Franke & Bergen (2020) include a cost term $c(u)$, which is fixed to zero for utterances starting with *some* or *all*, but may be positive for utterances starting with *none*. This is due to the consideration of the low choice rates of sentences starting with *none*. Crucially, they emphasize that the costs here are not to be confused with ‘processing costs’ from the psycholinguistic literature; instead, the costs characterize a general *dispreference* for particular expressions. Costs are subtracted. Table (7.1) illustrates how speakers and listeners reason about costs under different RSA models.

Table 7.1 indicates that “global” (GI) outputs the biggest value (0.28; 0.46) among all three models, regardless of cost, and “lexical uncertainty” (LU) performs not as well as GI but better than the vanilla RSA¹. This gradient increase in overall performance further implies that given that GI model is the most developed model, with QUD and the all the ambiguity readings fully incorporated, RSA models are generally on the right track in simulating pragmatic reasoning about disambiguation

¹This table is generated by me implementing Franke&Bergen’s (2020) programs and plugging in the cost values. It has nothing to do with “gao”, only with the “*some...some*” sentence.

inferences. That said, it appears that no significant difference is observed in between difference models in table 7.1 when cost is plugged in. Thus, I maintain that the vanilla RSA model is sufficient to serve the purpose of formalizing the intuitions of pragmatic reasoning about disambiguation inferences.

7.4 Chapter summary

In this chapter, I have modeled how *cost* influences pragmatic reasoning using the rational speech act framework (RSA). Essentially, implicatures are computed by recursive Bayesian reasoning in which speakers and listeners reason about each others' communicative strategies (Franke 2009, 2011, Frank & Goodman 2012, Goodman & Stuhlmuller 2013, Frank et al. 2016).

Cost-based competition is proposed to explain the derivation of disambiguation inference. A general competition principle is described as follows:

(19) Prediction: The more costly alternative's interpretation should attain.

(19) is what the experimental chapter (Chapter 8) is going to test: If (in a language L) A is ambiguous between meanings p and q , and if p can be unambiguously expressed by an alternative of A that is less costly than any alternative of A that unambiguously expresses q , then (with flat priors) A should typically communicate q , not p .

There are limitations in terms of the RSA simulation breakdowns, for example, the seemingly randomness of the numbers that I plugged in for each iteration steps, and the explanatory power of RSA – is it quantitatively *describing* the reasoning process or could it be analytical as well. Regarding the values assigned to *cost*, the analytical intuition about the *cost* parameter has been extensively discussed in Chapter 6. Regarding the *analytical* features of RSA, a more recent work by Enguehard & Spector (2021) approaches a similar puzzle. Their core formulation is illustrated in 20, further explained by 21.

(20) $\text{Info}_{P_0}(\llbracket m^+ \rrbracket) - \text{Info}_{P_0}(\llbracket m \rrbracket) > c$

- (21) a. Situations where the more costly message is used: the costly message is highly informative compared to the less costly one, so that the disadvantage it has in terms of cost is overridden.
- b. Situations where the less costly message is used: when the prior P_0 is not sufficiently biased so as to make the costly message optimal.

(20) says that the speaker will choose the more costly message only when the gain relative the less costly message exceeds the extra cost of the more costly message. Note that this is built up on scalar inference, in which the speaker *believes* both messages, whereas the Mandarin puzzle about *gao* is not concerned with scalar inference. Rather, it's based on intention. This has been extensively illustrated in section 4.3.

They started with an analytical intuition that presumably less frequent meanings are inclined to be lexicalized less than frequent meanings, therefore, the expected utility of lexicalization would vary. They end up by suggesting that by deriving the difference between the attested and the unattested lexicon from the messages' truth conditions, they have been able to use information-theoretic approach to the logical vocabulary of languages, which would ultimately offer an explanation for Horn's puzzle².

Enguehard & Spector (2021) suggests that in addition to disambiguation, RSA model can be really analytical – no specific values need to be plugged in during the pragmatic iteration. Precise values can be too speculative. This speaks to the simulations in Chapter 7, where specific numbers are assigned to the parameters for the purpose of illustrating *relative* cost, rather claiming that they are the *precise* costs or likelihoods.

Moreover, Enguehard & Spector (2021) sheds light on language universals. Their proposal about the expected utility of lexicalization is very reminiscent of the cross-linguistic observation about the degree expressions. They attempt to explain Horn's puzzle, namely why *some*, *all*, and *no* are lexicalized cross-linguistically, whereas *not all* is not. Their account involves probabilistic

²Horn's puzzle concerns the lexicalization of logical operators: the O corner, corresponding to 'nand' (= not and), 'nevery' (= not every), etc., is never lexicalized cross languages.

informativity, and they provide a model in terms of cost, utility, and frequency. Circling back to the Chinese data and the English data discussed in the current work, as well as in Grano (2012), we have seen that POS/EVAL is arguably lexicalized as *hen*, but COMP is not lexicalized, assuming that *geng/bijiao* does not correspond exactly to COMP (see Chapter 6 and Chapter 2). Why is that the case? One likely reason might be that the expected utility of lexicalizing COMP is significantly less than that of lexicalizing POS. In other words, there's less utility in lexicalizing COMP than lexicalizing POS, hence why we have “hen gao” but not “more gao”. Paradoxically, though, the simple “gao” ends up being interpreted as COMP, via competition with “hen gao”.

In the current discussion, I attempt to boil it down to a focused puzzle about *cost*. Built up on that, I further hypothesize that cost influences pragmatic reasoning, in a way that rational speakers tend to reason about the less-costly alternative whenever appropriate. In the following chapter, I test the hypothesis experimentally. I extend Buccola et al's (2018) artificial language learning paradigm to a speaker's perspective experiment, and use it to gathering independent evidence for earlier claims about competition works when alternatives differ in cost.

CHAPTER 8

EVIDENCE FROM ARTIFICIAL LANGUAGE LEARNING

As discussed in the previous chapters, how costly an alternative is turns out to be an empirical question, which may not always be purely linguistic. Without any independent measure of *cost*, the derivation of inferences could not be predicted. Conversely, this chapter deals with artificial language learning experiments, to gather independent evidence for earlier claims about competition working when alternatives differ in cost. The general goal is to assess what exactly is in competition and what role cost plays in disambiguation-based competition.

8.1 Competition principle and consequence

The proposed competition principle is illustrated below:

- (1) a. The use of ϕ implies that each alternative ψ of ϕ is less appropriate than ϕ , in the sense that ψ is not what the speaker intended.

(Following Buccola et al. (2018))

- b. In a context where both ϕ and a costly alternative ψ of ϕ can be true, an efficient speaker will most likely use ϕ .

(Inspired by Buccola et al. (2018))

Consequently, listeners interpret an ambiguous utterance ϕ as r_1 as opposed to r_2 . This disambiguation is realized via pragmatic reasoning in a way that each alternative ψ of ϕ is less appropriate than ϕ because it is not what the speaker intended.

As a general question, one may wonder: what is the source of the disambiguation-based competition principle? Is it purely linguistic (i.e., a part of the language faculty), or does it go beyond language? From a Gricean (and RSA) standpoint, we may expect it to be part of general rational behavior. Thus, we may expect to detect it even in non-linguistic tasks, such as those involved in artificial language learning.

I hypothesize that, as in (1a), in a highly artificial experiment, in which there is no sense of conversation going on, if even then humans still apply the proposed competition principle, then that's evidence suggesting that proposed disambiguation-based competition principle is beyond linguistic processes. This hypothesis (1a) has been extensively studied in (Buccola et al., 2018). The current investigation thus focuses on (1b), which is an implication as well as an extension of (1a).

Specifically, (1b) is a competition principle proposed from the speaker perspective, and it directly speaks to the degree expressions phenomena observed in the Chinese data. Suppose ϕ has two interpretations, corresponding to two different alternatives, and one of which ψ is significantly more costly than the other ψ' . In a context where both ϕ and ψ can be true, a rational speaker would favor ϕ over ψ because she believes that the listener will be cooperative and capable of drawing the target inference. By contrast, in a context where both ϕ and ψ' can be true, such preference would not be very strong because now the alternative ψ' is significantly *less costly*. A speaker may even favor ψ' over ϕ , since for this scenario it's not a matter of cost but a matter of specificity. ϕ and ψ' do not differ dramatically in terms of cost, yet ϕ is ambiguous and less specific than ψ' .

The speaker perspective experiment can provide independent evidence for the Chinese phenomena. Recall that Chapter 3 already shows that *gao* is ambiguous between 'tall' and 'taller', and Chapter 5 proposes that the 'taller' reading of *gao* surfaces because there is a cost asymmetry between the 'taller' alternative of *gao* (e.g., "bijao gao", "bi NP gao"), and the 'tall' alternative of *gao* (e.g., "hen gao").

Artificial language learning (ALL) experiments have a long history and are becoming more and more common and valued as a source of investigation into linguistic universals and cognitive preferences (Adger et al., 2019; Buccola et al., 2018; Culbertson, 2012; Culbertson & Schuler, 2019; Motamedi et al., 2019). Against this background, this speaker perspective experiment will provide proof-of-concept evidence for the proposal through measuring the extent to which *cost* influences pragmatic reasoning.

The current experiment is built up on Buccola et al. (2018). Specifically, Buccola et al. (2018) investigated how a nonce word can be pragmatically enriched when there is a single, more

specific/informative alternative available, and this enrichment disappears when a second, symmetric alternative is introduced. The role of *cost* is not discussed in Buccola et al (2018). The novelty of the current experiment is that it introduces cost as a factor that can be manipulated. Moreover, the current experiment simulates the speaker’s perspective.

8.2 Experiment: the speaker perspective

Different from Buccola et al. (2018)’s experiment with a listener perspective, in which participants play the role of a “listener” and reason about the meaning of an utterance, participants reason about utterances with respect to a given intended meaning in the speaker perspective experiment.

Compared to the listener version, the speaker version of the experiment directly speaks to the (vanilla) RSA model from Chapter 7. The participants, who are presumably rational speakers, are provided with an illustration of the *intended* meaning, and then choose an utterance that they consider appropriate. The Bayesian reasoning process is illustrated in the diagram (8.1) below.

Diagram 8.1 illustrates the predicted reasoning process for a rational efficient agent. For a simple yet ambiguous utterance that has two interpretations corresponding to two specific alternatives with significantly different costs (i.e., Alt_c and Alt_{lc}), the ambiguous utterance should hypothetically be favored over the specific yet costly alternative, given a costly message. Conversely, given a less-costly message, such preference is expected to be relatively small. This contrast is demonstrated in the top two green boxes. This is also the main prediction of the current experiment on *cost*.

As the reasoning proceeds to the next iteration, it becomes more and more probable that the ambiguous utterance is used in a costly situation than in a less-costly situation. This is illustrated by the bottom two green boxes. As a consequence, a rational efficient listener is more likely to infer the costly meaning from the ambiguous utterance, rather the less-costly meaning. This part of the hypothesis is listener-centered, thus it’s not testified in the current experiment. I leave it for future research. For a RSA-style mental simulation of the listener perspective, see Chapter 7.

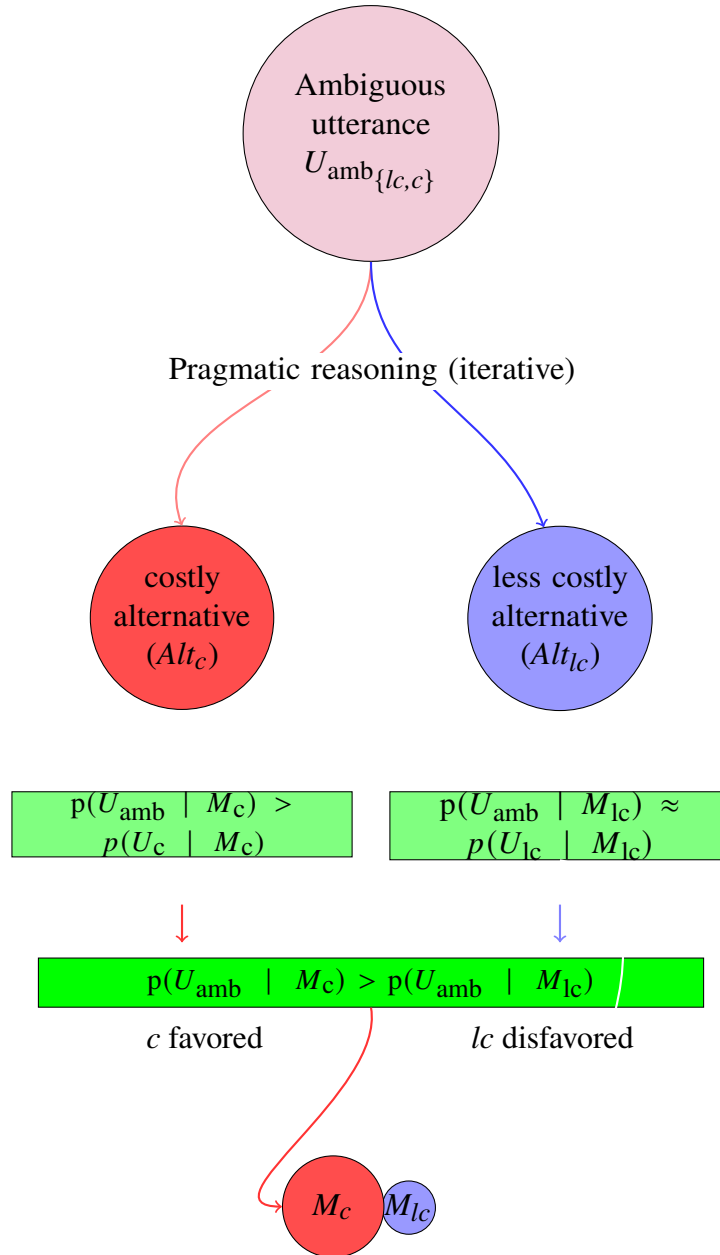


Figure 8.1: Predicted reasoning process – rational speakers

8.2.1 Task summary and hypothesis

The task is couched in a cover story about humans making contact with Martian aliens. Aliens want to teach humans (i.e., experiment participants) their language. First, the Martians teach participants phrases, and participants receive feedback as they learn. Afterwards, participants play a Martian game in which the aliens test their knowledge of the phrases they have learned. Martians are friendly. They are trying their best to teach participants their language. There is no deception or tricks throughout the experiment.

The goal of the task is to learn two classes of nonce phrases: *costly*-object-type phrases (e.g., *kad shegmifanzub*) versus *less costly*-object-type phrases (e.g., *kad*, *kad vun*). The less costly class phrases, such as *kad*, can be used by themselves to refer to an object (e.g., triangle, square, etc.). Even when they are used with an additional phrase to refer to an object, the complexity of that additional phrase is trivial (e.g., *kad vun*). By contrast, the costly class phrases *have to* be used with a complicated phrase to refer to an object (i.e., costly object¹), for example, *kad shegmifanzub*, making them significantly complex in terms of length, structure, and phonetics. Hence they are called *costly*-object-type phrases, relative to the *less-costly*-object-type phrases.

Crucially, there are two sub-types for the *less-costly*-object-type phrases: (i) a simple phrase that's always used by itself, but it has two interpretations. It can be used to refer to both the *costly* and the *less-costly* types objects. This implies that it can bring in ambiguities during communication; (ii) a simple phrase that's used with another simple phrase, and it specifically refers to one particular object (i.e., the *less-costly*-object) — *slightly* more complex than sub-type (i) yet more specific than sub-type (i). The experiment consists of two parts: learning part and gaming part. In the learning part, Martian aliens give participants a new phrase, then participants pick an object out of three objects. Participants receive feedback.

For the gaming part, the aliens play a game with participants. In this game, participants practice speaking with Martians. Participants are asked to use phrases they have learned in order

¹The notion of a 'costly object' is a paraphrase for a situation in which referring to an object costs effort and energy. The cost is quantified as the extent to which the linguistic representation counts as 'complicated'.

to communicate with the Martians. Participants see three objects, one of which is circled. Their job is to communicate the circled object to the Martians by using one of the phrases they have learned. Martians see the same three objects that participants see. Participants will pick a phrase that Martians will see. Participants are explicitly told that Martian speakers always try to be as *efficient* with their language as possible, and they tend to pick the *shortest* phrases that can do the job. Participants are instructed to try to communicate like a Martian. They do not receive any feedback in the gaming part.

Regarding group comparison, there are two phases for the experiment group. They learn both *costly* and *less-costly* types of phrases in phase 1. In phase 2, they only learn *less-costly* type of phrases. There is only one phase for the control group, in which they never learn the *costly*-type of phrases. Both groups of the participants will be tested on the phrases they have learned in the learning part.

The hypothesis for control group's response in the gaming part is as follows: if they implement the proposed competition principle, then they should go for the *less-costly*-object-type phrases that are specific. This tendency should be stronger in control group's phase 1 than in experimental group's phase 1. Specifically, the hypothesis for experimental group's response in the gaming part in phase 1 is as follows: (a) in a scenario where the *costly*-object is intended, the available candidates for a speaker include the less-costly-object-type phrases that are ambiguous, and the costly-object-type phrases that are specific. If cost influences speaker's reasoning and if the proposed competition principle gets implemented, then participants should go for the less-costly-object-type phrases that are ambiguous. (b) In a scenario where the *less-costly*-object is intended, the available alternatives are the less-costly-object-type phrases that are ambiguous, and the less-costly-object-type phrases that are specific. Now the participants should go more for the less-costly-object-type phrases that are specific. This is due to the fact that in this scenario, both alternatives are *less-costly*. The cost difference between the two is trivial. As a consequence, participants should reason about the phrase that's not ambiguous. In phase 2, since participants only learn less-costly-object-type phrases, regardless of which objects are intended, they should always go for the less-costly-object-type

group	phase 1	phase 2
EXP	COST (<i>lc</i> vs. <i>c</i>); AMBIGUITY (<i>amb</i> vs. <i>lc, c</i>)	AMBIGUITY (<i>amb</i> vs. <i>lc</i>)
CTRL	AMBIGUITY (<i>amb</i> vs. <i>lc</i>)	N.A.

Table 8.1: Groups, Conditions and Parameters

	M(<i>less-costly</i> object)	M(<i>costly</i> object)
U(<i>amb</i> phrases)	1	1
U(<i>lc</i> phrases)	1	0
U(<i>c</i> phrases)	0	1

Table 8.2: Literal semantics

	EXP	CTRL
phase 1	$P_{\text{case1}}(U_{\text{amb}} M_c) > P_{\text{case2}}(U_{\text{amb}} M_{lc})$	$P_{\text{case1}}(U_{\text{amb}} M_{lc}) \approx P_{\text{case2}}(U_{\text{amb}} M_{lc})$
phase 2	$P_{\text{case1}}(U_{\text{amb}} M_{lc}) \approx P_{\text{case2}}(U_{\text{amb}} M_{lc})$	N.A.

Table 8.3: Group comparisons, phase comparisons, and predictions

phrases that are specific.

To sum up, the experimental group is designed to measure the extent to which *cost* influences speakers’ pragmatic reasoning, thus phase 1 has the asymmetry between costly and less costly types of phrases, whereas phase 2 removes the *cost* factor and every alternative is *less-costly*.

	learning	gaming
phase1	<i>amb</i> 4; <i>lc</i> 4; <i>c</i> 4; <i>filler</i> 4 (16)	<i>amb-context-c-object</i> , <i>amb-context-lc-object</i> , <i>unamb-context-c-object</i> , <i>unamb-context-lc-object</i> (4x2=8)
phase2	same 12 (4 <i>c</i> excluded) in phase 1; <i>new-lc</i> 8 ² (20)	<i>amb-context-c-object</i> , <i>amb-context-lc-object</i> , <i>unamb-context-c-object</i> , <i>unamb-context-lc-object</i> (4x2=8)

Table 8.4: Calculation of trials and blocks

8.2.2 Method

Participant 60 adult Native English speakers are recruited from Prolific and compensated \$1.5-\$2 for their participation (rate: \approx \$6.57 per hour). 79 attempts are collected from Prolific and saved on Pavlovia, among which 19 attempts are labeled as “returned” by Prolific. Prolific keeps sending out experimenters’ requests until there are 60 attempts marked as “complete” and “awaiting review”. Among the 19 “returned” participants’ responses, 15 of which are excluded due to incomplete input, inaccurate input, or timed-out (duration longer than 45 minutes).³

For the experimental group, 6 participants’ responses are excluded, as the number of ambiguous-type-phrase response is smaller than 4, which is a good indicator suggesting that they tend to ignore the instructions (see subsection 8.2.2 for details), they are negligent in reasoning about *competition* or *efficiency* (see subsection 8.2.2 for details), and almost always go for the specific phrases. This calculation criteria is further verified through going through all of the control cases. The 6 participants are being *consistently* specific throughout the experiment, in a sense that they always chose ‘kad vun’ or ‘kad shegmifanzub’ and never chose the ambiguous phrase ‘kad’ even in control cases.

58 participants’ responses are used for data analysis (see Table 8.7). They are randomly assigned to one of the two groups, namely the experimental group (N = 29) and the control group (N = 29). Since participants are forced to learn the phrases in order to complete the experiment, and the task is relatively easy, nobody is excluded.

Procedure Participants see the following instruction in the beginning of the experiment:

(2) Instruction – the beginning of the experiment:

“*BACKSTORY:*

You have just made contact with Martian aliens! Now they want to teach you their language.

First, the Martians will teach you phrases, and you will receive feedback as you learn.

³All materials and results can be found at the osf repo https://osf.io/b6dwe/?view_only=aa691ac230db458ba676158909a70dc4

Afterwards, you will play a Martian game in which the aliens test your knowledge of the phrases you have learned. Martians are friendly. They are trying their best to teach you their language!”

Participants see the instruction below in the beginning of the gaming part:

(3) Instruction – the beginning of the gaming part:

“GAMING PART (speaker):

Good job! You have successfully learned the new phrases! Now the Martians want to play a game with you. In this game, you will practice speaking with Martians. You will use phrases you have learned in order to communicate with the Martians. You will see 3 objects, one of which is circled. Your job is to communicate the circled object to the Martians by using one of the phrases you’ve learned. Martians see the same 3 objects you see. You will pick a phrase that Martians will see. Martian speakers always try to be as EFFICIENT with their language as possible. In general, they tend to pick the SHORTEST phrases that can do the job. Try to communicate like a Martian! You will NOT receive any feedback in the next few trials. Press spacebar to continue.”

In the practice trials, participants see two types of stimuli: (i) a nonce phrase and a collection of objects and their task is to pick a shape; (ii) a collection of shapes with the intended object circled, a collection of nonce phrases, and their task is to pick a phrase. For the real trials, participants see stimuli in exactly the same format as they see in the practice trials.

Phrases are presented together with objects, which are horizontally displayed below the word (see Figure 8.2, Figure 8.3, Figure 8.4, Figure 8.6, Figure 8.5 for illustration). For the learning parts, all phrases are presented in a self-paced-reading format. Participants need to press spacebar to read the phrase, meaning that they have to press more when they learn a complex phrase than they do when learning a less complex one. This serves the purpose that each utterance is inherently associated with *cost*, and participants need to pay more effort to learn a costly phrase (i.e., pressing space-bar more times) than they do to learn a less costly phrase. Participants are asked to type 1, 2, 3 to indicate the object they believe to be associated with the phrase.

For the gaming parts, participants are told to indicate their response by typing a phrase that's already given on the screen. This is due to the consideration that the *cost* factor can hardly influence a computer's performance, since displaying 3 letters is not significantly different from displaying 14 letters for a computer. Nevertheless, for a human, memory retrieval is involved while typing a phrase with respect to an intended meaning (Ferreira, 2003; Ferreira & Henderson, 1998; Christianson, 2016). As a consequence, cost would play a role in the process of accessing the phrase.

The experiment consists of 2 learning phases and 2 corresponding gaming phases for the experimental group, and 1 learning phase and 1 gaming phase for the control group. Participants' answers as well as their response times are recorded on each trial. During the learning phase, participants learn new phrases by receiving feedback after each trial. The feedback is displayed in a transfer screen, with a message at the center of the screen. For correct responses, the transfer screen displays the prompt "Correct!"; for incorrect responses, it displays the prompt "Oops! That was wrong". Correct responses have 0.5s of feedback before the next trial, while incorrect responses have 1s of feedback. This serves the purpose of increasing attention to the task. During the gaming phase, the experiment proceeds without providing feedback. Before participants start the gaming phase, there is a reminder display. It reads "You will not receive feedback for the next few trials!"

At the end of the experiment, participants see a screen saying that "*Please DO NOT CLOSE YOUR BROWSER. Please WAIT for a few seconds until the results are sent and the session is closed.*".

Stimuli For the learning part, there are 4 trial types, which are all randomly interspersed. There are two trial types in the gaming part (see table 8.5 for illustration). For "critical case1", there are two geometric shapes but the circled one is the costly object. By contrast, in "critical case2", there are two geometric shapes and the circled one is the less costly object. As to control cases, there is one geometric shape and two organic distractor shapes, and it's always the geometric shape that's circled.

The object-alternative association is randomized as follows. there are two types of stimuli set (01

	critical case1	critical case2
objects	{distractor, <i>lc</i> -object, <u><i>c-object</i></u> }	{distractor, <u><i>lc-object</i></u> , <i>c-object</i> }
phrases	{ <i>amb</i> , <i>lc</i> , <i>c</i> }	{ <i>amb</i> , <i>lc</i> , <i>c</i> }

Table 8.5: Trials types in the Gaming part

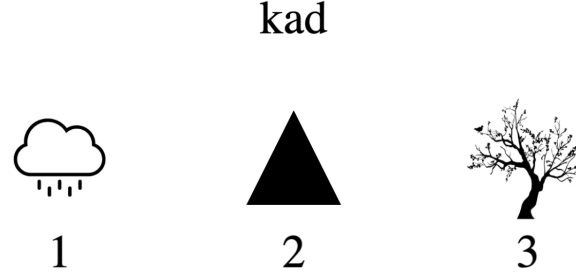


Figure 8.2: Learning trial: ambiguous phrase

- triangle as the costly object; 02 - square as costly object), and participants are randomly assigned in prolific, which is a platform for online participant recruitment for surveys and market research. For instance, suppose for 30 experimental participants, 15 take ‘stim01’, and the other 15 participants use ‘stim02’.

referent(s)	phrases
triangle	{ <i>kad</i> , <i>zud</i> , <i>kad shegmifanzub</i> , <i>kad mof</i> , <i>zud vun</i> }
square	{ <i>kad</i> , <i>zud</i> , <i>kad vun</i> , <i>zud shegmifankub</i> , <i>zud mof</i> }
cloud	{ <i>sot</i> , <i>heth</i> }

Table 8.6: Nonce phrases used in the real experiment stimuli

Figure 8.2 is a sample trial in which participants learn the ambiguous phrase *kad*. When they type “2”, they receive feedback “Correct!”. In a similar trial in which are two organic shapes and a geometric shape (e.g., square), when they see *kad* and choose the geometric shape, they also receive feedback “Correct!”.

Figure 8.3 illustrates a scenario in which participants learn the “costly” phrase *kad shegmifanzub*. Participants see three options: an organic shape and two geometric shapes. Since this costly phrase

kad shegmifanzub

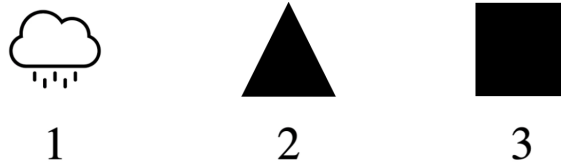


Figure 8.3: Learning trial: costly phrase

kad mof



Figure 8.4: Learning trial: less-costly phrase

is not ambiguous and only refers to a particular geometric shape, when they choose the wrong geometric shape, they receive feedback “Oops! That was wrong”. This set-up is to make sure that participants learn that the costly alternative never has two interpretations, instead it only refers to one shape. Similarly, Figure 8.4 is a sample trial in which participants learn a “less-costly” phrase such as *kad mof*. The options again include *two* geometric shapes, and the goal is to ensure that participants know it only means one particular shape.

Figure 8.6 and Figure 8.5 are sample trials used in the gaming/testing part. In particular, the control cases have two organic shapes (tree and cloud) and a geometric shape, which is always circled. Right below the objects are the three phrases that participants have learned in the learning phase. They are instructed to type a phrase that they think can communicate the circled object, and then click on the screen to proceed to the next trial.

On the top of the screen, the text “Remember to be *EFFICIENT*” constantly shows up to remind

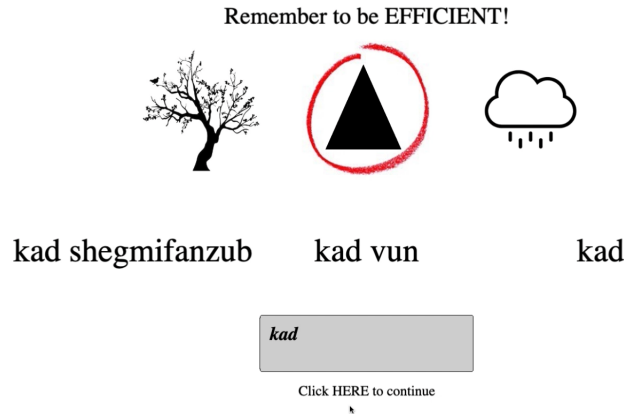


Figure 8.5: Testing trial: ctrl cases

the participants⁴. Under this context, suppose triangle is the costly object, there are presumably two correct answers: both *kad shegmifanzub* and *kad* can be used to refer to the circled object.

By contrast, two geometric shapes are given for the critical cases, and one of them is circled. Three phrases are given right beneath the shapes. This is an “ambiguous” context, in a sense that the two referents of the ambiguous phrase *kad* are simultaneously displayed to the participants, and both of the two alternatives of *kad* are available to the participants. Similarly, participants are asked to indicate their response by typing a phrase. Under this context, suppose triangle is the costly object and is circled, *kad shegmifanzub* and *kad* can both be used to refer to the circled object. The prediction is that compared to the control cases as illustrated in Figure 8.5, I expect to find more ambiguous phrase type responses in the critical cases (Figure 8.6), which would be a good indicator that participants are considering cost while reasoning about the alternatives. The unambiguous yet costly alternative *kad shegmifanzub* can convey the intended meaning. If they choose to type *kad*, a most likely reason is that *cost* plays a role in their pragmatic reasoning, and thus they favor *kad* over the costly alternative *kad shegmifanzub*. Put otherwise, they are being an *efficient* speaker, and relying on the listener’s pragmatic reasoning to lead them to the intended object.

⁴See subsection 8.2.2 for the text from the screen that describes how Martians are “efficient” with their language. They use the shortest phrase that does the job.



Figure 8.6: Testing trial: critical cases

Design Over the course of the experiment, each participant learns six or seven words depending on the condition as illustrated in Table 8.6. There are three transition displays: transition to practice trials, transition to learning trials, and transition to gaming/testing trials. For each transition, there are instructions indicating participants where they are and what they will see next.

Figure 8.7 and Figure 8.8 illustrate the sequence of phases. In particular, for the experiment group, competition is present in the first phase, and gets removed in the second phase. In phase 1, they learn an ambiguous phrase, and two alternatives corresponding to the two interpretations of the ambiguous word. One of the alternatives is costly (henceforth, *c*) and the other is less costly (henceforth *lc*). In the gaming part, they are tested on the phrases that they have learned in two major ways: in a ambiguous context in which one of the interpretations of the ambiguous phrase is intended (c.f. “critical cases”), and in a non-ambiguous context in which one of the interpretations of the ambiguous phrase is intended (“control cases”). In phase 2, the competition is removed. This is realized by having participants learn a new less-costly phrase for one of the geometric shape, so that now, both geometric objects have low-cost alternatives associated with them. The gaming part tests participants’ knowledge of the ambiguous phrase and the two alternatives, which are now both of equal, low cost.

By contrast, the control group is exposed to competition; it’s just that the alternatives that would

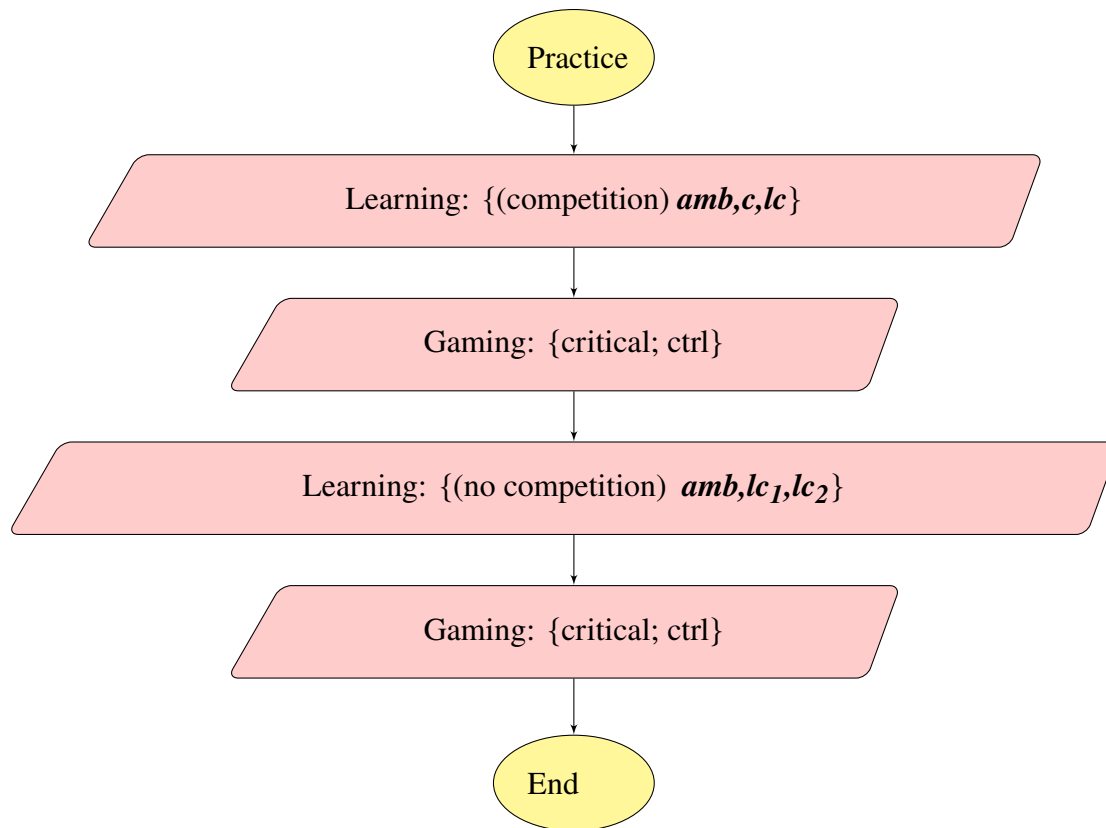


Figure 8.7: Experimental group: sequence of phases

compete are of equal cost, so competition is expected never to manifest. It's worth highlighting that this is an instance of symmetry. They only experience one single phase, in which they learn an ambiguous phrase, and two alternatives that are equally *less-costly*, hence symmetric. The gaming part for control group is the same as for the experimental group phase 1 and phase 2.

A between-experiment design is used. The experimental group learns words that differ in utterance costs; the control group learns words that are equally costly. Moreover, the two groups differ in terms of the number phases. The experimental group experiences two phases, whereas the control group group one phase. Other than that, the design of the two groups are identical.

Learning part Trials are presented in blocks to control for the amount of learning received for each word. Participants are exposed to a minimum of 6 blocks. The learning phase ends when

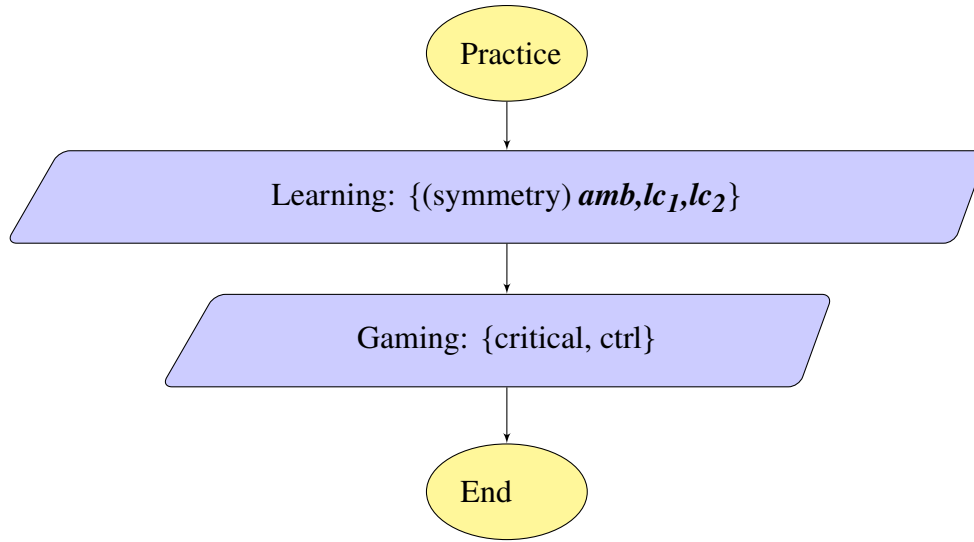


Figure 8.8: Control group: sequence of phases

participants respond correctly for all trials in a block. If they answered more than 6 blocks (i.e., 96 trials in phase 1 and 120 trials in phase 2) without reaching the learning criteria, the experiment continued normally but their responses will be discarded. No participants' responses are discarded as none of them used up the 6 blocks attempts in the learning phases.

Figure 8.9 illustrates a loop in a learning phase, in which participants learn an ambiguous phrase that has two interpretations (i.e., $Amb_{\{lc,c\}}$), corresponding to two nonce phrases: a less costly alternative phrase (i.e., *l-costly alt*) and a costly alternative phrase (i.e., *costly alt*). Additionally, participants learn a filler phrase. This is to ensure that whatever effects get observed are not merely due to participants learning a particular type of phrases. The loop repeats 6 times until participants correctly respond all the trials in a loop. When the loop ends, participants can proceed to the gaming part.

Testing part As mentioned in Table 8.5, there are two kinds of critical trials, one of which has the costly object circled (i.e., *c intended*), the other has the less-costly object circled (i.e., *lc intended*). As to the control trials, there is only one geometric shape, which is always the intended. The naming convention here (*c* and *lc*) is not very meaningful for the control group, as they never actually learn

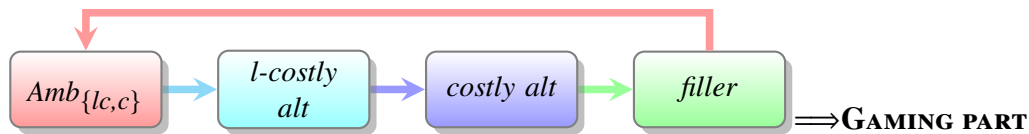


Figure 8.9: Zoom in: A flowchart of a learning phase in the artificial language learning experiment (the sequence is presented for illustration purpose; trials are randomly interspersed in the actual experiment)

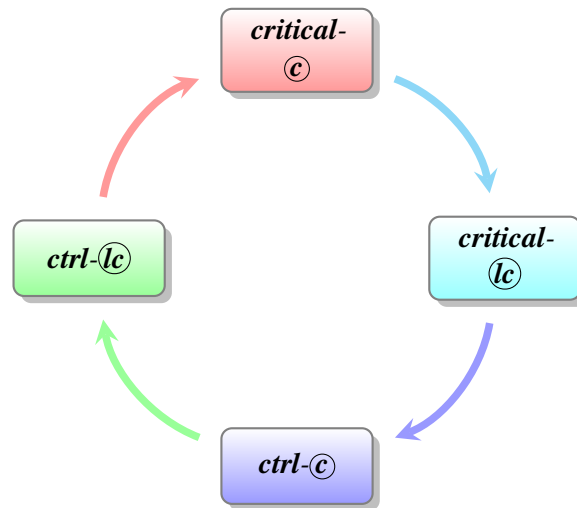


Figure 8.10: Zoom in: A flowchart of a gaming phase in the artificial language learning experiment (the sequence is presented for illustration purpose; trials are randomly interspersed in the actual experiment)

a costly-object-type phrase. It serves the purpose of data coding.

Figure 8.10 illustrates a loop in the gaming part, in which participants are tested through four trials. The loop repeats two times. This is to make sure that enough responses are collected and the responses are *consistent*.

8.2.3 Results

For experimental group, Figure 8.11 demonstrates experimental group's overall performance. The y-axis is the amount of ambiguous-type response out of all valid responses, and the x-axis shows that the data is grouped by phases. It shows that the two phase have similar trend and similar shape in general, but they differ in that the critical cases versus the control cases responses are more different

in phase2 than in phase1. In particular, overall fewer people go for ambiguous response for critical cases in phase 2 than in phase 1; conversely, more people go for ambiguous phrase for control cases in phase 2 than in phase 1, resulting in a greater contrast for phase 2 than for phase 1.

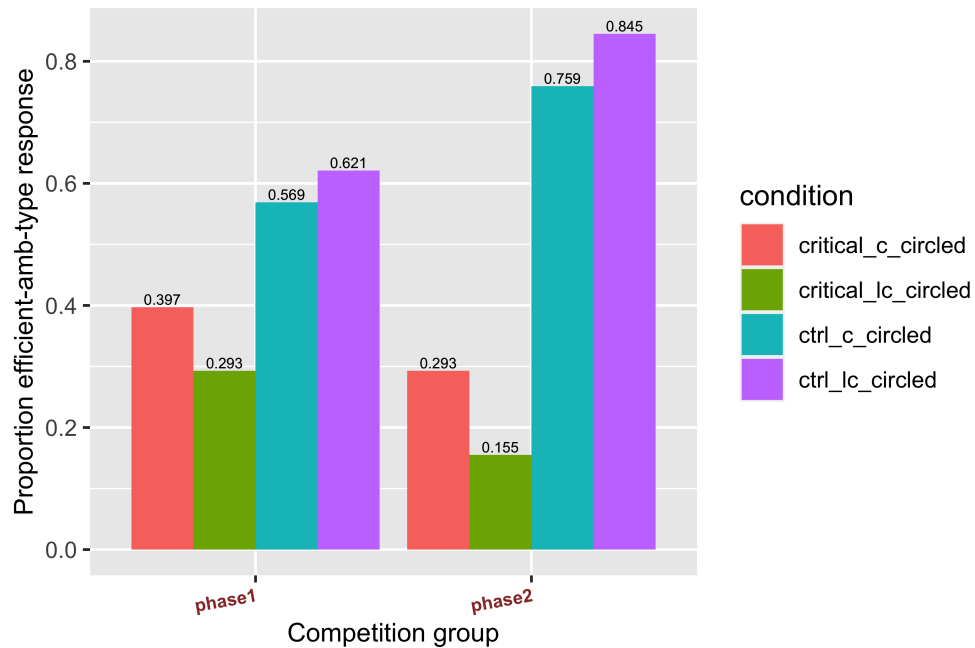


Figure 8.11: Overall performance - Experimental group

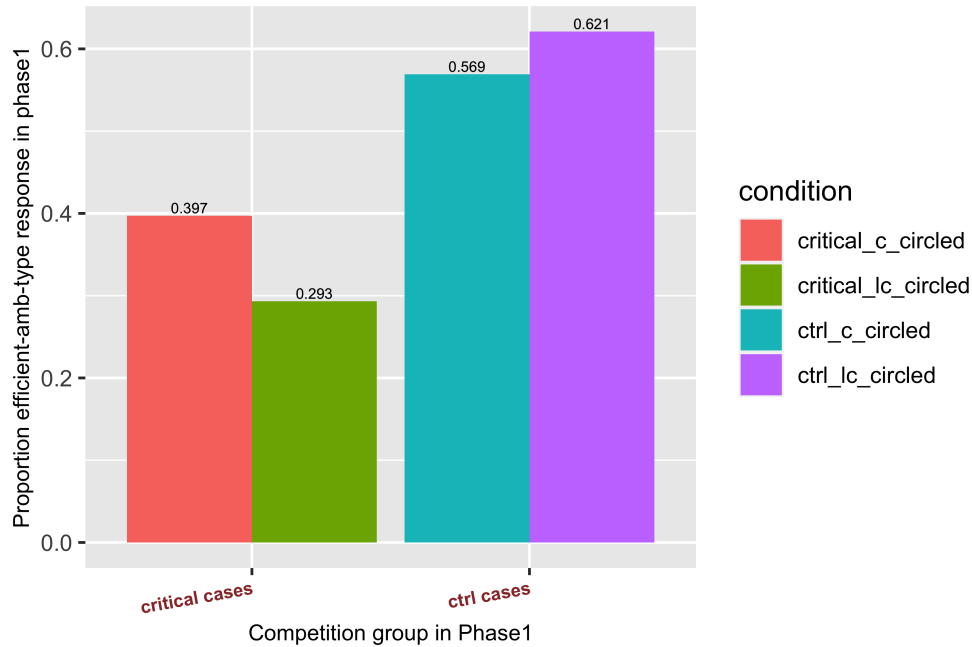


Figure 8.12: Within-group comparison: ambiguous responses phase1

Figure 8.12 illustrates a within experimental group comparison, in terms of the proportion that participants choose to type an ambiguous phrase in phase 1, with respect to all the valid responses (i.e., inaccurate responses and empty responses excluded). By eyeballing the data, we can see that the trend is in the right direction, but the contrast is not as pronounced as predicted. Ideally, an efficient speaker would go for *kad* all the time in critical cases, making the ambiguous-type-response proportion in a costly-object-intended scenario significantly different than that when a less costly is intended. Moreover, hypothetically there should be no dramatic difference in terms of the proportion of ambiguous-type-response in control cases.

What I actually find is that in critical cases, participants are more inclined to go for the ambiguous-type-response when a costly object is intended (39.7%), compared to their decision when a less-costly object is intended (29.3%), although the difference is not extremely pronounced. This appears to support the proposed competition principle in the following way. An ambiguous *context* is given, in which there are two salient messages and the costly one is intended. Further, participants

have the *prior knowledge* that to convey the intended message, there are two alternatives available: a costly phrase, and a simple one that can communicate both of the two salient messages. If a speaker (i.e., the participants) is caring about efficiency and considers the cost difference between the two competing alternatives, they should reason about *ambiguity* and thus go for the ambiguous yet simple alternative. By contrast, when the same ambiguous context is given but the *less-costly* message is intended, they should reason about the unambiguous alternative, because now the cost difference of the two competing alternatives is not *significant* enough for efficient speakers to compromise on specificity. This is borne out, as visualized in Figure 8.12 the red and the green bars.

Crucially, in control cases where ambiguity is absent and the only salient message is also the intended, no such contrast should be observed. Also, overall, participants should go more for the efficient-ambiguous-type responses in the control cases than they do in the critical cases. In the critical cases where the efficient-ambiguous-type-responses' two messages are both available, participants cannot help thinking that if they chose the ambiguous response, there is a chance that the Martian listeners would think that I intended the un-circled message. It might lead to miscommunication. As a consequence of communication concern, there are quite a few times where the costly yet specific responses are picked. However, such concern is less likely to get triggered as the context is no longer ambiguous. The results are as predicted, as demonstrated by Figure 8.12 the blue and the purple bars.

Figure 8.13 shows the within experimental group comparison in terms of the critical cases. The y-axis concerns the proportion of non-ambiguous responses in phase 1. The data is grouped by the object that's circled, namely a group in which a costly object is intended versus a group in which a less-costly object is intended. The red bar illustrates the proportion of costly-object-type response when a costly object gets circled, whereas the blue bar demonstrates the proportion of less-costly-object-type response given that a less-costly object is intended. The results are on the right track but they are not very clear-cut.

Figure 8.13 is labeled as “good-participant-type response”, since it stands in contrast with “*ideal-participant-type response*”. In an ideal case where the hypothesis is 100% accurate and where

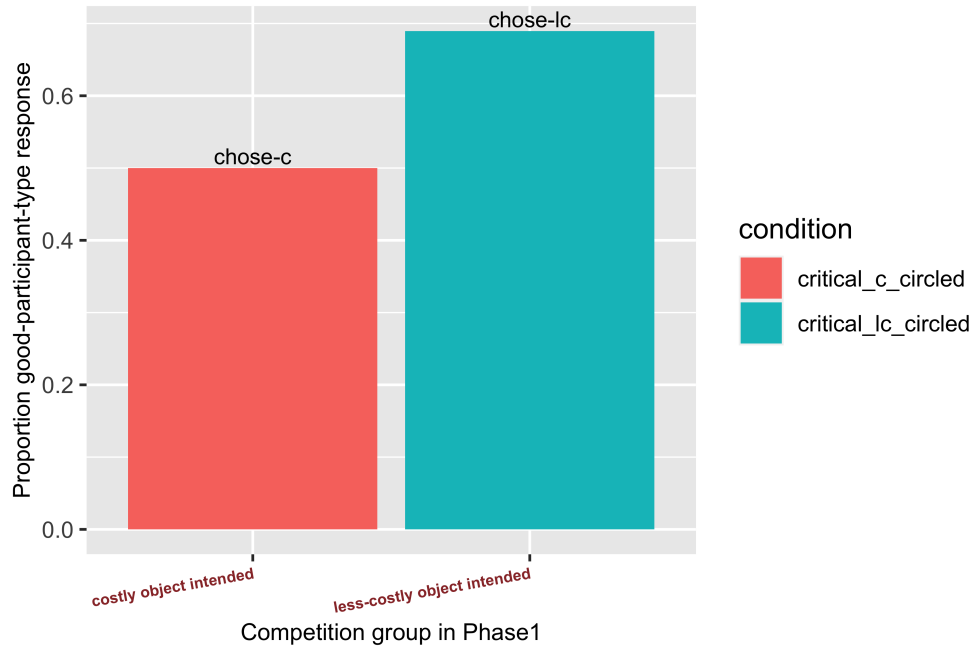


Figure 8.13: Within-group comparison: involving non-ambiguous responses phase1

participants are behaving exactly as predicted, there should be 0% costly yet specific response when costly object is intended, and conversely, 100% of the time participants should go for the specific (and less-costly) response when the less-costly object is the intended. Those are *ideal* speakers' performance. They are “optimal agent” in a sense that they always optimize efficiency, and take appropriate strategies to maximize information communication whenever necessary. In practice, however, various unexpected biases, memory limitations, fatigue, errors, etc. can affect performance, yielding more middling results. Consequently, what I observe from the experiment is that most of the time participants are *good enough* but not *ideal* speakers. Around 50% of the time participants responded with a specific phrase as opposed to an ambiguous one, when a costly object is intended. About 70% of the time participants chose the specific less-costly response, as opposed to the ambiguous phrase, when the less-costly object is circled. Overall, results in Figure 8.13 are in the right direction, but not as pronounced as would be predicted in an ideal world.

Figure 8.14 illustrates a between group comparison. The data is classified by groups: competition

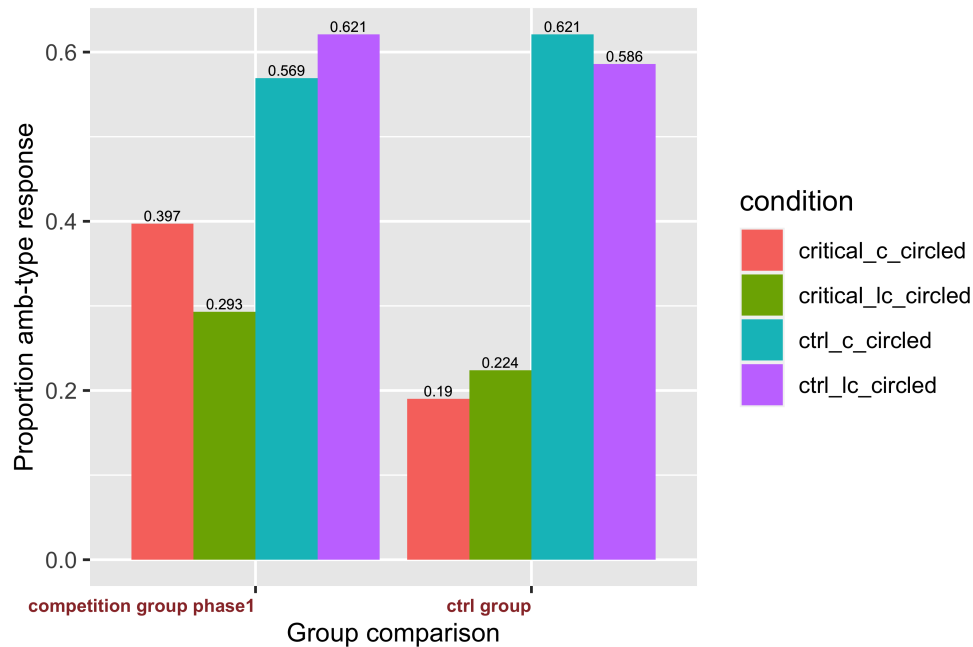


Figure 8.14: Between-group comparison

group phase 1 versus control group. The y-axis is the proportion of ambiguous type phrases. Overall, the difference between the critical cases' response and the control cases' response is greater in the control group than in the competition group phase1. Specifically, it's due to the fact that the proportion of ambiguous response is smaller in the control group critical cases than in the competition group critical cases. These are desirable results for the following reasons. For the control group critical cases, participants sensed that there is no cost difference between the ambiguous-simple-response and the specific-less-costly-response. Put otherwise, there is no competition of efficiency, thus there is no strong motivation for them to go for the ambiguous response. When it comes to control group control cases, participants' response mirror the competition group counterpart, since the control cases are consistently unambiguous.

Error bars represent standard error of the mean of the proportion efficient-ambiguous-type response. Figure 8.15 illustrates that in a critical case where there are two geometric shapes (i.e., ambiguous context), the experimental group phase1 shows difference depending on the intended

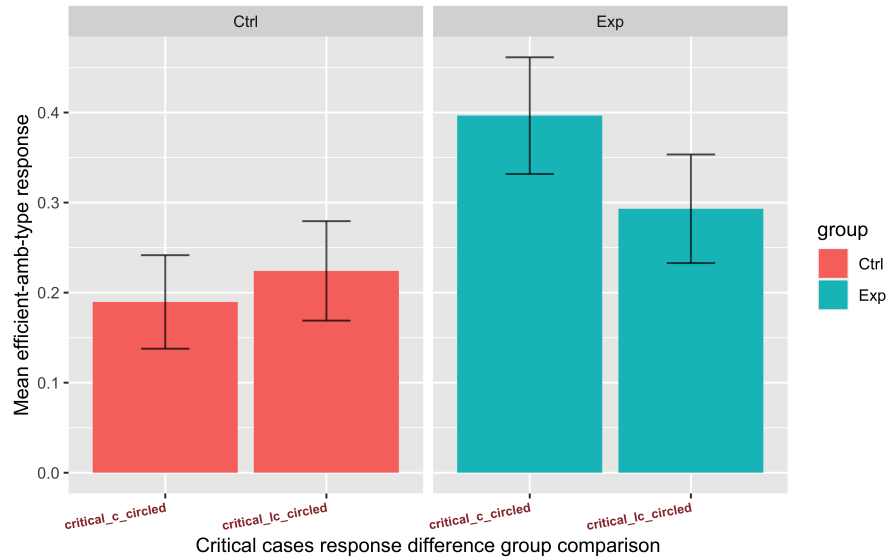


Figure 8.15: Error bar represents standard error of the mean: critical cases condition-based (costly vs. less-costly intended) *cost* effect difference between group

message: costly versus less costly, whereas such difference is smaller for control group. This is as predicted.

Figure 8.16 demonstrates that within experimental group, phase1 shows stronger tendency of going for ambiguous responses in critical cases, compared to phase2. This is in line with the prediction. Interestingly, this result is not predicted. I will discuss this result in the Discussion section. The difference of ambiguous response to a costly intended message and that to a less-costly one is smaller in phase1, compared to the difference in phase2.

Table 8.7 gives the basic information about the data that’s used in data analysis. There are 232 valid response, among which none of them are inaccurate, incomplete, or empty strings. Divided by 8 cases (2 repeats of 4 different cases { critical.c.circled, critical.lc.circled, ctrl.c.circled, ctrl.lc.circled}), it gives exactly 29 participants who are not consistently specific in the control group, and 29 in the experiment group (i.e., $N = 29$). That also speaks to the number of trials in a gaming block: 8 for Ctrl group, 8 for Exp.phase1 and 8 for Exp.phase2. “29*” means that the same 29 participants take part in both Exp.phase1 and Exp.phase2. The “mean” in Table 8.7 refers to the the

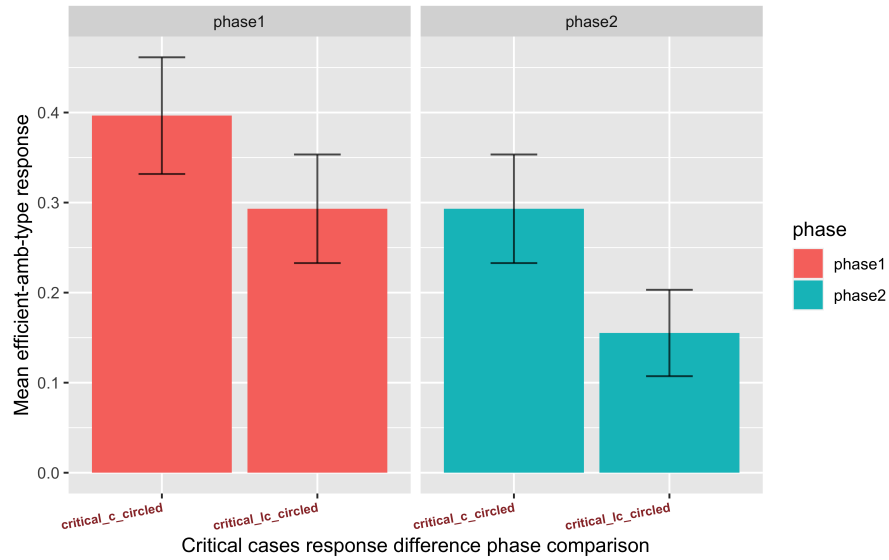


Figure 8.16: Error bar represents standard error of the mean: critical cases condition-based (costly vs. less-costly intended) *cost* effect difference between phase

mean of the efficient-ambiguous-type responses, and the “Std.Error” refers to the standard error of the mean. “condition” refers to different critical contexts in which the intended message can be an costly object or a less-costly one.

group	phase	valid.response	N	mean	Std.Error
Ctrl	phase1	232	29	0.207	0.0378
Exp	phase1	232	29*	0.345	0.0443
Exp	phase2	232	29*	0.224	0.0389
group	phase	condition		mean	Std.Error
Exp	phase1	critical.c.circled		0.397	0.0648
Exp	phase1	critical.lc.circled		0.293	0.0603
Exp	phase2	critical.c.circled		0.293	0.0603
Exp	phase2	critical.lc.circled		0.155	0.0480
Ctrl	phase1	critical.c.circled		0.190	0.0519
Ctrl	phase1	critical.lc.circled		0.224	0.0552

Table 8.7: Descriptives

Table 8.8 and Table 8.9 provide statistics analyses for two of the most crucial mixed effects:

between-group comparison of the proportion of ambiguous response in critical cases, and within-group cross-phase comparison of the proportion of ambiguous response in critical cases. Table 8.8 shows the output of the (One-way) ANOVA analysis and whether there is a statistically significant difference between the group means. We can see that the significance value is 0.0187 (i.e., $p = .019$), which is below 0.05. Therefore, there is a statistically significant difference in the mean proportion of ambiguous-type-response to critical cases between the experimental group (phase1) and the control group. As to the effect on different learning phases, we can see that it does significantly influence on participants' reasoning on ambiguous response ($p = .042$). This is a more modest effect, though. Specifically, they went more for the ambiguous phrase in critical trials in phase 1 than in critical trials in phase 2⁵.

Table 8.8 also summarizes a higher-order comparison of the mean of critical cases ambiguous responses when a costly message is intended versus that when a less costly message is intended. No statistically significant difference was found within experimental group phase1 ($p = 0.245$), within experimental group phase2 ($p = 0.076$), or within control group ($p = 0.65$). However, the p values show that the difference of the ambiguous response difference within experimental group ($p = 0.245$) is greater than that in control group ($p = 0.65$), as predicted. Surprisingly, for experimental group phase2, such difference ($p = 0.076$) turns out to be even greater than that for experimental group phase 1.

	Df	F value	Pr(>F)
critical.choseamb (Exp-phase1 vs. Ctrl)	1	5.61	0.0187*
critical.choseamb (Exp-phase1 vs. Exp-phase2)	1	4.19	0.0418*
critical.choseamb.Exp-phase1 (c.intended vs. lc.intended)	1	1.366	0.245
critical.choseamb.Exp-phase2 (c.intended vs. lc.intended)	1	3.206	0.076
critical.choseamb.Ctrl (c.intended vs. lc.intended)	1	0.207	0.65

Table 8.8: Comparison (cross-group; cross-phase)

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

⁵I did collapse both types of critical trials into one (triangle intended vs. square intended).

We can further find this out in the Multiple Comparisons table which contains the results of the Tukey post hoc test (Table 8.9). There was a statistically significant difference between groups as determined by one-way ANOVA ($F(1) = 5.61, p = .019$). A Tukey post hoc test revealed that the proportion of ambiguous response was statistically significantly lower without the presence of the cost competition factor ($p = .019$). There was statistically significant difference between phase2 and phase1 within the experimental group ($p = .042$). The proportion of ambiguous response in critical cases decreased significantly in phase 2, where the cost competition factor gets removed.

	diff	[lwr, upr]	<i>p</i> adj
critical.choseamb (Exp.phase1 - Ctrl)	0.138	[0.023 0.253]	0.019
critical.choseamb (Exp.phase2 - Exp.phase1)	-0.121	[-0.237 -0.005]	0.042

Table 8.9: Tukey multiple comparisons of means 95% family-wise confidence level

8.2.4 Other findings

Figure 8.17 concerns control group’s performance. It zooms in and demonstrates the proportion of ambiguous type responses in critical cases and in control cases. It’s a subset of Figure 8.14.

Figure 8.18 demonstrates control group’s non-ambiguous specific type responses (i.e., combined *c* and *lc*). It illustrates a comparison of critical cases and control cases. Around 80% of the time control group participants chose specific response in critical cases, whereas only half of it (around 40%) in control cases. This makes sense in that compared to control cases, for critical cases where using an ambiguous phrase is easier to have the concern of miscommunication, using a specific and not-that-costly alternative seems to be a safer choice. This also speaks to the similar concern raised in competition group.

Figure 8.19 illustrates how *cost* influences pragmatic reasoning in different groups. Figure 8.20 demonstrates the role of *cost* in different phases. For both Figure 8.19 and Figure 8.20, “critical cases” refer to the combination of two critical cases: one where the costly object is circled, and one where the less costly object is circled.

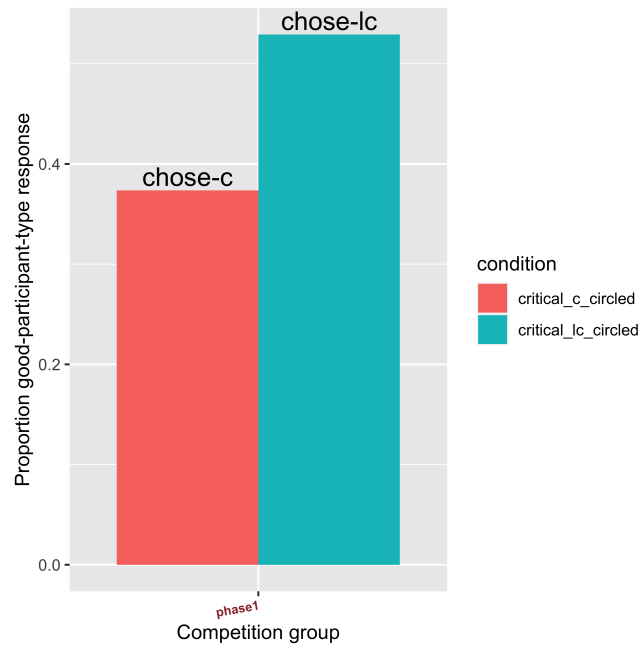


Figure 8.17: Within-group: ambiguous response (ctrl)

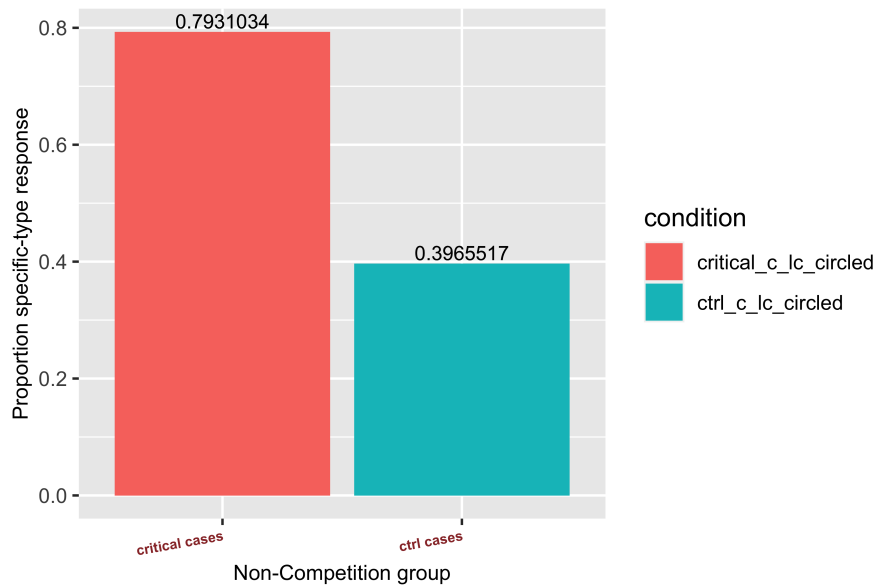


Figure 8.18: Within-group: non-ambiguous response (ctrl)

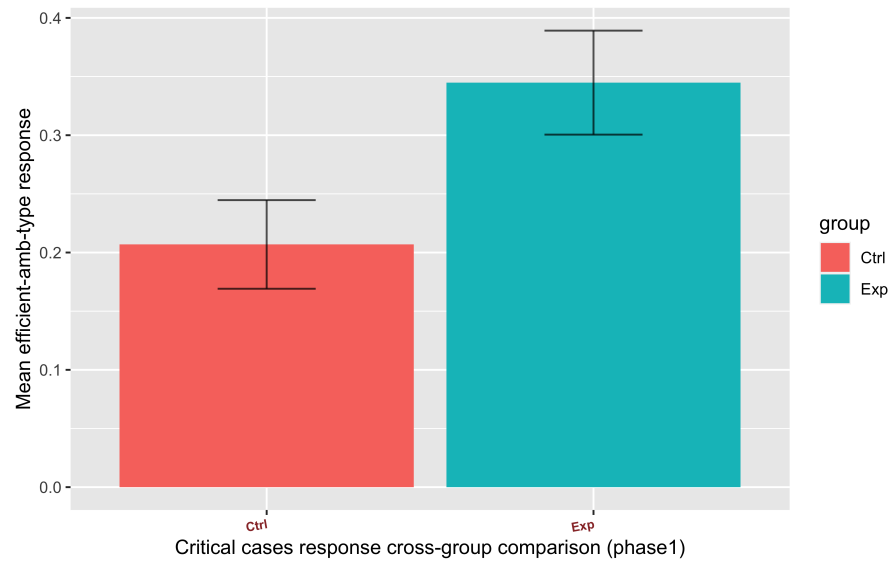


Figure 8.19: Error bar represents standard error of the *mean*: *cost* effect on group

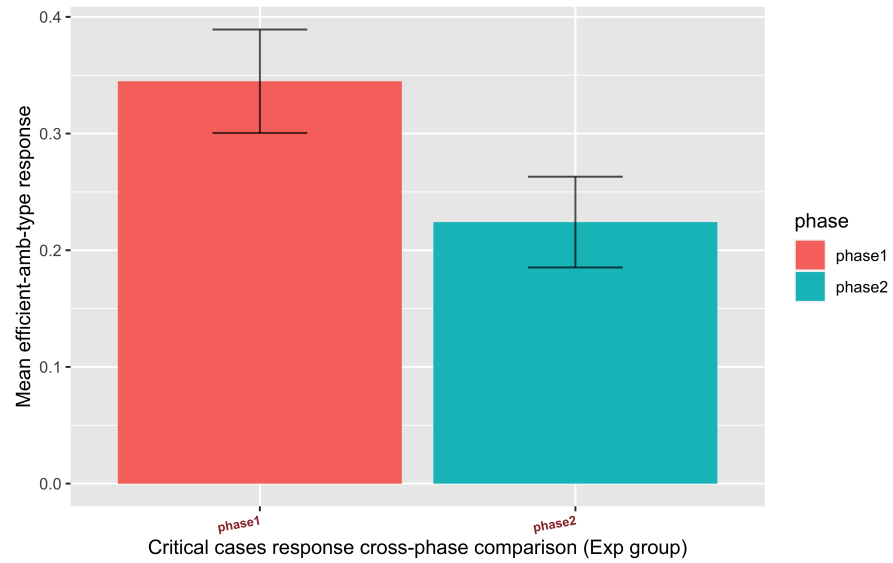


Figure 8.20: Error bar represents standard error of the mean: *cost* effect on phase

8.2.5 Discussion

As an extension of Buccola et al. (2018), overall the results of the current experiment are as predicted, namely it does appear to be the case that participants adhere to a competition principle, as Buccola et al. (2018) found, but also that cost – a novel factor in this experiment – plays a role in pragmatic reasoning. However, the results are not as *pronounced* as might be expected in an ideal world.

8.2.5.1 Limitations and explanations

There are some limitations. First, the overall performance of Experimental group as shown in Figure 8.11 indicates that participants seem to be reluctant to *significantly* update their choice in phase2, even when the competition is no longer present. One likely reason might be that there is an asymmetry of the amount of different available alternatives in participants' prior knowledge (i.e., phrases they learn through the learning phase). Although the amount of exposure is strictly controlled and well-balanced, by the time they get to the gaming part in phase2, they *know* that there are two alternatives to refer to the costly object (e.g., *kad shegmifanzub*, *kad mof*), yet only one alternative to refer to the less-costly object (e.g., *kad vun*). This may give them an (unwanted) impression that the costly object is somehow more important and salient for aliens. Thus they would keep using the same (ambiguous) phrase they learned in phase1 to refer to the costly object (e.g., *kad*).

I acknowledge that the artificial language learning experiment in this dissertation does not manage to tease apart 'underspecification' and 'ambiguity'. It's likely that participants treat '*kad*' as underspecification, because the experimental items are all subsets of the so-called "geometric shapes", whereas the distractor options are all organic shapes. Even though proposal focuses on addressing a puzzle about ambiguity and competition, the artificial language learning experiment does not speak directly to ambiguity.

Second, typing a response can be a feasible way used in a behavioral experiment to measure the extent to which *cost* plays a role in reasoning, nevertheless, this experiment did not restrict keys that are allowed during the experiment. As a consequence, some participants gave partial

responses that can be hard to interpret or classify. For example, there a participant who consistently typed “k”. It can be a mistake. But it can also be a sign that the participant is being even more efficient. Since *sot* is the filler phrase for that stimuli set, the participant might reason about the efficient-ambiguous-type response *kad* all the time, so they end up typing “k” as it’s distinct from “s” (for *sot*).

Third, a critical limitation of the current experiment is that the effect of question under discussion (henceforth QUD) has been raised in the language data (see Chapter 3), but has never been really tested in the experiment. The closest thing in the set-up of the current experiment is the hand drawn circle for the geometric shapes in the stimuli. It *can* be viewed as an implicit QUD. For future studies, I plan to revise the current version by tackling all kinds of technical issues and incorporating QUD into the design. Additionally, an explicit-cost-label version, and listener versions with and without cost-label are planned to be implemented to better our understanding of the role of *cost* in Gricean pragmatics.

Fourth, I used word length as a proxy for cost, but this is not necessarily expected on a Katzirian (structural) view of complexity. So how does this notion of cost connect back to competition between utterances in natural/non-artificial language?

And relatedly, the connection back to “gao” still needs to be made: “gao” is like “kad”, “hen gao” is like “kad vun”, and “bijao gao” is like “kad shegmifanzub”, but it’s not clear in what sense the extra cost of “bijao gao” is related to the extra cost (length) of “kad shegmifanzub”. This needs to be worked out in future research.

8.2.5.2 Implications

This experiment has interesting implications for the fast-mapping literature. Trueswell et al. (2013) suggests a strategy called “Propose but verify”: single conjectures from multiple exposures. This learning strategy is also called *fast-mapping*, reflecting the need for the learner to maintain a single hypothesized meaning that is then tested at the next learning instance for that word, rather than the models that track multiple meaning hypotheses for each word. They conducted labeling experiment

and eye-tracking measurement.

Similarly, Gerken (2006) attempts to tease apart similarity from complexity. They indicate consistency with the familiarization task and ask the question: is it “weird” to update your choice, or is it better to stick to your choice.

	di	je	li	we
le	leledi	leleje	leleli	lelewe
wi	wiwidi	wiwije	wiwili	wiwiwe
ji	jijidi	jijije	jijili	jijiwe
de	dededi	dedeje	dedeli	dedewe

Table 8.10: Gerken 2005:2 (columns: B; rows: A)

Table 8.10 AAB familiarization stimuli used by Marcus et al. (1999). The First column and diagonal were used as familiarization stimuli in the current experiments. Now consider two different subsets of four stimuli in Table 8.10: (1) those on the diagonal: AAB (generalizing & memorizing the entire set of 16); (2) those on the first column: AAB + end in *di* (subset of the entire set of 16). The question addressed in the two experiments include the following: (i) what information do infants discern in the stimuli in the first column of Table? (ii) do they make only the AAB generalization, only the “ends in *di*” generalization, both, or neither? In terms of materials, they use 3-syllable nonsense items, the conditions are illustrated below:

1. the column condition: AAB (*leledi*, *wiwidi*, *jijidi*, *dededi* or ABA (*ledile*, *widiwi*, *jidiji*, *dedide*)
2. the diagonal condition: AAB (*leledi*, *wiwije*, *jijili*, *dedewe*) or ABA (*ledile*, *wijewi*, *jiliji*, *dewede*)

The procedure is to have a familiarization phase (4 3-syllable strings) followed by 4 test trials (2 AAB and 2 ABA)⁶. The between subjects variable is familiarization condition (column versus diagonal); and the within subjects variable concerns test condition (consistent versus inconsistent

⁶Exp. 2 test phase: *di* replaced the B element in both AAB and ABA test strings

with training). Results show that infants showed a preference for stimuli consistent with their familiarization condition. Infants in the column condition made only the generalization involving the position of the syllable *di*. Infants in the diagonal condition, who were familiarized with a subset of the stimuli in which the only common feature was an abstract AAB or ABA pattern, were able to make the intended generalization (i.e., showed significant discrimination).

The current experiment can rely on their findings in that they have the potential to explain why people are inclined to propose a hypothesis for each utterance and stick to it while verifying it, instead of proposing multiple hypotheses and make revisions. More specifically, their findings can help explain that in the current experiment where there are two phases, participants are likely to stick to the phrase they have learned in the initial phase, if they run into the same scenario or message in phase. Strategically speaking, this can be viewed as an “efficient” behavior.

8.3 Chapter summary

This chapter provides a way to examine alternatives and to measure the extent to which cost influences pragmatic reasoning. I adopted the artificial language learning experiment design proposed in Buccola et al. (2018), and turned it into a speaker-perspective experiment, which also incorporated a cost asymmetry. Results are as predicted in general, but are not as pronounced as expected. Limitations are discussed.

To sum up, this chapter provides independent evidence for the argument about the role of cost/complexity in competition, through examining how natural language users respond to nonce phrase given ambiguous contexts. The fact that the competition principle is observed outside of natural language learning settings suggests that the notion of “alternatives” and “competition” is not only linguistic but can also get extended to human cognition.⁷

⁷Although aiming at involving “*conceptual*” alternatives (see Chapter 5), this experiment has almost nothing to do with conceptual alternatives or the language of thought. The experiment may actually be entirely linguistic (not conceptual). The reason for using artificial language is to be able to manipulate cost experimentally.

CHAPTER 9

CONCLUSION AND FUTURE WORK

In this chapter, the questions and puzzles mentioned in Chapter 1 are revisited. A summary of the contributions and the limitations of this dissertation is given. Finally, potential directions for future work are laid out.

9.1 Circling back to the question about competition

The core idea of this dissertation is proposed with the notion of competition serving as a backdrop. The investigation started with a literature review on the classic puzzle in competition, i.e.: the symmetry problem in the (scalar) implicature domain (Chapter 4). Apparently, little attention has been given to the non-scalar implicature domain. An ambiguity phenomenon observed in Chinese degree expressions was studied, with respect to Manner implicatures (i.e., non-scalar implicatures). The classic symmetry-puzzle was discovered and replicated in the non-scalar implicature domain (Chapter 5). The specialty of such a puzzle in non-scalar implicatures was stressed via the role of *cost* (Chapter 6). As proof of concept, Chapter 7 and Chapter 8 illustrated the extent to which *cost* influenced pragmatic reasoning.

The proposed competition principle has two major implications. First, the costly-referent-type interpretation sustains, given that an ambiguous utterance that has two interpretations, corresponding to a costly alternative and a less costly alternative. The reasoning behind this hypothesis is extensively simulated in the RSA Chapter (Chapter 7). Second, given a costly referent, and a costly utterance and an efficient yet ambiguous utterance, the latter is favored. Chapter 8 provides an independent evidence for this hypothesis, under an artificial language learning experiment setting.

9.2 Circle back to the question about degree

Against the theoretical background about symmetry and competition, the empirical puzzles observed in Chapter 2 and Chapter 3 become meaningful. This dissertation proposes a novel

approach to examine the (subtle) interpretation of degree expressions. I highlight the importance of linguistic context in a truth-value judgment survey.

From the degree semantics perspective, this dissertation is consisted of two major components: (i) the disambiguation model, which is proposed based on theoretical and empirical claims of adjectives in Chinese; (ii) the implementation of this disambiguation model in comparison-related reasoning settings. For future work about degree, I plan to explore the typological implication of the disambiguation model. All of these ingredients concern one phenomenon: competition in natural language meaning. They together tell one story that ambiguity is efficient: not only does the intended meaning get conveyed, conversational participants also get across the enriched meaning that the alternatives are not intended. Starting from there, I show that the classic symmetry problem is discovered in a new domain, and a novel theory is thus needed.

It's worth mentioning that the hypotheses and the analyses illustrated in this dissertation are confined to *gradable* adjectival constructions. For non-gradable adjectival constructions, there are cases in which they can be tensed and behave like verbal expressions. It's unclear whether the proposal still holds for non-gradable adjectival constructions such as *open* or *closed*. I leave cross-categorical inquiries for future exploration.

9.3 Circle back to the question about language universals

I plan to investigate whether degree expressions across languages showcase similar competition generalizations as those found in Chinese. This section takes an initial look at languages beyond Mandarin Chinese and seek the source of semantic universals.

9.3.1 Expand the typology table

The existence of morphemes like *hen* motivates us to expand the typology table discussed in chapter 2 (repeated below):

(1) Grano and Davis' observation (2018)

	<i>Positive form</i>	<i>Comparative form</i>	<i>Examples</i>
Pattern A	Adj	Adj	Japanese, ...
Pattern B	Adj	DERIV(Adj)	English,Irish,French,...
Pattern C	DERIV(Adj)	Adj	Impossible?
Pattern D	DERIV ₁ (Adj)	DERIV ₂ (Adj)	Impossible?

Pattern C and Pattern D have been claimed to be impossible, and Chinese has been argued to be classified as a Pattern A language. If these claims hold, there should be a positive morpheme available in Pattern A languages such as Japanese. If not, a similar pragmatic account should explain Japanese data, as have been proposed for the Chinese data in this dissertation.

9.3.2 Universals and tendencies

A mini fieldwork is necessary to fulfill the goal that, if the proposed disambiguation model still holds for languages that are unrelated to Mandarin Chinese, then it's solid to assert that the model is applicable cross linguistically. Moreover, this is also to provide evidence for the following generalizations, inspired by (Stassen, 1985; Grano, 2012; Coppock et al., 2020):

(2) Hypothesis - first attempt:

- **UNIVERSAL:** Bare (subjective) adjectives are underspecified with COMP and POS.
- **TENDENCY:** The way that they get expressed depends on what morphosyntactic machinery is available in that language:
 1. IF the language lacks such a COMP morpheme but there is a POS morpheme available, then the adjectives in plain form are more likely interpreted as COMP
 2. IF there is a dedicated COMP morpheme available in the language but no dedicated POS morpheme (or no simple way to express POS), then the adjectives in plain form are more likely interpreted as POS

3. IF a language that does not have a COMP morpheme and also does not have an easy way to express POS, then a bare adjective should be totally underspecified. There is no competition.

An implication of the above generalization is that TENDENCY-3 is cognitively costly, compared to the other two situations. Hypothetically, there should be a separate strategy for this type of language to circumvent the ambiguity.

9.3.3 Candidate languages

Here are some candidate languages which do not have an *-er* morpheme in comparatives. If the claimed universal and tendency can be justified in the following languages, one may find it convincing to make the generalization that the proposed disambiguation model is not confined to Mandarin Chinese.

- (3) *Mo tobi*
I big-PRES
I am big YORUBA

- (4) a. *Mo lo*
I go-PRES.IND
I go

b. *Onisowo ni mi*
merchant COP I
I am a merchant YORUBA

- (5) *Mo tobi ju u*
I big-PRES exceed him
I am bigger than him YORUBA

For Yoruba, it is described in Stassen (1985) that syntactically, adjectives such as ‘big’ (3) behave more like verbs (4a) than like nouns (4b), in a sense that they can tensed. This description is reminiscent of adjectives in Mandarin Chinese, which are also verby. Crucially, to convey a thought ‘I am bigger than him’, adjective *tobi* ‘big’ is used with plain form.

- (6) *I -pandu a -pandu-kanna tipi -ga undi*
 this -fruit that -fruit-PARTICLE sweet -one is
 This fruit is sweeter than that fruit TELUGU

- (7) *Ramarav podugu-vadu ka -du*
 R. tall-one not -is
 Ramarav is not tall TELUGU

- (8) *I-pandu tipi -di ka-du*
 this-fruit sweet -one not-is
 This fruit is not sweet TELUGU

For Telugu, Stassen (1985) shows that adjectives used in comparative constructions such as (6), and in (negated) positive constructions such as (7, 8) are both in plain form: *tipi* ‘sweet’, *podugu* ‘tall’. This is, again, very reminiscent of adjectives in Mandarin Chinese.

- (9) *Lam nan-pa -yin-las rta agrul-ma-thub*
 road bad-one -be-PARTICLE horse walk-not-able
 The road was bad and the horses could not walk TIBETAN

- (10) *Rta -nas khyi chun-ba yin*
 horse -from dog small-one is
 A dog is smaller than a horse TIBETAN

For Tibetan, bare form adjectives are used in (negated) positive constructions such as (9) as well as in comparative constructions such as (10), according to the description in Stassen (1985). Chinese behaves in a very similar way.

- (11) *Muri khipi-na kenne kagesso*
 water deep-CONV.CONCESS we go over-FUT
 Although the water is deep, we will cross it KOREAN

- (12) *Na-eso to kheda*
 I-from he big-PRES
 He is bigger than me KOREAN

For Korean, plain form adjectives are found in both positive constructions such as (11) and in comparative constructions like (12), according to Stassen’s description (1985). Adjectives in Chinese, again, behave in very much the same way.

9.4 Chapter summary

Although it's finally time for me to conclude the dissertation and write the Chapter summary of the conclusion chapter, I don't consider what I write here to be the final word, but rather a beginning of the investigation of a research program.

This dissertation has illustrated attempts to explain a disambiguation-based competition phenomenon in degree expressions. I propose that for non-scalar implicatures, the classic symmetry problem still exists. This dissertation argues that cost influences the pragmatic reasoning about alternatives, in a sense that gradient costs break the symmetry of the relevant alternatives, resulting in a particular interpretation surface. I provide a way to approach the puzzles, using both the classic structural analysis (Katzir, 2007) and the more recent conceptual alternative analysis (Buccola et al., 2021). As an extension of Buccola et al (2019), this dissertation proposes a speaker-perspective artificial language learning experiment as a proof of concept justifying the effect of cost. Additionally, probabilistic pragmatic framework RSA (Frank & Goodman, 2012a) is applied to simulate and visualize the proposed competition principle.

Although the domain of my investigation is far from complete, I hope I have drawn a basic picture that can capture our general intuition of how competition in natural language works. I would like to conclude the dissertation with a concise summary of the contributions: methodological/empirically – truth value judgment survey showing that “Anna gao” is genuinely ambiguous; theoretically – uncovering the symmetry problem in NSI/manner implicature, and using the RSA to incorporate cost; experimentally – a novel extension of artificial learning experiments to test the role of cost in competition.

BIBLIOGRAPHY

BIBLIOGRAPHY

- Adger, D, A Martin, J Culbertson, K Abels & T Ratitamkul. 2019. Cross-linguistic evidence for cognitive universals in the noun phrase. *Linguistics Vanguard* .
- Anand, P. & A. Nevins. 2004. Shifty operators in changing contexts. In R.B. Young (ed.), *R.b. young (ed.), proceedings of salt 14*, 20–37.
- Asherov, Daniel, Danny Fox & Roni Katzir. 2021. On the irrelevance of contextually given states for the computation of scalar implicatures. In *Proceedings of the linguistic society of america 2021 annual meeting*, upcoming.
- Beck, Sigrid, Sveta Krasikova, Daniel Fleischer, Remus Gergel, Stefan Hofstetter, Christiane Savelsberg, John Vanderelst & Elisabeth Villalta. 2009. Crosslinguistic variation in comparison constructions. *Linguistic Variation Yearbook* 9(1). 1–66.
- Beltrama, Andrea. 2016. *Bridging the gap: Intensifiers between semantic and social meaning*: University of Chicago dissertation.
- Beltrama, Andrea. 2018. Subjective assertions are weak: Exploring the illocutionary profile of perspective-dependent predicates. In *Proceedings of sinn und bedeutung*, vol. 22 1, 160–173.
- Bergen, Leon, Roger Levy & Noah D. Goodman. 2016. Pragmatic reasoning through semantic inference. *Semantics and Pragmatics* 9. doi:10.3765/sp.9.20.
- Bhatt, Rajesh & Roumyana Izvorski. 1998. Genericity, implicit arguments and control. In *Proceedings of the 7th student conference in linguistics. paper presented at scil*, vol. 7, .
- Buccola, Brian, Isabelle Dautriche & Emmanuel Chemla. 2018. Competition and symmetry in an artificial word learning task. *Frontiers in Psychology* 9 (2176) .
- Buccola, Brian, Manuel Križ & Emmanuel Chemla. 2021. Conceptual alternatives. *Linguistics and Philosophy* <https://doi.org/10.1007/s10988-021-09327-w>.
- Bylinina, Lisa. 2014. *The grammar of standards: Judge-dependence, purpose-relativity, and comparison classes in degree constructions*: Universiteit Utrecht dissertation.
- Bylinina, Lisa. 2017. Judge-dependence in degree constructions. *Journal of Semantics* 34. 291—331.
- Cable, S. 2005. *Binding local person pronouns without semantically empty features*. MIT.
- Capell, A. 1957. *A new fijian dictionary*. Wilson Guthrie and Company Limited.
- Cappelen, H & J. Hawthorne. 2009. *Relativism and monadic truth*. Oxford University Press.
- Carcassi, Fausto, Shane Steinert-Threlkeld & Jakub Szymanik. 2021. Monotone quantifiers emerge via iterated learning. *Cognitive Science* 45(8). e13027.

- Carcassi, Fausto & Jakub Szymanik. 2021. ‘most’ vs ‘more than half’: An alternatives explanation. *Proceedings of the Society for Computation in Linguistics* 4(1). 334–343.
- Charlow, Simon. 2016. Scalar implicature and exceptional scope. *Ms., Rutgers University*.
- Chemla, Emmanuel, Brian Buccola & Isabelle Dautriche. 2019. Connecting content and logical words. *Journal of Semantics* 36(3). 531–547.
- Chen, K. & H. Tao. 2014. The rise of a high transitivity marker ‘dao’ in contemporary Chinese: Co-evolution of language and society. *Chinese Language and Discourse* 5(1). 25–52.
- Chierchia, Gennaro, Danny Fox & Benjamin Spector. 2012. Scalar implicature as a grammatical phenomenon. In *Semantics*, vol. 3, 2297–2331. Mouton de Gruyter.
- Chomsky, Noam. 1957. Logical structures in language. *American Documentation*.
- Chomsky, Noam. 1995. *The minimalist program*. The MIT Press.
- Christianson, K. 2016. When language comprehension goes wrong for the right reasons: Good-enough, underspecified, or shallow language processing. *The Quarterly Journal of Experimental Psychology* 69. 817–828.
- Coppock, E. 2018. Outlook-based semantics. *Linguistics and Philosophy* 41(2). 125–164.
- Coppock, Elizabeth, Elizabeth Bogal-Allbritten & Golsa Nouri-Hosseini. 2020. Universals in superlative semantics. *Language* 96(3). 471–506.
- Cresswell, M. 1976. The semantics of degree. In B.H. Partee (ed.), *Montague grammar* 261–292, Academic Press.
- Culbertson, Jennifer. 2012. Typological universals as reflections of biased learning: Evidence from artificial language learning. *Language and Linguistics Compass* 6(5). 310–329.
- Culbertson, Jennifer & Kathryn Schuler. 2019. Artificial language learning in children. *Annual Review of Linguistics* 5. 353–373.
- Dalrymple, M., M. Kanazawa, Y. Kim, S. Mchombo & S. Peters. 1998. Reciprocal expressions and the concept of reciprocity. *Linguistics and Philosophy* 21. 159–210.
- Dalrymple, Mary, Makoto Kanazawa, Sam Mchombo & Stanley Peters. 1994. What do reciprocals mean? In *Semantics and linguistic theory*, vol. 4, 61–78.
- Deal, A. R. & V. Hohaus. 2019. Vague predicates, crisp judgments. In *Proceedings of Sinn und Bedeutung (sub)*, vol. 23, 347–364.
- Egan, A. 2010. *Disputing about taste* chap. Disagreement, 247–286. Oxford University Press.
- Egan, A., J. Hawthorne & B. Weatherson. 2005. *Epistemic modals in context* chap. Contextualism in philosophy: Knowledge, meaning and truth, 131–169. Oxford University Press.

- Enguehard, Emile & Benjamin Spector. 2021. Explaining gaps in the logical lexicon of natural languages: A decision-theoretic perspective on the square of aristotle. *Semantics and Pragmatics* 14.
- Erlewine, M. Y. & H. Kotek. 2016. A streamlined approach to online linguistic surveys. *Natural Language and Linguistic Theory* 34(2). 481–495.
- Faller, Martina T. 2002. *semantics and pragmatics of evidentials in cuzco quechua*: Stanford University dissertation.
- Fang, Huilin. 2016. Subjectivity and evaluation in standard setting: a study on mandarin ‘hen’. In Aaron Kaplan, Abby Kaplan, M.K. McCarvel & E.J. Rubin (eds.), *Proceedings of the west coast conference on formal linguistics* 34, .
- Fang, Huilin & T.-H. Jonah Lin. 2008. The mandarin you existential: a verbal analysis. In *Ustwpl*, vol. 4, 43–56.
- Farkas, Donka F. & Kim B. Bruce. 2010. On reacting to assertions and polar questions. *Journal of Semantics* 27(1). 81—118.
- Fernald, T.B. 2000. *Predicates and temporal arguments*. New York: Oxford University Press.
- Ferreira, F. 2003. The misinterpretation of noncanonical sentences. *Cognitive Psychology* 47. 164–203.
- Ferreira, F. & J.M. Henderson. 1998. Syntactic reanalysis, thematic processing, and sentence comprehension. In J.D. Foder & F. Ferreira (eds.), *Reanalysis in sentence processing*, Dordrecht: Kluwer Academic.
- von Fintel, Kai & A.S. Gillies. 2010. Must...stay...strong! *Natural language semantics* 18. 351–383.
- Fleisher, Nicholas. 2013. The dynamics of subjectivity. In *Semantics and linguistic theory*, vol. 23, 276–294.
- Fox, Danny. 2007. Free choice and the theory of scalar implicatures. In Uli Sauerland & Penka Stateva (eds.), *Presupposition and implicature in compositional semantics*, 71–120. Palgrave Macmillan. doi:10.1057/9780230210752_4.
- Francez, I. & A. Koontz-Garboden. 2015. Semantic variation and the grammar of property concepts. *Language* 91(3). 533–563.
- Francez, Itamar & Andrew Koontz-Garboden. 2017. *Semantics and morphosyntactic variation: qualities and the grammar of property concepts*. Oxford: Oxford University Press.
- Frank, Michael C. & Noah D. Goodman. 2012a. Predicting pragmatic reasoning in language games. *Science* 336(6084). 998–998. doi:10.1126/science.1218633.
- Frank, Michael C. & Noah D. Goodman. 2012b. Predicting pragmatics reasoning in language games. *American Association for the Advancement of Science* 336(6084). 998.

- Franke, M. 2011. Quantity implicatures, exhaustive interpretation, and rational conversation. *Semantics and Pragmatics* 4(1). 1–82.
- Franke, M. & Gerhard Jäger. 2016. Probabilistic pragmatics, or why bayes' rule is probably important for pragmatics. *Zeitschrift für Sprachwissenschaft* 35(1). 3–44.
- Franke, Michael & Leon Bergen. 2020. Theory-driven statistical modeling for semantics and pragmatics: A case study on grammatically generated implicature readings. *Language* 96(2).
- Fujita, Koji. 1994. Middle, ergative and passive in english: A minimalist perspective. In *Mit working studys in linguistics*, vol. 22, chap. The morphology-syntax connection, 71—90. MITWPL, Department of Linguistics and Philosophy.
- Gamut, L.T.F. 1991. *Logic, language and meaning, volume 2: Intensional logic and logical grammar*, vol. 2. Chicago, IL: University of Chicago Press.
- Geach, P.T. 1965. Assertion. *The Philosophical Review* 74(4). 449–65.
- Gerke, Berit & Nino Grillo. 2009. How to become passive. In *Explorations of phase theory: Features, arguments, and interpretation at the interfaces*, Kleanthes Grohmann. Mouton de Gruyter.
- Gerken, LouAnn. 2006. Decisions, decisions: infant language learning when multiple generalizations are possible. *Cognition* 98. B67–B74.
- Geurts, Bart. 2010. *Quantity implicatures*. Cambridge University Press, Cambridge.
- Gillon, B. 1990. Ambiguity, generality, and indeterminacy: Tests and definitions. *Synthese* 85. 391–416.
- Gillon, B. 2004. Ambiguity, indeterminacy, deixis, and vagueness — evidence and theory. In Steven Davis & Brendan S. Gillon (eds.), *Semantics: A reader*, 157–190. Oxford University Press.
- Gobeski, Adam. 2019. *quality noun*: Michigan State University dissertation.
- Grano, Thomas. 2008. Mandarin hen and the syntax of declarative clause typing. *Unpublished manuscript*. Accessed online: < http://home.uchicago.edu/~tgrano/grano_hen.pdf>. First accessed 4.
- Grano, Thomas. 2012. Mandarin *hen* and Universal Markedness in gradable adjectives. *Natural Language and Linguistic Theory* 30. 513–565. doi:10.1007/s11049-011-9161-1.
- Grano, Thomas & Stuart Davis. 2018. Universal markedness in gradable adjectives revisited. *Natural Language & Linguistic Theory* 36(1). 131–147.
- Grano, Thomas. & Christopher Kennedy. 2012. Mandarin transitive comparatives and the grammar of measurement. *Journal of East Asian Linguistics* 21. 219–266.
- Grice, H. P. 1989. *Studies in the way of words*. Harvard University Press.

- Grice, H.P. 1975. Logic and conversation. In Peter Cole & Jerry L. Morgan (eds.), *Syntax and semantics*, vol. 3, 41–58. Academic Press, New York.
- Heim, Irene. 1985. *Notes on comparatives and related matters*. University of Texas, Austin.
- Hemforth, B., B. Mertins & C. Fabricius-Hansen. 2014. *Introduction: Meaning across languages*, vol. 44 chap. Introduction: Meaning Across Languages, 41–58. Springer, Cambridge.
- Hohaus, Vera. 2015. *Context and composition: How presuppositions restrict the interpretation of free variables.*: Tübingen: TOBIAS-lib Publikationssystem. dissertation.
- Hohaus, Vera. 2018. How do degrees enter the grammar? language change in samoan from [-dsp] to [+dsp]. In *Proceedings of triplea*, vol. 4, 106–120.
- Hohaus, Vera & M.R. Bochank. 2020. The grammar of degree: Gradability across languages. *Annual Reviews of Linguistics* 6. 235–259.
- Horn, Laurence R. 1972. *On the semantic properties of logical operators in English*. Los Angeles, CA: University of California, Los Angeles dissertation.
- Huang, C-T James. 2015. On syntactic analyticity and parametric theory. In *Chinese syntax in a cross-linguistic perspective*, Oxford University Press.
- Huang, C-T James, Audrey Li & Yafei Li. 2009. *The syntax of chinese*. The Cambridge University Press.
- Jackendoff, R. 2007. *Language, consciousness, culture*. MIT Press.
- Karttunen, L. 1972. Possible and must. In J. Kimball (ed.), *Syntax and semantics*, vol. 1, 1–20. Academic Press, New York.
- Katzir, Roni. 2007. Structurally-defined alternatives. *Linguistics and Philosophy* 30(6). 669–690. doi:10.1007/s10988-008-9029-y.
- Kennedy, Christopher. 1999. *Projecting the adjective: The syntax and semantics of gradability and comparison*. New York: Routledge. Garland.
- Kennedy, Christopher. 2007. Vagueness and grammar: the semantics of relative and absolute gradable predicates. *Linguistics and Philosophy* 30(1-45).
- Kennedy, Christopher. 2013. Two sources of subjectivity: qualitative assessment and dimensional uncertainty. *Inquiry: an interdisciplinary journal of philosophy* 56:2—3(258—277).
- Kennedy, Christopher. 2019. Points of comparison: English, chinese, japanese. In Qiongpeng Luo (ed.), *Degree semantics: an east asian perspective*, Nanjing University.
- Kennedy, Christopher & Malte Willer. 2016. Subjective attitudes and counterstance contingency. In *Semantics and linguistic theory*, vol. 26, 913–933.
- Khoo, Justin. forthcoming. Quasi indexicals. *Philosophy and phenomenological research* .

- Klein, Ewan. 1982. The interpretation of adjectival comparatives¹. *Journal of Linguistics* 18(1). 113–136.
- Kolbel, M. 2003. Faultless disagreement. In *Proceedings of the aristotelian society* 104, 53–73.
- Korotkova, N. 2016. Disagreement with evidentials: a call for subjectivity. In J.Hunter, M.Simons & M.Stone (eds.), *Jersem: The 20th workshop on the semantics and pragmatics of dialogue*, 65–75.
- Krasikova, Sveta. 2008. Comparison in chinese. In O. Bonami & P. Cabredo Hofherr (eds.), *Empirical issues in syntax and semantics*, vol. 7, 263–281.
- Kratzer, A. 2009. Making a pronoun: Fake indexicals as windows into the properties of pronouns. *Linguistic Inquiry* 40(2). 187–237.
- Larson, R., M. denDikken & P. Ludlow. 1997. Intensional transitive verbs and abstract clausal complementation. In A. Grzankowski & M. Montague (eds.), *Non-propositional intentionality*, Oxford: OUP.
- Lasersohn, Peter. 2005. Context dependence, disagreement, and predicates of personal taste. *Linguistics and Philosophy* 28(6). 643—686.
- Lasersohn, Peter. 2017. *Subjectivity and perspective in truth-theoretic semantics*. Oxford University Press.
- Lassiter, Daniel & Noah D. Goodman. 2017. Adjectival vagueness in a bayesian model of interpretation. *Synthese* 194(10). 3801–3836.
- Li, Charles N. & Sandra A. Thompson. 1981. *Mandarin chinese: a functional reference grammar*. Berkeley: University of California Press.
- Li, Xiao. 2017. Measure scales and gradability: on the semantics of the possessive property concept construction in mandarin chinese. In *Proceedings of the 53rd annual meeting of the chicago linguistics society*, Chicago: Chicago Linguistic Society.
- Li, Xiao. 2018. Measurement scales and gradability: on the semantics of the possessive property concept construction in mandarin chinese. In *North american conference on chinese linguistics 30*, The Ohio State University.
- Liu, C-S L. 2010a. The chinese ‘geng’ clausal comparative. *Lingua* 120. 1579–1606.
- Liu, C-S L. 2010b. The positive morpheme in chinese and the adjective structure. *Lingua* 120. 1010–1056.
- Liu, Chen-Sheng Luther. 2018. Projecting adjectives in Chinese. *Journal of East Asian Linguistics* 27. 67–109. doi:10.1007/s10831-018-9166-4.
- MacFarlane, J. 2014. *Assessment sensitivity: relative truth and its applications*. Oxford University Press.
- Maienborn, C. 2005. A discourse-based account of spanish ser/estar. *Linguistics* 43(1). 155–180.

- Martin, J. & P. White. 2005. *The language of evaluation: Appraisal in english*. Palgrave Macmillan.
- Matthewson, L., H. Davis & H. Rullmann. 2007. Evidentials as epistemic modals: evidence from statimcets. *Linguistic Variation Yearbook* 7. 203–256.
- Milner, G.B. 1967. *Fijian grammar*. University of London.
- Mitchell, Jonathan Edward. 1986. *The formal semantics of point of view*: University of Massachusetts, Amherst dissertation.
- Moltmann, F. 1997. Intensional verbs and quantifiers. *Natural language semantics* 5. 1–52.
- Moltmann, F. 2008. Degree structure as trope structure: a trope-based analysis of positive and comparative adjectives. *Linguist and philosophers* 32. 51–94.
- Moltmann, F. 2010. Relative truth and the first person. *Philosophical Studies* 150. 187–220.
- Moltmann, F. 2012. Two kinds of first-person-oriented content. *Synthese* 184(2). 157–177.
- Moracchini, Sophie. 2018. Evaluativity and structural competition. In *Proceedings of semantics and linguistic theory (salt)* 28, vol. 28, 727–746.
- Morzycki, Marcin. 2009. Degree modification of gradable nouns: Size adjectives and adnominal degree morphemes. *Natural language semantics* 17(2). 175–203.
- Morzycki, Marcin. 2011. Metalinguistic comparison in an alternative semantics for imprecision. *Natural language semantics* 19(1). 39–86.
- Morzycki, Marcin. 2012. The several faces of adnominal degree modification. In J. et al. Choi (ed.), *Proceedings of the west coast conference on formal linguistics 29th*, 187–95.
- Morzycki, Marcin. 2018. *Modification—key topics in semantics and pragmatics*. Cambridge University Press, Cambridge.
- Motamedi, Yasamin, Marieke Schouwstra, Kenny Smith, Jennifer Culbertson & Simon Kirby. 2019. Evolving artificial sign languages in the lab: From improvised gesture to systematic sign. *Cognition* 192. 103964.
- Moulton, K. 2009. *Natural selection and the syntax of clausal complementation*: University of Massachusetts at Amherst dissertation.
- Munn, Alan. 1999. First conjunct agreement: Against a clausal analysis. *Linguistic Inquiry* 30. 643–668.
- Munn, Alan & Cristina Schmitt. 2005. Number and indefinites. *Lingua* 115. 821–855.
- Nouwen, Rick. 2007. *Predicates of (im)personal taste*. Utrecht Institute for Linguistics OTS.
- Partee, B.H. 1989. Binding implicit variables in quantified contexts. In C. Wiltshire, B. Music & R. Graczyk (eds.), *Papers from the 25th regional meeting of the chicago linguistic society*, 342–365.

- Pearson, H. 2013a. A judge-free semantics for predicates of personal taste. *Journal of Semantics* 30(1). 103–154.
- Pearson, H. 2013b. *The sense of self: topics in the semantics of de se expressions*: Harvard dissertation.
- Peirce, J, JR Gray & S Simpson. 2019. Psychopy2: Experiments in behavior made easy. *Behav Res* 51. 195–203.
- Pustejovsky, James. 1995. *The generative lexicon*. The MIT Press.
- Qing, Ciyang & Michael C. Frank. 2014. Gradable adjectives, vagueness, and optimal language use: A speaker-oriented model. In *Proceedings of semantics and linguistic theory (salt) 24*, vol. 24, 23–41.
- Rett, J. 2008a. Antonymy and evaluativity. In M. Gibson & T. Friedman (eds.), *Proceedings of salt xvii*, CLC Publications.
- Rett, J. 2008b. *Degree modification in natural language*: Rutgers University dissertation.
- Rett, J. 2014a. Measure phrase equatives and modified numerals. *Journal of Semantics* 32. 425–475.
- Rett, J. & S.E. Murray. 2013. A semantic account of mirative evidentials. In NY. CLC Publications, Ithaca (ed.), *Proceedings of semantics and linguistic theory (salt-23)*, 453–472.
- Rett, Jessica. 2010. Equatives, measure phrases and npis. In *Logic, language and meaning*, 364–373. Springer.
- Rett, Jessica. 2015. *The semantics of evaluativity*. Oxford University Press.
- Rett, Jessica. 2019. Manner implicatures and how to spot them. *International Review of Pragmatics* in press.
- Saebo, Kjell Johan. 2009. Judgment ascriptions. *Linguist and philosophers* 32. 327–352.
- Sassoon, G. 2010. Measurement theory in linguistics. *Synthese* 174. 151–180.
- Sauerland, Uli. 2004. Scalar implicatures in complex sentences. *Linguistics and Philosophy* 27(3). 367–391. doi:10.1023/B:LING.0000023378.71748.db.
- Schutz, A.J. 1875. *The fijian language*. University of Hawaii Press.
- Schwarzschild, R. 2005. Measure phrases as modifiers of adjectives. *Recherches Linguistiques de Vincennes* 34(207-228).
- Smith, Ryan Walter. 2020. Similative plurality and the nature of alternatives. *Semantics and Pragmatics* 13. 15.
- Spector, Benjamin. 2017. The pragmatics of plural predication: Homogeneity and non-maximality within the rational speech act model. In *Proceedings of the 21st amsterdam colloquium*, 435–444.

- Stalnaker, Robert. 1979. Assertion. In Peter Cole (ed.), *Syntax and semantics*, Academic Press, London.
- Stassen, Leon. 1985. *Comparison and universal grammar*. Oxford: Basil Blackwell.
- von Stechow, Arnim. 1984. Comparing semantic theories of comparison. *Journal of semantics* 3(3). 1–77.
- Stephenson, Tamina. 2002. *Toward a theory of subjective meaning*: MIT dissertation.
- Stephenson, Tamina. 2007. Judge dependence, epistemic modals, and predicates of personal taste. *Linguistics and philosophy* 30(4). 487–525.
- Stojanovic, Isidora. 2007. Talking about taste: Disagreement, implicit arguments, and relative truth. *Linguistics and philosophy* 30(6). 691–706.
- Sun, CC, P Hendrix, JQ Ma & RH Baayen. 2018. Chinese lexical database (cld): a large-scale lexical database for simplified chinese. *Manuscript submitted for publication*.
- Sybesma, R. 1999a. *The mandarin vp*. Netherlands: Springer.
- Sybesma, Rint. 1999b. *The Mandarin VP*. Kluwer.
- Tovena, L. 2001. Between mass and count. In Megerdooomian K. & L.A. Bar-el (eds.), *Proceedings of the west coast conference on formal linguistics 20th*, 565–78. Somerville, MA: Cascadilla Proceedings Project.
- Trueswell, J.C., T.N. Medina, A. Hafri & L.R. Gleitman. 2013. Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive psychology* 66(1). 126–156.
- Van Rooij, Robert. 2011. Vagueness and linguistics. In *Vagueness: A guide*, 123–170. Springer.
- Vendler, Z. 1967. *Linguistics in philosophy*. New York: Cornell University Press.
- Willet, Thomas. 1988. A cross-linguistic survey of the grammaticization of evidentiality. *Studies in Language* 12(1). 51–97.
- Wurmbrand, S. 2015. *Fake indexicals, feature sharing, and the importance of gendered relatives*. MIT Linguistics Colloquium.
- Xiang, M. 2008. *Some topics in comparative constructions*: Michigan State University dissertation.
- Xiang, Ming. 2005. *Some topics in comparative constructions*: Michigan State University dissertation.
- Xun, E., G. Rao, X. Xiao & J. Zang. 2016. The development of bcc corpus in the background of big data. *Corpus Linguistics* 3(1).
- Zakkou, J. 2015. *Tasty contextualism*: Humboldt University of Berlin dissertation.
- Zhang, Linmin. 2019a. The semantics of comparisons in Mandarin Chinese. In *GLOW in Asia*, vol. 12, 643–652. <https://ling.auf.net/lingbuzz/004755>.

- Zhang, Linmin & Jia Ling. 2021. The semantics of comparatives. *Journal of Semantics* <https://ling.auf.net/lingbuzz/005223>.
- Zhang, Yiwen. 2019b. Nominal property concepts and substance possession in chinese. Unpublished PhD qualifying paper, Indiana University.
- Zhang, Yiwen & Thomas. Grano. 2019. Stay positive! the mandarin hen puzzle meets possessed property concepts. In Qiongpeng Luo (ed.), *Degree semantics: an east asian perspective*, Nanjing University.