

NORTHWESTERN UNIVERSITY

Disjunctive Questions: Experimental Evidence in Prosody and Discourse

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Linguistics

By

Samantha Heidenreich

EVANSTON, ILLINOIS

December 2019

© Copyright by Samantha Heidenreich 2019

All Rights Reserved

Abstract

A disjunctive interrogative has two different semantic interpretations: as an alternative (Alt.) question, or yes-no (Y/N) question, dependent on prosody. These interpretations also differ in the types of licit answers: Alt. questions demand ‘exactly one’ of the listed alternatives, while Y/N questions can have many answers, including *Yes*, *No*, *Both*, *Neither*, often in combination with any of the disjuncts proposed. Semantic theories that model disjunctive questions and their licit answers rely on many assumptions. First, the models are framed according to assumed canonical prosodic contours, which have been untested in the experimental field. Certain responses are deemed ‘licit’ or ‘illicit’ without regard to syntactic variation or intonational differences. A disjunctive question is assumed unbiased without prosody, but it is assumed that inserting *either* forces the Y/N interpretation. The relationship between Y/N questions and Alt. questions is assumed to be a superset/subset relationship.

Four experiments were conducted in order to examine the underlying assumptions of disjunctive questions and their licit responses. A production experiment revealed the true restrictions on Alt. question contours; Alt. questions do not need a phrase break between disjuncts, but do rely on a contour in which the first disjunct ends in a relatively high pitch range, while the second disjunct (or full utterance) ends in a relatively low pitch range.

The text experiment exposed the inherent bias of the ambiguous string: a disjunctive question is biased to the Y/N interpretation when prosody is absent. Furthermore, licit responses to a disjunctive question lie on a continuous scale, rather than a categorical one. If the response is valid for one of the two interpretations, then participants always seem to rate the response as at least slightly acceptable.

The perception experiment indicated that *either* can be inserted into either type of question and the meaning of the question will be derived from the intonation. The production experiment results support this claim, as participants voluntarily inserted *either* into both Alt. and Y/N productions, when not given a specific utterance prompt.

The artificial language experiment tested participants against different language conditions of the Alt. and Y/N question dichotomy: some conditions had two separate lexemes for the “or” in the question, while others had one; some conditions had monosyndetic constructions (the “or” occurred between disjuncts) while others had bisyndetic constructions (the “or” occurred before both disjuncts), or even a juxtaposition construction (no word for “or”). Results indicated that participants think of Alt. questions as a specific subset of Y/N questions; the text experiment supports this conclusion as well.

The data revealed shortcomings in frameworks intending to model interrogatives and their licit responses. For example, Inquisitive Semantics relies on prosodic focus to distinguish between Alt. questions and Y/N questions. The production experiment revealed both types of questions can be produced with and without prosodic focus on both disjuncts. Commitment Space Discourse, however, models Alt. questions and Y/N questions the same way; it lacks a method to ensure the ‘exactly one’ stipulation on Alt. responses is upheld. Both frameworks required modifications in light of the experimental results. Various mechanisms within each framework were explored in order to amend their shortcomings; however, neither can fully account for the data from the experiments.

ACKNOWLEDGMENTS

It seems insufficient to condense the years of toil and hard work into a page of acknowledgments. Truly, there are many people whose labors go into the completion of a single dissertation. I would not have been able to see this project to its completion without a host of support from my family, friends, and committee members. While these acknowledgements are by no means thanks enough for their efforts, I hope to adequately convey my own appreciation for their respective parts in this culmination of years of research.

First, my eternal gratitude goes out to my committee members: Shari Speer, Carl Pollard, Alexis Wellwood, and my chair, Janet Pierrehumbert; for supporting me through this lengthy process, as well as for graciously giving insight and feedback every step of the way. Their willingness to continue in this endeavor despite locational inconsistencies gave me strength to persevere through many other obstacles. I also must acknowledge Chun Liang Chan, who held videoconference meetings with me on multiple occasions to facilitate the creation of my online experiments; and Matt Goldrick and Masaya Yoshida, both of whom were on the receiving end of my deluge of questions concerning the dissertation process while living remotely.

Then, I must thank several people for their support in non-academic ways: for Tiffany Thornton and Patti Doud, who each watched my two sons for hours or days at a time, so that I could devote the necessary time and energy into this project; for my parents and my husband's parents, who, besides providing childcare support, also lent encouragement in various other ways. I owe you all an enormous debt of gratitude. Finally, my most ardent thanks to Alex, my husband; I truly would not have finished this dissertation without your constant support, companionship, and faith in me. You probably didn't know what you were getting into when you married a graduate student; thank you for going on this journey with me.

DEDICATION

To Alex, who believed in me even when I didn't believe in myself; and to Samuel and David, so they can know that anything is possible.

TABLE OF CONTENTS

Abstract.....	3
ACKNOWLEDGEMENTS.....	5
DEDICATION.....	6
TABLE OF CONTENTS.....	7
LIST OF TABLES.....	12
LIST OF FIGURES.....	17
Chapter 1. Introduction.....	21
Chapter 2. Background.....	31
2.1 Disjunctive Question and Answer Assumptions.....	31
2.2 Disjunctive Question Intonation.....	31
2.3 Disjunctive Question Grammaticalization.....	39
2.4 Disjunctive Question Responses.....	40
2.5 Theoretical Work.....	42
2.5.1 Inquisitive Semantics.....	42
2.5.2. Commitment Space Semantics.....	57
2.6 Typology.....	68
2.7 Overview of Terminology.....	72
2.8 Summary.....	74
Chapter 3. Production Experiment.....	76
3.1 Introduction.....	76
3.2 Methodology.....	76

	8
3.2.1 Design and Materials.....	76
3.2.2 Procedure.....	79
3.2.3 Participants.....	80
3.3 Predictions.....	81
3.4 Results.....	84
3.4.1 Error Analysis.....	84
3.4.2 Standard Error.....	86
3.4.3 Stimuli Analysis.....	86
3.4.4 Prosodic Annotation.....	89
3.4.5 Final Contour.....	93
3.4.5.1 Phrasal Accent: Phonological Properties.....	93
3.4.5.2 D2 Pitch Accent: Phonological Properties.....	95
3.4.5.3 Final Contour: Phonetic Properties.....	96
3.4.5.4 Summary.....	99
3.4.6 Prosodic Focus.....	100
3.4.7 Phrasal Breaks.....	105
3.5 Discussion.....	113
3.6 Conclusion.....	123
Chapter 4. Text Experiment.....	126
4.1 Introduction.....	126
4.2 Methodology.....	126
4.2.1 Design and Materials.....	126
4.2.2 Procedure.....	129

4.2.3 Participants.....	9 130
4.3 Predictions.....	130
4.4 Results.....	135
4.4.1 Filler Analysis.....	135
4.4.2 Stimuli Analysis.....	136
4.5 Discussion.....	141
4.6 Conclusion.....	149
Chapter 5. Perception Experiment.....	151
5.1 Introduction.....	151
5.2 Methodology.....	151
5.2.1 Design and Materials.....	151
5.2.2 Procedure.....	153
5.2.3 Participants.....	154
5.3 Predictions.....	154
5.4 Results.....	157
5.4.1 Filler Analysis.....	157
5.4.2 Stimuli Analysis.....	158
5.5. Discussion.....	165
5.5.1 Alt. Questions.....	165
5.5.2 Y/N Questions.....	168
5.6 Conclusion.....	173
Chapter 6. Artificial Language Experiment.....	177
6.1 Introduction.....	177

	10
6.2 Methodology	177
6.2.1 Design and Materials	177
6.2.2 Language Conditions	183
6.2.3 Procedure	188
6.2.4 Participants	189
6.3. Predictions	189
6.3.1 Superset/Subset Relationship	189
6.3.2 Complementary Relationship	191
6.4 Results	192
6.4.1 Filtering	192
6.4.2 Analysis	194
6.4.2.1 Entire Testing Phase	194
6.4.2.2 Disjunctive Connective	196
6.4.2.3 Response to Question	200
6.5 Discussion	204
6.5.1 Type Errors	204
6.5.2 Ranking Performance by Disjunctive Connective	208
6.5.3 Ranking Performance by Entire Testing Phase	212
6.6 Conclusion	213
Chapter 7. General Discussion	215
7.1 Introduction	215
7.2 Theory predictions and experimental observations	215
7.3 Alternative Questions: phrasal breaks and first disjunct ends ‘high’	227

	11
7.3.1 Inquisitive Semantics and Alternative Questions.....	228
7.3.2 Commitment Space Discourse and Alternative Questions.....	230
7.4 The presence of <i>either</i> in disjunctive questions.....	238
7.5 The tune of the response.....	241
7.5.1 Inquisitive Semantics and response tunes.....	241
7.5.2 Commitment Space Discourse and response tunes.....	244
7.6 Presence of <i>either</i> in questions altering answerhood acceptability.....	246
7.6.1 Inquisitive Semantics and the interaction with <i>either</i>	248
7.6.2 Commitment Space Discourse and the <i>either</i> interaction.....	251
7.7 Continuous Scale of Acceptability for Answers.....	253
7.7.1 Inquisitive Semantics and the continuous scale.....	254
7.7.2 Commitment Space Discourse and answers on a continuum.....	258
7.8 Location of Disjunction in Question.....	264
7.8.1 Inquisitive Semantics and disjunct location.....	266
7.8.2 Commitment Space Discourse and question disjunct location.....	267
7.9 Conclusion and Further Work.....	269
References	272
Appendix A.....	282
Appendix B.....	291

LIST OF TABLES

Table 2.1: Percent correct of all items broken into utterance/context combinations.....	37
Table 2.2: Disjunctive Questions and licit answers according to Interrogative Semantics.....	56
Table 2.3: Disjunctive Questions and licit answers according to CSD.....	68
Table 2.4 Summary of terminology of disjunctive questions for IS, CSD, and Typology.....	73
Table 3.1: Predictions of tune based on previous empirical results.....	82
Table 3.2: Predictions of tune based on Inquisitive Semantics theory.....	82
Table 3.3: Predictions of tune based on Commitment Space Discourse theory.....	83
Table 3.4: Breakdown of Errors in Production Data.....	85
Table 3.5: Best fit model parameters for the production data.....	87
Table 3.6: Breakdown of Errors from 31 participants.....	91
Table 3.7: Best fit model for filtered production data.....	91
Table 3.8: Glmer with D2ip as dependent variable.....	94
Table 3.9: Final phrasal accent across question Type, including and excluding errors.....	94
Table 3.10: Second Disjunct pitch accent distribution across question type.....	95
Table 3.11: Second Disjunct accent placement across question type, excluding errors.....	96
Table 3.12: Lmer with final contour (scaled in ST) as dependent variable.....	97
Table 3.13: Linear Discriminant Analysis based on change from D2Pitch to D2ipPitch.....	99
Table 3.14: Predictions and Observations for the Final Contour of Disjunctive Questions.....	100
Table 3.15: Occurrence of pitch accent types on the second disjunct, by question type.....	101
Table 3.16: Average F0 across L* productions on the second disjunct.....	102

	13
Table 3.17: The percent of first disjunct pitch accent distribution across question type.....	102
Table 3.18: Theories and predictions on prosodic focus occurrence.....	104
Table 3.19: Treatment of Y/N questions under IS and CSD.....	105
Table 3.20: Predictions of phrasal breaks across theories by question type.....	106
Table 3.21: Phrasal Break patterns for both types of disjunctive questions.....	106
Table 3.22: Logistic model with number of lists as dependent factor.....	107
Table 3.23: Linear model with duration of first disjunct as dependent factor.....	109
Table 3.20 (updated): Predictions and observations of phrasal breaks.....	110
Table 3.24: Breakdown of first disjunct pitch accents without a phrasal break.....	111
Table 3.20 (updated): Predictions and observations of phrasal breaks.....	112
Table 3.25: Theory predictions, observations, and conclusions from the production data.....	114
Table 3.26: Logistic mixed effect model with dependent variable as question type	117
Table 3.27: Logistic mixed effect model scaled in ST (base 100 Hz).....	120
Table 4.1: Permissible answers to alternative question interpretation, by theory.....	131
Table 4.2: Permissible answers to Y/N interpretation, by theory	132
Table 4.3: Filler Scoring.....	135
Table 4.4: Average rating by question type and answer type.....	137
Table 4.5: Model of best fit for ordinal data.....	138
Table 4.6: Average Rating across answer types.....	141
Table 4.7: Answer types categorized into tiers by average observed rating.....	142
Table 4.8: Metrics of fifteen participants used in RT analysis, in comparison to full dataset....	146
Table 4.9: Linear mixed-regression model of best fit for reaction data of 15 participants.....	146

Table 4.10: Average reaction time (in seconds) for stimuli by location of the disjunct in the question, whole dataset.....	147
Table 4.11: Linear mixed-regression model of best fit for reaction data of all participants.....	148
Table 5.1: Inquisitive Semantics predictions across question-answer combinations.....	154
Table 5.2: Commitment Space Discourse Predictions.....	156
Table 5.3: Predictions on answer conditions for alternative question tune with <i>either</i>	156
Table 5.4: Filler Scoring.....	157
Table 5.5: Fillers removed.....	158
Table 5.6: Model of best fit for ordinal perception data.....	159
Table 5.7: Average rating over <i>either</i> condition and answer condition.....	161
Table 5.8: Average rating over question tune and answer condition.....	162
Table 5.9: Predictions and observations of Alt tune across <i>either</i> condition and answer conditions.....	165
Table 5.10: Predictions and observations of Y/N tune across <i>either</i> condition and answer conditions.....	168
Table 5.11: Regression model statistics for Y/N data.....	171
Table 5.12: Regression model statistics for Y/N data with interaction.....	171
Table 5.13: Comparison of average ratings when looking at all Y/N questions vs. those with rising accents on both disjuncts, across <i>either</i> condition and answer conditions.....	172
Table 6.1: Words used in the experiment and counterbalancing measures.....	179

Table 6.2: Delineation of each language condition according to # of distinct connectives and type of construction of connective.....	187
Table 6.3: Best fit model for the entire dataset	195
Table 6.4: Percent correct in testing phase when “or” was the missing factor.....	197
Table 6.5: Model of best fit for testing phase data when disjunctive connective was missing...197	
Table 6.6: Percent correct in testing phase when the response was the missing factor.....	200
Table 6.7: Model of best fit for testing phase data when the response was missing.....	201
Table 6.8: Percent correct of Y/N questions when the response was missing.....	203
Table 6.9: Comparison of percent correct of Y/N questions when response was missing.....	203
Table 6.10: Percent correct of testing phase where the response was the missing factor.....	204
Table 6.11: Percent correct, percent type error, and total percent stimuli accounted for by the two types when “or” was the missing factor, across conditions.....	205
Table 6.12: Participant 52 data from the testing phase.....	206
Table 6.13: Language conditions ranked by percent % when participants had to fill in the disjunctive connective.....	210
Table 6.14: Language conditions ranked by % correct in testing phase, overall.....	212
Table 7.1: Theory predictions, observations, and conclusions from the production data.....	216
Table 7.2: Permissible answers to alternative questions and various interpretations of Y/N questions across Inquisitive Semantics (IS) and Commitment Space Discourse (CSD).....	217
Table 7.3: Average rating across answer condition, categorized into tiers of acceptability.....	218
Table 7.4: Acceptability ratings for text and perception experiments.....	219
Table 7.5: Results from experiments that are currently not accounted for by formal theory.....	227

	16
Table 7.6: Criteria list for Inquisitive Semantics.....	230
Table 7.7: Criteria list for Commitment Space Discourse.....	238
Table 7.6 (updated): Criteria list for Inquisitive Semantics.....	240
Table 7.7 (updated): Criteria list for Commitment Space Discourse.....	240
Table 7.8: Ratings from Perception Experiment, across question tune and answer condition...	241
Table 7.6 (updated): Criteria list for Inquisitive Semantics.....	244
Table 7.7 (updated): Criteria list for Commitment Space Discourse.....	246
Table 7.9: Ratings across answer condition, tune, and <i>either</i> in perception experiment.....	247
Table 7.6 (updated): Criteria list for Inquisitive Semantics.....	251
Table 7.7 (updated): Criteria list for Commitment Space Discourse.....	252
Table 7.10: Average ‘closure’ slope by question type, answer condition, and given rating.....	255
Table 7.6 (updated): Criteria list for Inquisitive Semantics.....	258
Table 7.11: Continuum of Acceptability, based on text and perception results.....	259
Table 7.12: Rankings of alternative question responses in CSD.....	261
Table 7.7 (updated): Criteria list for Commitment Space Discourse.....	264
Table 7.13: Average ‘closure slope’ of alternative question.....	266
Table 7.6 (updated): Criteria list for Inquisitive Semantics.....	267
Table 7.7 (updated): Criteria list for Commitment Space Discourse.....	269

LIST OF FIGURES

Figure 2.1: Maximal sets for (15).....	45
Figure 2.2: Applying Exclusive Strengthening.....	51
Figure 2.3: Closed intonational pattern (two lists).....	54
Figure 2.4: Open intonational pattern (two lists).....	55
Figure 2.5: Block intonational pattern (one list).....	55
Figure 2.6: Update of Commitment Space.....	58
Figure 2.7: Polar Question in CSD.....	61
Figure 2.8: Alternative Question Disjunction in CSD.....	63
Figure 2.9: Disjunction question as disjunction of two monopolar questions.....	64
Figure 2.10: Relationship of licit responses for different Y/N questions across IS and CSD.....	73
Figure 3.1: Alternative Question scenario example.....	77
Figure 3.2: Y/N question scenario example.....	78
Figure 3.3: Baseline example (simple polar question).....	78
Figure 3.4: Mean of Correct value grouped by disjunct Type.....	88
Figure 3.5: Mean of Correct value grouped by Half.....	89
Figure 3.6: Mean of Correct value grouped by Disjunct Type.....	92
Figure 3.7: Mean of Correct value grouped by Half.....	93
Figure 3.8: L*+H pitch accent on second disjunct (chameleon) of Alt. question.....	95
Figure 3.9: Average F0 change (in semitones) of the Final Fall.....	98
Figure 3.10: Counts of Recordings without a phrasal break.....	108

	18
Figure 3.11: Counts of Recordings with a phrasal break.....	108
Figure 3.12: Mean length of first disjunct by type and gender.....	109
Figure 3.13: Counts of boundary tones separated by disjunct type.....	118
Figure 3.14: Counts of final phrase accent separated by disjunct type.....	118
Figure 3.15: Counts of phrase break between disjuncts, separated by disjunct type.....	119
Figure 3.16: Counts of accent on first disjunct, separated by disjunct type.....	119
Figure 3.17: Mean scaled pitch of second disjunct, grouped by disjunct type.....	121
Figure 3.18: Mean scaled pitch of final phrase accent grouped by disjunct type.....	121
Figure 3.19: Mean scaled pitch at the end of the first disjunct, grouped by disjunct type.....	121
Figure 3.20: Average pitch across full utterance.....	122
Figure 4.1: Mean of acceptability rating by answer condition.....	139
Figure 4.2: Mean of acceptability rating by disjunct location in question.....	139
Figure 4.3: Mean of acceptability rating by answer condition and disjunct location.....	140
Figure 4.4: Reaction Time mean in seconds, grouped by location of disjunct in the question...147	
Figure 5.1: Mean Rating collapsed across location of disjunct in the question.....	160
Figure 5.2: Mean Rating collapsed over answer condition.....	160
Figure 5.3: Mean rating collapsed over answer condition and <i>either</i> condition.....	161
Figure 5.4: Mean Rating over question tune and answer condition.....	164
Figure 5.5: Mean rating of question-answer pairs with the Alt tune.....	166
Figure 5.6: Mean rating of question-answer pairs with the Y/N tune.....	169
Figure 5.7: Mean rating for Y/N tune questions.....	172

Figure 6.1: Introduction to declarative phrase in training.....	179
Figure 6.2: Introduction to interrogative phrase in training.....	180
Figure 6.3: Exposure phase, alternative question example.....	180
Figure 6.4: Feedback phase, response completion example.....	181
Figure 6.5: Feedback phase, question completion example.....	182
Figure 6.6: Histogram of total attempts by participant.....	193
Figure 6.7: Percent correct across significant factor: Disjunct Type.....	195
Figure 6.8: Percent correct across the significant factor: Missing Factor.....	195
Figure 6.9: Percent correct across version when “or” was missing.....	198
Figure 6.10: Percent correct of connective across Version and Disjunct Type.....	199
Figure 6.11: Percent correct of response across Disjunct Type.....	202
Figure 7.1: Plain disjunct answer condition across text and perception experiments.....	221
Figure 7.2: Yes + disjunct answer condition across text and perception experiments.....	221
Figure 7.3: Cleft disjunct answer condition across text and perception experiments.....	222
Figure 7.4: IS account of an alternative question without a phrase break.....	229
Figure 7.5: Alternative question: Disjunction of two monopolar questions.....	231
Figure 7.6: Alternative question: Disjunction of monopolar questions.....	231
Figure 7.7: Polar question: Disjunctive question of two monopolar questions.....	232
Figure 7.8: Alternative Question with licit continuations.....	233
Figure 7.9: Constituent question as disjunction of monopolar questions.....	237
Figure 7.10: He is going to Spain↓, analyzed in the IS framework.....	242

	20
Figure 7.11: He is going to Spain [↑] , analyzed in the IS framework.....	242
Figure 7.12: Average closure slope on acceptability scale for Alternative Questions.....	256
Figure 7.13: Ratings averaged over disjunct location in question, by answer condition.....	265

Chapter 1: Introduction

Disjunctive interrogatives are interesting phenomena in English. The same string of words can have two different semantic interpretations, dependent on the prosody. Take (1) for example; (1a) is the ambiguous string, while (1b) and (1c) contain two different paraphrases for that string.

- (1) a. Is Mary allergic to dairy or soy?
 b. *Which one of these is Mary allergic to: dairy or soy?*
 c. *Is Mary allergic to any of these: dairy or soy?*

The first interpretation (1b) expresses an alternative question (Alt.). In this case, it is supposed that Mary is allergic to something, and that the thing she is allergic to is one of the two disjuncts. Importantly, this type of semantic interpretation of the ambiguous string (1a) makes answers such as *neither* or *both* unacceptable without some type of dismissal of this supposition (i.e. *Well actually, she is allergic to both*). The second interpretation (1c) expresses a yes/no question (Y/N). Here, the distinction between the two disjuncts is not important or salient. Additionally, acceptable responses are less limited, allowing either of the two disjuncts, as well as *neither* or *both*. This type of interpretation may also allow polarity particles (*Yes, No*) as part of the answer. Thus, the two interpretations not only differ in meanings, but in the types of appropriate responses.

Disjunctive interrogatives can be disambiguated by the prosody of the sentence. Previous accounts on the prosody of disjunctive questions (Bartels 1999) stated that an Alt. interpretation would need an emphasis on both disjuncts, have a H- phrasal break between disjuncts, and end with a falling pitch. This intonational structure is called the canonical representation for an Alt. question. A Y/N interpretation would not have particular emphasis on either disjunct, with rising

intonation at the end of the phrase- the canonical representation for a Y/N question. Although the canonical representations had not been tested under the scrutiny of experimental work, these same representations are being used in the formulation of semantic theories of questions. Pruitt & Roelofsen (2013) provided the first experimental study on disjunctive questions; they aimed to discern whether the pitch accents, final contour, or combination of the two contribute to the interpretation. Both the final contour of the Alt. (prosodic phrasing) and the accentual characteristics play a crucial role in disambiguation. However, the study used question tunes that were synthetically pieced together from the aforementioned canonical representations (2) in order to gain interpretational judgments. The underlying assumption of canonical representations remains to be validated.

(2) Canonical Representations of Alt. and Y/N Questions

a. Alternative Question intonation

Would you like mineral water or lemonade?

(H*/L* H-) (H* L-L%)

b. Y/N Question intonation

Would you like mineral water or lemonade?

([H*/L*/L*+H]₁ H*/L* H-H%)

In fact, production data of 24 disjunctive question minimal pairs across three participants (Heidenreich 2014b) revealed that participants don't always use the tunes in (2) to convey a specific interpretation. Although the canonical structures were the most frequently produced patterns, there was also substantial variation from it in the Alt. and *especially* in the Y/N. For

¹ Indicates optional, not necessary, pitch accent

Alt. questions, it seems that emphasis on both disjuncts is necessary but not sufficient. It was also noted that the first disjunct had to end high (either with a H*, L+H*, or H-) and the second disjunct had to end low (either with a L-or L%. etc.); a H- phrasal break was not necessary for the interpretation. The Y/N questions, meanwhile, were extremely variable in their productions and did not appear to have any singular defining characteristics, other than that they were not produced like the Alt.'s. These results do not support the current canonical representations, nor any semantic theory that uses such representations as the means to disambiguate disjunctive interrogatives. Thus, although it is known that the prosodic structure disambiguates the two interpretations, the exact prosodic structures (or the aspects of the prosodic structure) that disambiguate the string are as yet undefined. The aim of this dissertation is to test current underlying assumptions about the production and comprehension of disjunctive interrogatives against experimental data and weigh semantic theories and their predictions against this data.

The first underlying assumption, the canonical representations of Alt. and Y/N question tunes, is addressed in a production experiment. Heidenreich (2014b) revealed that participants were successfully able to produce different intonations for each interpretation when given only a textual context (no audio). They were, however, given an exact phrase to utter. The production experiment in this paper used image context to elicit the two different interpretations; participants were not given exact phrases to utter, nor any sample audio. Despite this lack of direction, participants were able to successfully produce both Alt. and Y/N tunes based on image context. The tune contours showed more variation than canonical representations (Bartels 1999), indicating that theories that rely on *only* the canonical representations (especially of Alt. questions) will not account for the entire spectrum of permissible Alt. contours.

Another underlying assumption of disjunctive interrogatives is that a disjunctive string can instead undergo a grammatical change in order to differentiate between the two interpretations. By adding *either* into a disjunctive string, the sentence can now only be interpreted as a Y/N question (Huddleston 1994, Haspelmath 2007). While (3a) is an ambiguous string, (3b) can only have the interpretation ‘Is Mary allergic to any of these things: dairy or soy?’

(3a) Is Mary allergic to dairy or soy?

(3b) Is Mary allergic to either dairy or soy?

Only author intuition and introspection are given as proof for this claim; no experimental data supports the analysis. Two experiments, a text experiment and a perception experiment, test this assumption against native speaker acceptability ratings. Once again, the underlying assumption does not hold up to experimental scrutiny. The experiments reveal participants’ acceptance of *either* in both contexts, not simply the Y/N context. The production experiment also reveals that participants will put *either*, unprompted, into both contexts², further supporting the notion that *either* does not lexically differentiate between the two. Nevertheless, *either* does have an unforeseen interaction with the acceptability of responses, increasing the acceptability of typically ‘illicit’ responses for that interpretation, and increasing acceptability of responses that incorporate uncertainty (e.g. responses uttered with the uncertainty contour at the end of the utterance, represented as L-H%).

However, licit responses to each interpretation have also been accepted as categorically sound, despite the lack of empirical work to support the answerhood conditions. An Alt.

² Any single participant would put *either* into one of the two contexts; no participant put *either* into both contexts.

question ought only to accept a disjunct as an answer; any other response ought to be invalid (Aloni & van Rooy 2002). But theories do not account for lexical or intonational differences in how this disjunct may be realized. The text and perception experiments also gauge acceptability of these different types of ‘disjunct’ answers (4a-c).

(4) Is Mary allergic to dairy or soy?

(4a) Falling Disjunct (realized with falling prosody): “She’s allergic to dairy.”

(4b) Rising Disjunct (realized with a continuation rise): “She’s allergic to dairy.”

(4c) Cleft Disjunct: “It’s dairy.”

Although (4a-c) are all ‘disjunct’ answers, their acceptability as a licit response differs across the two question interpretations. The text and perception experiments offer further insight into a continuous, rather than categorical, scale of acceptability regarding responses to disjunctive questions. Furthermore, the acceptability of a response, as shown in both the text and perception experiments, can change depending on where in the question the disjunction is located, challenging conventional wisdom that a response is always ‘licit’ or ‘illicit’ regardless of where the disjunct occurs in the question.

Finally, disjunctive strings are disambiguated differently across languages. Mauri (2008) notes that, typologically speaking, many languages do not have a disjunctive connective at all, but rather encode the concept of alternative by an overt ‘irreality’ marker (expressions encoding possibility, future, question, etc.). Other languages, such as Finnish, Polish (5), Mandarin Chinese, and Basque, have two specific disjunctive connectives; one for the standard alternative relation (this is the disjunction found in declaratives and polar questions, 5a) and the other for the interrogative alternative relation (the disjunction in an alternative question, 5b).

(5) Polish, Slavic, Indo-European (Mauri 2008)

a. Zazwyczaj pisze **lub** czytam az do pozna
 usually write.PRS.1SG **ALTNs**³ read.PRS.1SG until to ate

“Usually I write or I read until late.”

b. Idziemy jutro do szkoły **czy** zostajemy w domu?

Go.PRS.1PL tomorrow to school **ALTi** stay.PRS.1PL at home

“Do we go to school tomorrow or do we stay at home?”

In a standard disjunction (e.g. polar question), no choice is needed; a choice is not essential (Mauri 2008). By contrast, an interrogative disjunction always conveys the alternative (question) relation: the two disjuncts are presented as alternatives, there is an immediate request for a choice between the two disjuncts, and the addressee is asked to specify one of the alternatives (Haspelmath 2007, Mauri 2008). Additionally, the two have a superset/subset relationship, whereby the standard disjunction can occur in all clause types, but the interrogative disjunction only occurs in interrogatives (Haspelmath 2007, Mauri 2008). Importantly, Haspelmath dispels the idea that a language with two distinct words for the standard and interrogative disjunction indicates an inclusive vs. exclusive relationship (2007); in fact, no language is known to delineate the inclusive/exclusive relationship with two distinct words. Ariel & Mauri (2019) analyzed a corpus that contained over 1,000 examples of *or* in the Santa Barbara Corpus of spoken American English and found several different variations in the semantic meaning of *or*, rather than a simple inclusive/exclusive relationship. While this corpus does provide insight into the

³ **ALTS** indicates the standard disjunction; **ALTi** indicates the interrogative disjunction

semantic meaning(s) of *or*, nearly all of the *or* productions were in declarative sentences, and therefore the corpus cannot be used in the study of disjunctive interrogatives.

The superset/subset relationship is also played out in the answerhood conditions, where a standard disjunction allows a wide variety of answers (*Yes, Both, Neither*, etc.), while an interrogative disjunction is restricted to a subset of those answers (single disjunct).

Other languages, such as Somali, have two different words, but the standard disjunctive connective occurs bisyndetically (occurs before both disjuncts, 6a), while the interrogative disjunctive connective occurs only between disjuncts, monosyndetically (6b).

(6) Somali, Cushitic, Afro-Asiatic (Saeed 1993, 275)

- a. *amá wuu kéeni doonaa amá wuu sóo.díri doonaa*
ALTS 3SG bring that **ALTS** 3SG send that
 ‘Either he will bring it or he will send it.’
- b. *ma tégaysaa misé waad jóogaysaa?*
 INT go:2SG **ALT**i here stay:2SG
 ‘Are you going or are you staying?’

Languages like Korean (7) omit a connective in the alternative case altogether (7a), relying on question particles alone. The languages do still have a disjunctive connective for the standard alternative relation (7b).

(7) Korean (Sohn 2001, 307 and 305, respectively)

- a. *Wuli ka ka-l kka-yo↑ Mia lul ponay-l kka-yo?↓*
 we NM go-PRS whether-POL Mia AC send-PRS whether-POL
 ‘Shall we go, or shall we send Mia?’
- b. *Kiho ka w-ass-kena/tunci Nami ka w-ass-e*
 Kiho NM come-PST-**ALTS** Nami NM come-PST-INT
 ‘Either Kiho or Namie came.’

The number of languages that utilize these various methods to disambiguate disjunction varies. Mauri (2008) concludes the typological study by conjecturing that if a language has an interrogative connective, then it must have a standard connective. This prediction is supported in the final experiment in the paper, an artificial language experiment. Participants struggled to learn a language that had a word for the interrogative connective but none for the standard connective, supporting the conclusion that a disambiguation technique that does not occur in a natural language may not occur because it is more difficult to learn. Furthermore, participants had better performance in the artificial language experiment when their language condition incorporated a standard disjunction connective that had more lexical appearances in the sentence than the interrogative disjunction connective; the language conditions that were based on Somali (standard disjunction: bisyndetic; interrogative disjunction: monosyndetic) or Korean (standard disjunction: monosyndetic; interrogative disjunction: no connective) were easier to learn than those in which both connectives appeared in the same manner (i.e. the language condition based on Finnish; standard disjunction and interrogative disjunction: monosyndetic, but differentiated by two different lexical words). When participants had to enter the correct disjunctive connective, the language conditions' ease of learnability aligned with the predictions associated with a superset/subset relationship. When participants had to enter a correct response to a question, they performed significantly better for Alt. questions than for Y/N questions. Since type frequency between the two question types was held constant, the higher accuracy for Alt. questions (the subset) supports the superset/subset hypothesis (Heidenreich 2014a).

Two conversation-driven semantic frameworks that focus not only on questions, but their licit answers, are examined in conjunction to the experiments outlined. The first framework, Inquisitive Semantics (Ciardelli et. al. 2013a, Ciardelli et. al. 2013b, Roelofsen 2013b, among

others), has roots in alternative semantics, and was developed in order to account for all types of interrogatives. The semantic theory incorporates the exchange of information into its model. The framework has incorporated specific mechanisms in order to handle the ambiguity in disjunctive strings (Roelofsen and van Gool 2010), which allows the theory to not only predict the different semantic interpretations, but also the different answerhood conditions for each interpretation. The second framework, Commitment Space Discourse (Krifka 2015, Krifka 2016), was formulated as a response to Inquisitive Semantics. The theory aimed to account for bias in questions without resorting to extraneous means to highlight one option over the other (Krifka 2015). This framework relies on speech acts to update a commitment space (common ground) to include information agreed on by the interlocutors and restrict projected continuations. Both models provide means by which the two interpretations of the questions have different answerhood conditions- namely, the Alt. interpretation has the two disjuncts as possible answers, while the Y/N interpretation has additional answer possibilities. Additionally, both frameworks use only the canonical representations of the question intonations in order to disambiguate the two semantic interpretations. When the underlying assumptions on these intonational contours prove to be inaccurate, based on the data from the experiments, the models undergo adjustment in order to account for the new factors. Direction for future work is also discussed.

By the end of the dissertation, disjunctive interrogatives obtain experimental data to support the prosodic contour disambiguation, different answerhood conditions, and subset/superset relationship between the Alt. and Y/N interpretations. Additionally, previous assumptions concerning disjunctive interrogatives are refuted, such as the grammaticalized use of *either* and the categorical distinction between a licit and illicit answer. Future work on

disjunctive interrogatives, including question-driven semantic theories, can forego the representations made via introspection alone in favor of these experimental results.

The dissertation is divided into the following sections: Chapter 2 contains an extensive background of disjunctive questions, a typological overview, and an in-depth look at the theories examined, as well as their predictions on question tune and licit responses; Chapter 3 contains the methodology, results, and discussion on the production experiment, as well as outlining the specific predictions from the two semantic theories and the comparison to the empirical results; Chapters 4 and 5 follow the same outline for the text and perception experiments, respectively; Chapter 6 contains the methodology, results, and discussion on the artificial language learning experiment and the implications on the relationship between Y/N and Alt. interpretations; Chapter 7 is a discussion combining the results of all experiments, including the modifications necessary for the semantic theories in order to account for the results. Chapter 7 also incorporates the conclusions from the study with suggestions for future work.

Chapter 2: Background

2.1 Disjunctive Question and Answer Assumptions

The previous introductory chapter revealed key assumptions concerning disjunctive questions and answers, many of which lack any empirical basis. In previous work, these assumptions are the foundational building blocks of current semantic theory. The bulk of the dissertation tests these assumptions in a series of experiments. Chapter 2 provides additional background and explores the empirical data concerning disjunctive questions, as well as the theoretical data and formalisms. First, the intonation of disjunctive questions across experimental data is explored, offering predictions for the different contours for Alt. vs. Y/N questions in the production experiment (Chapter 3). Next, assumptions concerning the grammaticalization and ambiguity of disjunctive questions are explored; the appropriate responses to disjunctive questions are also examined, and these assumptions form the crux of the text experiment (Chapter 4) and the perception experiment (Chapter 5). Then, both Inquisitive Semantics and Commitment Space Semantics offer formalisms in the question-answer exchange, providing predictions for licit answers to both interpretations. These predictions are compared to the experimental data in Chapters 3, 4, and 5. Finally, a brief typological overview of disjunction is provided in order to test the type of relationship between the Alt. question and Y/N question (artificial language experiment, Chapter 6) and infix all experimental discoveries into the broader relationship between the two.

2.2 Disjunctive Question Intonation

The exact relationship between alternative questions and polar yes/no questions has been characterized in different ways in literature. Bartels (1999) described the intonational pattern of alternative questions as a “coherent class of utterances, whereas yes/no-questions are not” (83).

Alternative questions do seem to follow a more specified tune, while their polar question counterparts remain varied. The typical intonational feature of alternative questions frequently mentioned in literature is the final fall at the conclusion of the utterance (Schubiger 1958; Quirk et al 1985). However, the full canonical alternative question intonation (8), according to Pruitt & Roelofsen (2013), incorporates a pitch accent on each disjunct, with a H- phrasal accent after the first disjunct and a L- phrasal accent after the second disjunct. The utterance then ends with a falling intonational phrase boundary.

(8) Pruitt & Roelofsen's (2013) canonical alternative question intonations

Would you like mineral water or lemonade?

- a. (H* H-) (H* L-L%)
 b. (L* H-) (H* L-L%)

This intonation pattern does not license a yes-no response; rather one of the two disjuncts must be given. Importantly, the speaker suggests that exactly one of the disjuncts is true. The canonical disjunctive yes/no intonation (9) allows for more variation, such as presence/absence of pitch accents and their values. If disjunctive yes/no questions pattern like other yes/no questions, then disjunctive yes/no questions may even allow rises and falls as final contours as context permits (Hedberg et. al. 2004).

(9) Pruitt and Roelofsen's (2013) canonical disjunctive yes/no question prosody

Would you like mineral water or lemonade?

- a. ((H*) H* H-H%)
 b. ((L*) L* H-H%)

Yes/no questions do not canonically have prosodic phrase breaks or phrase accents between disjuncts (Pruitt & Roelofsen 2013); however some argue that many yes/no questions indeed

employ both (Farkas & Roelofsen 2015, Ciaredelli 2013a, Roelofsen 2013b, Krifka 2015). A pitch accent may be present on the second of two disjuncts, but when present it usually coincides with the nuclear accent of the utterance (i.e. the second disjunct is utterance final).

Historically, focus and semantic theories have assumed that pitch accents carry the weight in disambiguating between the two structures (Han and Romero 2004a,b, Aloni and van Rooy 2002, Romero and Han 2003, Beck and Kim 2006). The pitch accents are said to be markers of ‘focus stress’ on the disjuncts (Han and Romero 2004a,b). Bartels (1999) acknowledges the importance of accentual differences but emphasizes that the final contour- specifically the low phrasal accent- is an intonational morpheme inextricably tied to alternative question meaning. Pruitt & Roelofsen (2013) conducted a study in which they split the important intonational factors into two dimensions, ACCENT and FINAL. The ACCENT dimension of a yes/no question included the single pitch accent on the nuclear stress and no prosodic breaks (represented by ‘S’ for ‘single’); for an alternative question, it included the pitch accents on both disjuncts and the phrasal accent between them (represented by ‘M’ for ‘multiple’). FINAL described the final contour of the utterance; for yes/no questions, this was given as H-H% (↑), whereas alternative questions were L-L% (↓). The factors were crossed in a two by two design, with non-canonical intonational structures (e.g. ACCENT dimension of alternative questions with FINAL dimension of yes/no questions) created by splicing together the respective parts. Participants heard a sentence with one of the four contours and were required to choose the best paraphrase: an alternative question paraphrase, a yes/no question paraphrase, and ‘other’. The study determined that both the target prosodic features are influential, with the final contour the most important of the features in disambiguation.

There were a few specific limitations of this study. The non-canonical intonational patterns (10) were produced via splicing⁴. White et. al. (2009) found that non-natural speech (even excellent quality stimuli) failed to produce the facilitative effect of contextually appropriate accent patterns in human speech. In their study, many participants were unaware until the debrief that the speech they heard was not natural; yet the speech still failed to produce facilitative effects. That is, when the meaning of the utterance is ambiguous, it is possible that discontinuous, spliced, or synthetic speech may not sway interpretation correctly, even with appropriate intonation patterns.

(10) Non-canonical intonational patterns in Pruitt & Roelofsen (2013)

- a. Multiple accents with rise at the end (M↑)

Did Sally bring wine or bake a dessert?

(H* H-) (L* H-H%)

- b. Single accent with fall at the end (S↓)

Did Sally bring wine or bake a dessert?

((L*) H* L-L%)

In addition, the study contained stimuli with disjuncts that occur only at the end of sentences, which combined the phrasal accent with the phrasal boundary. While the final fall has been thought to be important in alternative questions (Rando 1980, Quirk et. al. 1985, among others), Bartels (1999) argued that the phrasal accent, more than the phrasal boundary, was important in

⁴ Splicing was performed at the beginning of the word with the nuclear accent, and encompassed the entire final contour (pitch accent, phrase accent, and boundary tone) in order to switch e.g. the final rise in a single accent utterance with a final fall (10b).

determining the type of disjunctive question. The study also confounded the phrasal accent and pitch accent dimensions by combining the pitch accents of the two disjuncts with the possible phrasal accent between them. As previous research hypothesized that the phrasal accent L- at the end of the second disjunct in alternative questions is important, so too might be the phrasal accent putatively between the disjuncts. Importantly, the phrasal accents ought to be analyzed separately from pitch accents; it may prove that one informs the interpretation more so than the other. Finally, the study used only canonical intonational structures, despite yes/no questions specifically exhibiting such variability, and despite any empirical test to determine what constitutes 'canonical'. Questions that are intoned differently from either canonical pattern may reveal which factors sway interpretation to one type of question versus the other.

Heidenreich (2014b) attempted to tease apart some of these issues by performing a small production experiment to elicit alternative questions and yes/no questions without introducing the participants to any canonical tune. Twenty-four alternative questions, yes-no questions, and alternative declarative sentences each were recorded by three participants, for a total of seventy-two utterances for each participant. The questions themselves were designed in order to answer some outstanding questions. For example, the disjuncts occurred only at the beginning or middle of the sentence, thereby separating the phrasal accent from the phrasal boundary. Each disjunct was at minimum three syllables, with the primary stress on the antepenultimate syllable, to separate the pitch accents and the (potential) phrasal accent between the disjuncts. In addition, each disjunct was chosen to maximize sonority so that pitch might be tracked accurately.

Participants read a paragraph context and then were presented with the utterance, which they were asked to pronounce consistent with the context. For Y/N question contexts, the paragraph made it clear that the difference between the disjuncts is not salient (11a). Meanwhile,

the alternative question context insured that exactly one of the disjuncts could be chosen after asking the question (11b). The declarative context was also added as a baseline (11c).

(11) a. Y/N question context: “Your workplace has rented out a ballpark today. You do not go, although your coworker Marie does. Two of your coworkers, Isabel and Andrea, also really like baseball. You aren’t sure if they ended up going. You ask Marie, ‘Did Isabel or Andrea go to the baseball game today?’ ”

b. Alt question context: “You were going to a baseball game with your friend Marie, when you started feeling sick. You gave Marie your ticket and told her to invite one of her friends who like baseball, either Isabel or Andrea. Later that night, you want to know who went to the baseball game. You ask Marie, ‘Did Isabel or Andrea go to the baseball game today?’ ”

c. Declarative context: “Your friend Marie has an extra ticket to a baseball game today. Two of your friends want the ticket, and you told them to decide amongst themselves who would go. When another friend, Jason, asks you who else went to the game, you say ‘Isabel or Andrea went to the baseball game today.’ ”

The small study found that the L- phrasal accent was present in alternative questions (83.3%), and pitch accents were present on each disjunct (100%). The pitch accents varied on both disjuncts, although there was a trend for the first disjunct to end with a H- or L*+H (if a phrasal break was unclear; contours may be ambiguous between L*+H and L* H-, c.f. Beckman 1996). By comparison, the yes/no questions did not have a L- phrasal accent on the second disjunct

(97.2%) but were more often than not produced with a final rise (86.1%). Additionally, almost all yes/no questions⁵ contained only one phrase/phrasal boundary.

Heidenreich (2014b) used the production data from one participant to conduct a perception experiment in which a textual context was paired with a recorded utterance, and participants had to choose whether or not the utterance fit naturally with the story. Participants saw twelve examples of consistent context-utterance pairs (six alt-alt and six yes/no-yes/no) and twelve examples of inconsistent pairs (six alt-yes/no and six yes/no-alt). The results were analyzed according to if the participant correctly chose if the pair was consistent or inconsistent (Table 2.1).

	Alt. Sound	Y/N Sound
Alt Text	90.1%	56.7%
Y/N Text	68.4%	80.1%

Table 2.1: Percent correct of all items broken into utterance/context combinations

Participants were better at identifying consistent pairs vs. inconsistent pairs (they were more likely to say that the pair was consistent, vs. inconsistent, in general), which suggests that participants allow for accommodation when a tune does not match the preceding context. The alternative question productions were more uniform than the yes/no productions, which would explain why participants were better at identifying whether they fit naturally with the context (90.1%) or did not fit naturally with the context (68.4%). Additionally, when contexts were appropriate for alt. questions, yes/no questions that were further away from the canonical representations in (9) were more easily accommodated. Thus it seems easier to accommodate a

⁵ The only two exceptions to this claim were two productions that seemed to be produced as alternative questions (they had the previously stipulated ‘canonical’ alternative question tune)

yes/no utterance (which has a variety of contours) into an inconsistent context than to accommodate an alt. question (which has a specific contour) into an inconsistent context. In fact, no particular yes/no intonation type was responsible for the errors; rather it seemed that because yes/no contours *can be* so variable, participants had a harder time both fitting them into their appropriate contexts and choosing their inappropriate contexts.

The study also included one alternative question production that had a low phrasal accent, but a high phrasal boundary (12). While participants were perfect in scoring the production with its matching context as correct, they were significantly worse at correctly choosing the yes/no context as inconsistent (21.4%).

(12) Is Leah listening to philosophers or psychologists at the conference this weekend?

(L* H-) (H* L- H%)

While these results suggest that it is the L-, rather than a L%, that determines whether an utterance is an alternative question, they also suggest that a high phrasal boundary allows listeners to accommodate an alternative utterance into a yes/no context better than a canonical alternative tune. As the L-H% final contour is also associated with uncertainty (Pierrehumbert & Hirschberg 1990), it is possible that participants took the final contour as ‘more to come’ with the Y/N context and thus marked it (incorrectly) as consistent with the context.

Thus, the most recent empirical work has theorized that there are two important distinctions for an alternative question tune: the first disjunct must end with a high pitch (whether a H-, or even a L*+H if a phrasal accent is missing), and the second disjunct must end with a low pitch (L-). Alternative question productions may have one or two phrasal accents, as long as these conditions are met; this is contrary to current canonical representations of alternative questions, which stipulate a mandatory phrasal accent between disjuncts. Yes/no

questions can have nearly any pitch accent/phrasal accent combination, as long as it does not fall into the alternative question categorization described; current canonical representations do not allow a phrasal accent between disjuncts, nor pitch accents on non-nuclear syllables. However, these conclusions were drawn from productions of only three participants; a larger study must be conducted in order to determine whether the assumptions of canonical representations are inaccurate. The production experiment in Chapter 3 was conducted with this aim in mind.

2.3 Disjunctive Question Grammaticalization

In English, a disjunctive question is ambiguous in text form (Aloni & van Rooy 2002). As discussed above, intonation may disambiguate between the two interpretations. However, many have argued that a Y/N question allows the lexical item *either* before the first disjunct in a disjunctive question, while an Alt. question does not (Huddleston 1994, Haspelmath 2007). Huddleston (1994) states “*Is it either a boy or a girl?* is unambiguously a polar question (420)”. Haspelmath (2007) concurs, stating “*Do you want either tea or coffee?* cannot be an alternative question (26).” That is, (13a) is ambiguous without intonational or additional textual context, while (13b) can only have the Y/N interpretation (‘Is Mary allergic to any of these things: dairy or soy?’).

(13a) Is Mary allergic to dairy or soy?

(13b) Is Mary allergic to either dairy or soy?

There are two assumptions involved in the text form of disjunctive questions; the first being that a disjunctive question is ambiguous in text form (not biased in any particular interpretational direction); the second assumption is that *either* sways the interpretation exclusively to the Y/N interpretation. The text and perception experiments (in Chapters 4 and 5, respectively) were conducted in order to test these assumptions. The text experiment is text only (no audio) and

compares licit answers to questions like (13a) to those from questions like (13b). The perception experiment is audio only, comparing licit answers to questions with *either* to questions without *either*, but with the added cross factor of intonational tune (an Alt. question tune vs. a Y/N question tune).

2.4. Disjunctive Question Responses

The acceptability of responses to disjunctive questions has been discussed in many papers (Aloni & van Rooy 2002, Biezma & Rawlins 2012, among others), but this acceptability has been implicitly accepted by the writers, rather than being tested within a group of native speakers. The consensus is that an Alt. question will accept one of the disjuncts (but not both) as an answer, while disallowing *neither*, or any combination of disjuncts with *Yes* or *No*. Y/N questions, however, will allow polarity particles (*Yes*, *No*) as well as a wide range of alternatives (*Neither*, *Both*, one of the disjuncts, etc.). There is an assumption that any licit Alt. question answer would also be licit for its Y/N question counterpart.

While a disjunct answer (the acceptable answer for an Alt. question) would appear to be licit in all instances of its Y/N question counterpart, a cleft disjunct answer poses a potential problem. Historically, clefts have been described as carrying an exhaustivity implication, although whether this arises via pragmatics (Horn 1981, 2004) or semantics (Percus 1997, Velleman et. al. 2012) has been debated. Pollard and Yasavul (2015) argued that clefts specify an antecedent discourse reference, and “the appearance of an exhaustivity implication only arises when such clefts are used to answer *wh*-questions” (1). This theory has been supported by empirical work by DeVeugh-Geiss et. al. (2018) on clefts, which found that participants obtained an exhaustive reading only if they took clefts to anaphorically refer to an implicit question. Therefore, if the interrogative in (14) is given (without prosody at the moment; let us

assume it is in text form and both interpretations are available), a cleft disjunct answer (14a) may take an exhaustive reading. This interpretation, against typical disjunctive question response assumptions (Aloni & van Rooy 2002, Biezma & Rawlins 2012), would suggest that a cleft disjunct answer only be appropriate for the Alt. question interpretation, rather than a Y/N question interpretation.

Similarly, a disjunct answer with a continuation rise (14b, called a rising disjunct answer in the perception experiment, Chapter 5) communicates more than a disjunct answer with the typical falling (assertive) prosody (14c).

(14) *Is Mary allergic to dairy or soy?*

- a. Cleft disjunct answer: *It's dairy.*
- b. Rising disjunct answer: *She's allergic to dairy*↑.
- c. Falling disjunct answer: *She's allergic to dairy*↓.

A continuation rise (L-H%) has been known to signal a ‘forward reference’ (Pierrehumbert & Hirschberg 1990), interpreted as ‘this utterance will be completed by a subsequent utterance’. This type of response, while still a ‘disjunct’ answer, is not compatible with an Alt. question (as the Alt. question itself demands exactly one disjunct, and the response contains exactly one disjunct, indicating subsequent information is not only unnecessary but also potentially illicit). Therefore, the assumption that any ‘disjunct’ response is acceptable for an Alt. question may prove to be inaccurate.

The text and perception experiments explore the generally licit answers to both types of questions, for which empirical work has generally been lacking (even the DeVeugh-Geiss et. al. 2018 experiment was not based on responses to questions, but rather determining whether the image revealed enough information to support the cleft utterance). Current semantic theories on

interrogatives and their answers offer predictions for licit answers for both types of disjunctive questions.

2.5 Theoretical Work

In comparison to empirical work, semantic theories offer slightly different predictions on intonational contours, as well as offering predictions for licit answerhood conditions. Two distinct theories, Inquisitive Semantics and Commitment Space Discourse, are examined in order to test their predictions against the results of the experiments. Both these theories provide a formal framework to model disjunctive questions *and* their licit responses. Inquisitive Semantics was developed first (Ciaredelli et. al. 2013a), as a new dynamic semantics framework that aimed to handle all types of questions. Commitment Space Semantics was developed as a response to Inquisitive Semantics (Krifka 2015), attempting a more elegant solution to handle bias in questions, while also (perhaps inadvertently) predicting slightly different answerhood conditions for the two disjunctive question interpretations. As discussed in the introduction, any framework that attempts to differentiate between the two types of disjunctive questions must incorporate both aspects of question and answer, and also correctly model empirical data.

2.5.1 Inquisitive Semantics

Inquisitive semantics (IS) is a framework of meaning that captures both informative and inquisitive content (Ciaredelli et al 2013a, Ciardelli 2015, Roelofsen 2013a, Roelofsen 2013b) in an integrated way. It uses classical truth-conditional notions of meaning to be both speaker- and hearer-oriented, formulating what it means to behave rationally by exchanging information rather than just providing information. In order to arrive at the predictions for licit answers in Inquisitive Semantics, a brief overview of Inquisitive Semantics is given.

The framework is built on the discourse context, or *common ground*, that has been established in the discourse so far. The common ground is modeled as a set of possible worlds (the worlds that are compatible with the established information). In Inquisitive Semantics, the meaning of a sentence is defined as a proposition (or *discourse context*), which is modeled as nonempty downward-closed sets of states (Roelofsen 2013b) or possibilities (Roelofsen and van Gool 2010). Ciardelli et al. (2013a) explain that the intended effect of uttering a sentence in a certain discourse context is to restrict that discourse context to precisely those worlds that satisfy the truth-conditions of the uttered sentence, i.e. to those worlds in which the sentence is true. So in uttering a sentence φ , a speaker provides information that the actual world lies in $U[\varphi]$, and also steers the common ground toward a specific state in $[\varphi]$ (Roelofsen 2013b). Thus, complete sentences express propositions, which are sets of states, or possibilities.

To model disjunctive questions, we will focus on a very specific type of discourse where participants exchange information by raising and resolving issues. Therefore, the contexts must not only contain the information established, but also the issues raised. The meaning of a sentence needs to embody both its informative content (potential to provide information) as well as its inquisitive content (potential to raise issues). Thus the meaning of a sentence should embody its *information exchange potential*.

Inquisitive Semantics formulates a notion of discourse contexts by defining information states (s) as a set of possible worlds, i.e. $s \subseteq \omega$ (where ω denotes the set of all possible worlds). For any discourse context c , info_c denotes the information state that represents that information available in c . In discourse context c , the actual world is located somewhere in info_c , i.e. every subset $s \subseteq \text{info}_c$ is at least as informed as info_c and thus an enhancement of info_c (an enhancement of any state is defined as a subset of that state). An issue in c can be modeled,

therefore, as a non-empty set \mathcal{J} of enhancements of info_c . Formal definitions of this model are given in Definitions 1-4 (for more information on supporting the formulation of this model, refer to Ciaredelli et. al. 2013a,b).

Definition 1 (Issues)

Let s be an information state, t an enhancement of s , and \mathcal{J} a non-empty set of enhancements of s .

Then we say that \mathcal{J} is an issue over s if and only if:

1. \mathcal{J} is downward closed: if $t \in \mathcal{J}$ and $t' \subset t$ then also $t' \in \mathcal{J}$
2. \mathcal{J} forms a cover⁶ of s : $\bigcup \mathcal{J} = s$

Definition 2 (Settling an Issue)

Let s be an information state, t an enhancement of s , and \mathcal{J} an issue over s . Then we say that t settles \mathcal{J} if and only if $t \in \mathcal{J}$

Definition 3 (Discourse contexts)

1. A discourse context c is a non-empty, downward closed set of states
2. The set of all discourse contexts will be denoted by C

Definition 4 (The information available in a discourse context)

For any discourse context c : $\text{info}_c := \bigcup c$

⁶ A *cover* implies that every world in info_c be must included in at least one element of \mathcal{J} ; this is in order to ensure that it is even possible to satisfy the request of \mathcal{J} without discarding the actual world (Ciaredelli et. al. 2013a).

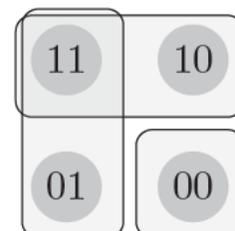
Just like in the classical case, a meaning f can be identified with a unique static object, $f(c\tau)$, which is called the *proposition* associated with f . Propositions are non-empty, downward closed sets of states. Inquisitive Semantics claims there is a one-to-one correspondence between propositions and meanings. When adding a proposition to the current discourse context c , the proposition merges with c (where merge is defined as set intersection) to form a new discourse context c' . The requirement of downward closure for propositions ensures that if a given proposition is settled by a state s , then it will also be settled by any more informed state s' that is a subset of s (Roelofsen 2013b, Ciardelli et. al. 2013a).

The IS framework employs visualizations in order to illustrate their definition of propositions and worlds. Because there could potentially be many states that settle a proposition $[\varphi]$, only the maximal elements of $[\varphi]$ are included in the figures associated with IS (Ciardelli et. al. 2013a, Roelofsen and van Gool 2010, Roelofsen 2013b). These maximal elements of $[\varphi]$ are the states that contain the least information while still settling $[\varphi]$ (defined as being the easiest to reach). The maximal elements are known as alternatives. The figures shown henceforth will only depict those alternatives, or maximal states.

Thus, an open disjunctive interrogative (how IS classifies a Y/N disjunctive question with a phrasal break), such as (15), can be visualized with Figure 2.1 (adapted from Ciardelli et. al. 2013a). In this depiction, four possible worlds exist: one world where both Peter and Maria will attend the meeting, one where only Peter will attend, one where only Maria will attend, and one where neither will attend, visualized by circles labeled as 11, 10, 01, and 00, respectively.

(15) Will Peter[↑] attend the meeting, or Maria[↑]?

Figure 2.1(right): Maximal sets for (15)



The three alternatives (maximal states, shown by boxes enclosing the labeled circles) for a Y/N disjunctive question are the set of all worlds where Peter will attend the meeting, the set of all worlds where Maria will attend the meeting, and the set of all worlds where neither Peter nor Maria will attend. That is, these are the least informative alternatives that settle the proposition expressed by (15); more informed states (e.g. *Maria will attend but Peter will not*) presumably could also settle the proposition due to the downward closure requirement of propositions. *Less* informed maximal states (i.e. *Yes*, which would encompass states 11, 10, and 01) are not permitted. Thus the framework seemingly explains the licit set of responses for one type of Y/N disjunctive question (henceforth called an open disjunctive interrogative, as it is named in this framework). This set of licit responses will be tested against the empirical data from the experiments.

The framework proposed does run into a snag: while an open disjunctive can be explained by this framework, a closed disjunctive interrogative (the term used to define an alternative question), cannot. Alternative questions have a non-at-issue implication that exactly one of the disjuncts is supposed to hold (Karttunen and Peters, 1976; Hand and Romero, 2004a; Beck and Kim, 2006, among others). This non-at-issue implication can, however, be accounted for by extending the framework to include focus features and prosody, that can *suggest* or *highlight* states, or possibilities (Roelofsen and van Gool, 2010). This extension uses machinery from alternative semantics (Hamblin, 1973; Kratzer and Shimoyama, 2002; Alonso-Ovalle, 2006). Each expression of a certain type (e , s , t , σ and τ) is mapped to a model-theoretic object, which belongs to a certain domain. There is a domain D_e of individuals, a domain D_s of indices, and a domain D_t consisting of truth values 0 and 1. For every complex type ($\sigma\tau$) there is a domain $D_{(\sigma\tau)}$ consisting of all functions from D_σ to D_τ . The semantic value of an expression α

of type τ is denoted by $[[\alpha]]$, which is always a set of objects in $D\tau$. Sentences are expressions of type (st).

This extension resolves an old puzzle concerning polar questions, exemplified by (16a) and (16b):

(16) a. Is the door open?

b. Is the door closed?

According to IS, (16a) and (16b) are equivalent: they both express a proposition consisting of two possibilities, the possibility that the door is open, and the possibility that the door is closed. The difference lies in the replies to these questions: *yes* to (16a) means the door is open, while *yes* to (16b) means that the door is closed. This difference is captured if we assume (16a) *highlights* the possibility that the door is open, (16b) *highlights* the possibility that the door is closed, and the interpretation of *yes* confirms the highlighted possibility, while *no*, if felicitous, simply rejects the possibilities highlighted.

Therefore, the semantic value of a sentence α , denoted by $[[\alpha]]$ consists of two components, $[[\alpha]]_P$ and $[[\alpha]]_H$. Both $[[\alpha]]_P$ and $[[\alpha]]_H$ are sets of possibilities; $[[\alpha]]_P$ (P-set) is the proposition that α expresses, and $[[\alpha]]_H$ (H-set) consists of the possibilities that α highlights. All semantic values, including P-sets and H-sets, are composed by means of a pointwise function application:

(17) Pointwise Function Application

If $[[\alpha]] \subseteq D(\sigma\tau)$ and $[[\beta]] \subseteq D\sigma$ then

$$[[\alpha\beta]] := [[\beta\alpha]] := \{ d \in D\tau \mid \exists a \in [[\alpha]]. \exists b \in [[\beta]], d = a(b) \}$$

Lexical items are mapped to singleton sets (18), and disjunction introduces alternatives (19).

(18) $[[\text{Ann}]] := \{\text{Ann}\}$, etc.

(19) For any type τ , if $[[\alpha]], [[\beta]] \subseteq D\tau$, then $[[\alpha \text{ or } \beta]] := [[\alpha]] \cup [[\beta]]$

In addition, the interrogative complementizer, Q , always operates on an expression α of type (st), and the resulting clause $[Q \alpha]$ is always again of type (st), which shifts the syntactic category but not semantic type (Roelofsen and van Gool, 2010). This proposal expressed by $[Q \alpha]$ consists of the possibilities of α itself *plus* the possibility that α excludes:

(20) $[[Q \alpha]] := [[\alpha]] \cup [\downarrow\alpha]$

If we assume that $[Q \alpha]$ highlights that possibilities that α itself highlights, *not* the possibility that α excludes, then we are able to capture the contrast between these two polar questions (21a, b).

(21)a $[Q\text{-is the door open}]$

Proposes: open/closed

Highlights: open

yes \Rightarrow the door is open

no \Rightarrow the door is closed

(21)b $[Q\text{-is the door closed}]$

Proposes: open/closed

Highlights: closed

yes \Rightarrow the door is closed

no \Rightarrow the door is open

By separating proposition meanings into what is expressed and what is highlighted, the framework can now begin to account for the following three types of disjunctive questions: (The

c in the logical form of the closed intonation pattern represents a closure feature, which is linked with the rising-and-falling pitch contour of the utterance. This will be discussed momentarily.)

These answers are judged as licit (or illicit) by Roelofsen and van Gool; thus their predictions will be weighed against the empirical data from the experiments. Note that all the patterns listed below are ‘narrow-scope’ in form, where ‘narrow-scope’ here indicates that the disjunction is at the NP level, rather than the S level.

(22) Narrow scope block intonation pattern

Logical form: [Q-does [Ann or Bill]_F play]

Acoustic signal: Does Ann-or-Bill[↑] play the piano?

- Responses:
- a. No.
 - b. Yes.
 - c. (Yes,) Ann does.
 - d. (Yes,) Bill does.

(23) Narrow scope⁷ open intonation pattern

Logical form: [Q-does [Ann]_F or [Bill]_F play]

Acoustic signal: Does Ann[↑] or Bill[↑] play the piano?

- Responses:
- a. No.
 - b. #Yes.
 - c. Ann does.
 - d. Bill does.

⁷ Wide-scope open and closed intonation also occurs; while having different logical forms, they behave like their narrow-scope counterparts in terms of the acoustic features and the effects of intonation on answerhood conditions

(24) Narrow scope closed intonation pattern

Logical form: [Q-does [Ann]_F or [Bill]_F play]_C

Acoustic signal: Does Ann[↑] or Bill[↓] play the piano?

- Responses:
- a. #No.
 - b. #Yes.
 - c. Ann does.
 - d. Bill does.

Adding in the effect of highlighting now accounts for why the block intonation pattern licenses a *yes* response, while an open intonation pattern and a closed intonation pattern do not:

(25) Does [Ann or Bill]_F play?

- a. Highlights the possibility that Ann or Bill plays.
- b. *yes* ⇒ at least one of them plays
- c. *no* ⇒ neither Ann nor Bill plays

(26) Does [Ann]_F or [Bill]_F play?

- a. Highlights the possibility that Ann plays and the possibility that Bill plays
- b. *yes* ⇒ presupposition failure (the question highlights more than one possibility)
- c. *no* ⇒ neither Ann nor Bill plays

To understand why the closed intonation pattern does not license a *no* response as well, we need to return to the idea of closure. Closure (also called presuppositional closure, c.f. Roelofsen 2013a) is the IS interpretation of how to account for the ‘exactly one’ stipulations⁸ in

⁸ Inquisitive Semantics refers to this ‘exactly one’ stipulation as a presupposition (Roelofsen and van Gool 2010)

alternative questions; that if the question is “Does Ann or Bill play the piano?”, at least one of them plays and they do not both play. A third component is added to the meaning of sentence: $[[\alpha]]_s$ (S-set), which is the set of possibilities/updates that α suggests. S-sets that do not bear a closure-feature are empty; the semantic contribution is defined below:

(27) The effect of closure:

$$[[\alpha c]]_P := [[\alpha]]_P$$

$$[[\alpha c]]_H := [[\alpha]]_H$$

$$[[\alpha c]]_s := \mathcal{EX}([[\alpha]]_H)$$

The exclusive strengthening operator \mathcal{EX} is further defined for any set of possibilities Π and for any possibility $\pi \in \Pi$, the exclusive strengthening of π relative to Π is defined as:

$$(28) \mathcal{EX}(\pi, \Pi) := \pi - \cup\{\rho \mid \rho \in \Pi \text{ and } \pi \not\subseteq \rho\}, \text{ where } \rho \text{ is a possibility}$$

$$(29) \mathcal{EX}(\Pi) := \{ \mathcal{EX}(\pi, \Pi) \mid \pi \in \Pi \}$$

Figure 2.2 (adapted from Roelofsen and Van Gool, 2010) illustrates the results of exclusive strengthening.

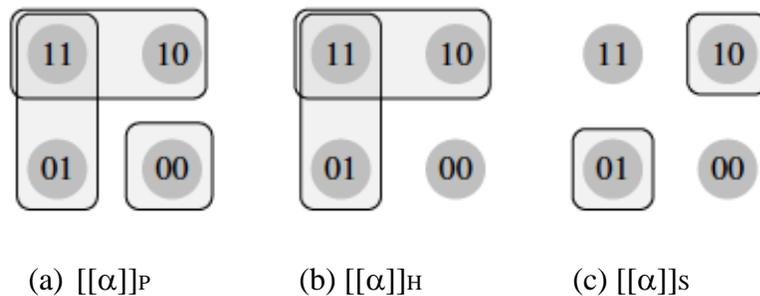


Figure 2.2: Applying Exclusive Strengthening

Thus, closure has the following effect on the narrow-scope⁹ closed intonation pattern:

- (30) Does Ann[↑] or Bill[↓] play the piano?
- a. $[[\alpha]]_P$ proposes three possibilities: Ann plays, Bill plays, or neither
 - b. $[[\alpha]]_H$ highlights two possibilities: Ann plays or Bill plays
 - c. $[[\alpha]]_s$ removes the overlap between the highlighting suggestions, resulting in the two distinct possibilities in Figure 2.2(c). These are the correct responses, as licensed in (24).

In order to negate a closed intonation pattern, *no* is licensed in the declarative disjunctive but not the interrogative disjunctive:

- (31) Ann[↑] or Bill[↓] plays the piano.
- a. No, neither of them does.
- (32) Does Ann[↑] or Bill[↓] play the piano?
- b. #No, neither of them does.
 - c. Actually, neither of them does.

The difference here is that the declarative truly asserts that at least one of Ann and Bill plays, while the interrogative *suggests* this (as represented by highlighting). So in the IS framework, *no* can be used to deny an assertion, but not to cancel a suggestion. Rather, a weaker disagreement

⁹ Wide-scope closed intonation pattern has the same result, as it highlights the same possibilities. Closure here is not interpreted as exhaustivity (Roelofsen and van Gool 2010), as in Zimmerman 2000.

particle (*actually, in fact*) is required to cancel a suggestion. This accounts for the licensing and interpretation of *no* in response to disjunctive interrogatives with block or open intonation; it also accounts for the contrast between (31) and (32). While *no* denies the possibilities that (32) highlights, it is not felicitous because it also cancels the suggestion that (32) expresses (namely, the ‘exactly one’ suggestion).

The correct licensing of *no* is imperative, as a final operator, the presuppositional closure (pres-closure) operator, is applied to disjunctive questions. This operator implies that at least one of the given alternatives for a disjunctive question holds; in block and open intonational patterns, neither disjunct (*no*) is a possible alternative for the sentence.

Summary

The IS framework takes into account three factors when dealing with disjunction: phrasing, final pitch contour, and word order. A block intonation pattern has a single phrase (or list item), while open and closed intonation patterns have a phrase break between disjuncts (two list items). The logical form reflects this difference of focus; open and closed intonation patterns will have focus on both disjuncts, while the block intonation pattern will have focus over the whole disjunctive phrase (and not each disjunct individually). The final pitch contour (rising vs. falling) can indicate the possibility that none of the items holds (*no* is licensed) or that exactly one of the items is supposed to hold (*closure* feature). The word order determines whether it is a declarative or interrogative list. The IS framework operates on the assumption that in an interrogative list, one of the given alternatives holds (in block and open intonation patterns, *no* is a licensed alternative). The given interpretation procedure for identifying the type of disjunctive question (Roelofsen 2013b) is given below:

1. Determine basic list items
 - a. Detect the prosody phrase boundaries
2. Determine whether the list is open (or block) or closed
 - a. If the list is open or block, *no* is a licensed alternative (list completion)
 - b. If the list is closed, apply exclusive strengthening
3. Determine whether the list is declarative or interrogative
 - a. (Because we are only dealing with interrogatives), apply pres-closure

Thus, IS defines three distinct types of interrogative disjunction questions, with their possible answers defined graphically below (figures taken from Roelofsen 2013b).

(33) Does Ann[↑] or Bill[↓] play the piano? (Figure 2.3)

- Responses:
- a. #No.
 - b. #Yes.
 - c. Ann does.
 - d. Bill does.

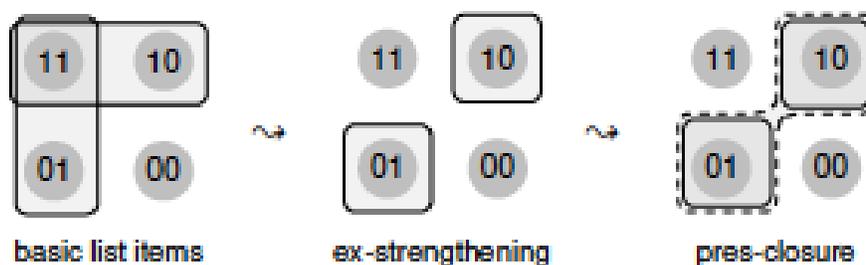


Figure 2.3: Closed intonational pattern (two lists)

(34) Does Ann[↑] or Bill[↑] play the piano? (Figure 2.4)

- Responses:
- a. No.
 - b. #Yes.
 - c. Ann does.
 - d. Bill does.

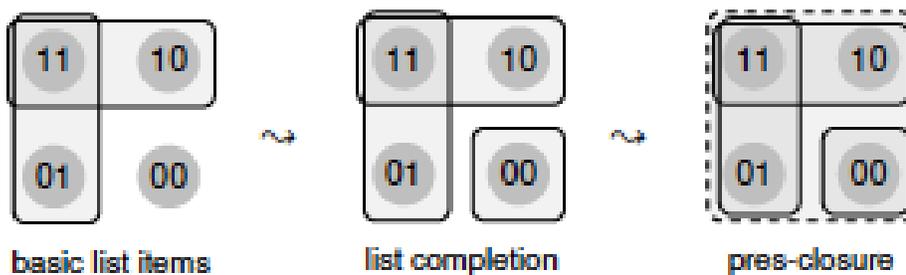


Figure 2.4: Open intonational pattern (two lists)

(35) Does Ann-or-Bill[↑] play the piano? (Figure 2.5)

- Responses:
- a. No.
 - b. Yes
 - c. (Yes,) Ann does.
 - d. (Yes,) Bill does.

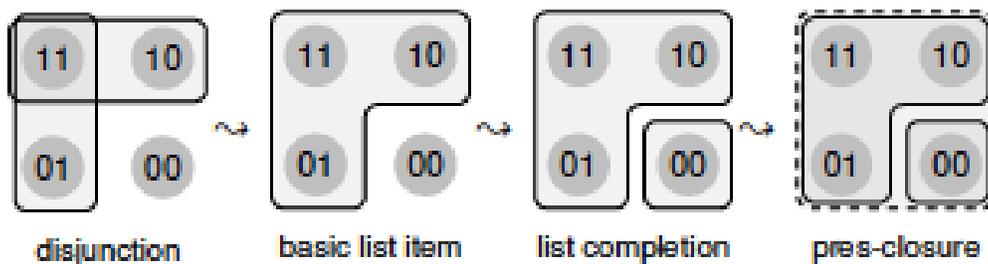


Figure 2.5: Block intonational pattern (one list)

One final note is the reminder that the figures indicate maximal states; an answer may be more specific within a defined maximal state (e.g. *Yes, Bill does* occurs within the maximal state *Yes* in Figure 5; hence why response (35d) is licit); therefore, we might assume that if (35) licenses both (c) and (d), then it ought to also license a response (e) *Yes, they both do*. Similarly, the open intonational pattern with two lists ought to license the more specific state *They both do*, though a *Yes* before this response is not allowed (just as a *Yes* is not allowed before either of the disjunct answers, *Ann does* and *Bill does*, c.f. (34) and its responses). Table 2.2 gives a full breakdown of each type of disjunctive question that Inquisitive Semantics accounts for, as well as the licensed answers.

Disjunctive Question	Number of Lists	Final Contour	Intonation Pattern Type	Licensed Answers				
				Yes	No	Ann does	Bill does	Both
Does Ann [↑] or Bill [↓] play the piano?	Two	Falling	Closed			✓	✓	
Does Ann [↑] or Bill [↑] play the piano?	Two	Rising	Open (Open complex)		✓	✓	✓	✓
Does Ann-or-Bill [↑] play the piano?	One	Rising	Block (Open simple)	✓	✓	✓*	✓*	✓*

Table 2.2: Disjunctive Questions and licit answers according to Interrogative Semantics

An * indicates an optional Yes could occur before these licensed answers.

2.5.2 Commitment Space Semantics

Commitment Space Semantics, or Commitment Space Discourse, is a formal framework that captures both information agreed on by the interlocutors as well as projected continuations of the exchange (Krifka 2015). Specifically, it allows for modeling interrogatives by having questions restrict the viable continuations by the other participants. The framework uses speech acts, which are functions from commitment states to commitment states (Krifka 2015, Szabolcsi 1982, Cohen & Krifka 2014). The framework has been proposed as an alternative to Inquisitive Semantics because of how it easily accounts for bias in questions (Krifka 2015), without resorting to additional operators such as ‘highlighting’ or ‘suggesting’. However, the theory needs to correctly model and predict disjunctive questions and their answers if it is to be taken as a superior framework.

A commitment state in this framework is a set of propositions that are publicly shared and agreed upon by the participants. The basic function of the speech act is to alter this commitment state. The formal object that includes the commitment state and the expected or legal continuations in discourse is the commitment space (C), as modeled by a set of commitment states.

(36) C is a commitment space if C is a set of commitment states, $\cap C \neq \emptyset$ and $\cap C \in C$

$\cap C$ is called the root of C (written as \sqrt{C}), the set of all propositions that participants have positively committed to up to the current point in conversation. Suppose speaker 1 (S_1) utters an assertion φ ; the commitment state that adds the proposition φ to the commitment state \sqrt{C} is defined as the union between \sqrt{C} and φ (Figure 2.6, modeled after Figure 1 in Krifka 2015).

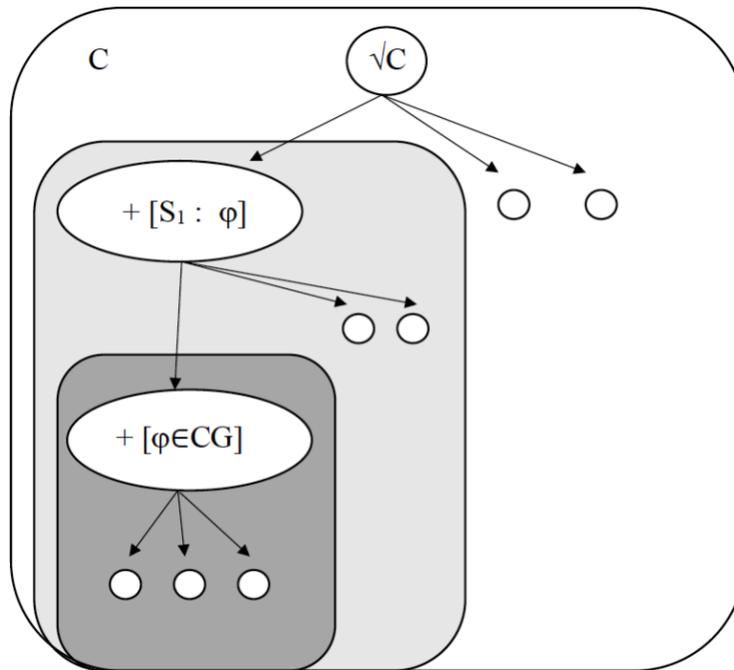


Figure 2.6: Update of commitment space

The first commitment state is \sqrt{C} ; an intermediate commitment state indicates the point at which S_1 has uttered φ and thus is liable for the commitment of the proposition. That is, when a speaker asserts a proposition, they undertake responsibility for what is claimed by committing themselves to the truth of the proposition. In the remaining figures, speaker S committed to the truth of proposition φ will be expressed as ‘ $S \vdash \varphi$ ’, à la Krifka 2015, Krifka 2016. The final commitment state is when φ has been added to the common ground (CG), which translates to φ now included in the new \sqrt{C} . Pragmatically licit updates must not be entailed by \sqrt{C} and, ideally, must be consistent with the propositions in \sqrt{C} (a proposition and its negation cannot be part of the same commitment state, but sometimes people have inconsistent beliefs and make inconsistent commitments).

Certain conversational moves amount to a rejection; therefore, a record of the moves in the conversation is needed. This is modeled as a sequence of commitment states $\langle C_0, C_1, \dots, C_n \rangle$, which is the commitment space development (CSD). The framework lays out how to update the CSD (37), indicate the actor of a speech act (38), and reject a speech act (39).

(37) Update of a commitment space development with a speech act \mathfrak{A} :

$$\langle \dots, C \rangle + \mathfrak{A} = \langle \dots, C, C + \mathfrak{A} \rangle$$

(38) Update of a commitment space development with speech act \mathfrak{A} by actor S:

$$\langle \dots, C_s \rangle +_s \mathfrak{A} = \langle \dots, C_s, [C + \mathfrak{A}]_s \rangle$$

(39) Rejection of last speech act, as expressed by operator \mathfrak{R}

$$\langle \dots, C^*, C'^* \rangle +_s \mathfrak{R} = \langle \dots, C^*, C'^*, C_s \rangle$$

The + in the previous three functions is shorthand for functional application, e.g. (40):

- (40) a. $c + \mathfrak{A}\varphi = \mathfrak{A}\varphi(c)$, where $\mathfrak{A}\varphi = \lambda c[c \cup \varphi]$
 b. $C + \mathfrak{A} = \mathfrak{A}(C)$, where $\mathfrak{A} = \lambda C\{c \in C \mid \sqrt{C + \mathfrak{A}} \subseteq c\}$
 c. $\langle \dots, C^* \rangle +_s \mathfrak{A} = \mathfrak{A}_s(\langle \dots, C^* \rangle)$, where $\mathfrak{A}_s = \lambda \langle \dots, C^* \rangle \langle \dots, C, [\mathfrak{A}(C)]_s \rangle$
 d. $\langle \dots, C^*, C'^* \rangle +_s \mathfrak{R} = \mathfrak{R}_s(\langle \dots, [C] \dots \rangle)$, where $\mathfrak{R}_s = \lambda \langle \dots, C'^*, C^* \rangle, \langle \dots, C^*, C'^*, C_s \rangle$

In addition, a speech act phrase, or Illocutionary Act phrase (ActP) will distinguish between assertions and questions, and it represents the final prosody of the phrase. For assertions, this means the head of the ActP will be rendered by “.”, which is expressed by the

falling prosody of the intonational phrase that corresponds to the speech act (Truckenbrodt 2015, Krifka 2015). An ActP head of ‘?’ is triggered by rising prosody of the intonational phrase that corresponds to the speech act. This simple analysis, so far, does not explain how Alt. questions have the same final falling prosody as assertions (which will be addressed momentarily). First, let us consider a simple, non-disjunctive, polar question.

In Commitment Space Discourse, a question indicates that a speaker is requesting an assertion of a particular type from the other speaker (Krifka 2015). This type of speech act is considered a meta speech act, in that the question does not change the root of the commitment space (c.f. Figure 2.6, in which an assertion does change the root of the commitment space), but restricts the possible continuations of the commitment space “to those in which the other speaker makes an assertion of an appropriate type” (Krifka 2015, 335). Commitment Space Discourse uses the classical analysis of polar questions, i.e. a choice between the proposition and its negation (Hamblin 1973). For example, if speaker one (S₁) were to say to speaker two (S₂) “Did I win the race”, the current framework would express it as (41), seen visually in Figure 2.7 (adapted from Figure 8 in Krifka 2015):

(41) Polar question analyzed as a choice between two alternatives

“Did I win the race?”

$\langle \dots, C^* \rangle + S_1$ to S_2 : *Did I win the race?*

$= \langle \dots, C^*, [\{\sqrt{C}\} \cup C + S_2 \vdash \varphi \cup C + S_2 \vdash \neg\varphi]_{S_1} \rangle$

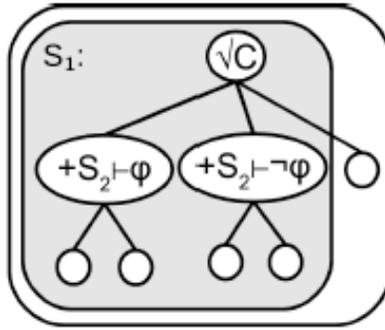


Figure 2.7: Polar Question in CSD

The future developments of C are restricted by S_1 in such a way that the only legal continuations are the commitments by S_2 that S_1 won the race (42), that S_1 did not win the race (43), or “moves that are entailed by these commitments (Krifka 2015, 335)”.

$$(42) (41) + S_2: \text{Yes.} = (34) + S_2 \vdash \varphi$$

$$(43) (41) + S_2: \text{No.} = (34) + S_2 \vdash \neg\varphi$$

When a question is analyzed via the standard treatment of polar questions (as presenting an option between two alternatives), the question is termed bipolar in this framework. However, Krifka (2015, 2016) argues that this type of question may also be biased toward one answer, and thus a monopolar interpretation is also available. In a monopolar question, a speaker proposes just one legal continuation to the addressee. So a question like *Did I win the race?* has both bipolar and monopolar interpretations available. (44) reveals the monopolar interpretation of *Did I win the race?* Notice that in a monopolar interpretation, the *yes* answer becomes straightforward, whereas the *no* answer requires a prior rejection (45).

$$(44) \langle \dots, C^* \rangle + S_1, \text{ to } S_2: \text{Did I win the race?}$$

$$= \langle \dots, C^*, [\{\sqrt{C}\} \cup C + S_2 \vdash \varphi]_{S_1} \rangle$$

(45) Responses to *Did I win the race?* (44)

a. S_2 : *Yes.* = (35) + $s_2 S_2 \vdash \varphi$

b. S_2 : *No.* = (35) + $s_2 \mathfrak{R}$ + $s_2 S_2$: *No.*

$$= \langle \dots, C^*, [\{\sqrt{C}\} \cup C + S_2 \vdash \varphi]_{S_1}, C^*, [C + S_2 \vdash \neg\varphi]_{S_2} \rangle$$

This type of interpretation is supported by the fact that a question like *Did I win the race, or not?* can also be uttered. Krifka (2015) argues that the very existence of questions like (46) is evidence that questions like (41) actually have the monopolar reading (44); otherwise such disjunctions would be redundant. The bipolar reading is only available via special circumstances (Krifka (2016) gives the example of an accented auxiliary as moving the interpretation from monopolar to bipolar).

(46) S_1 to S_2 : *Did I win the race, or not?*

$$= [[[\text{ActP } \textit{Did I win the race}]]]_{S_1 S_2} \vee [[[\text{ActP } \textit{Did I not win the race}]]]_{S_1 S_2}$$

$$= \lambda C^*[\{\sqrt{C}\} \cup C + S_2 \vdash \textit{'S}_1 \textit{ won the race'}] \cup [\{\sqrt{C}\} \cup C + S_2 \vdash \neg \textit{'S}_1 \textit{ won the race'}]_{S_1}$$

$$= \lambda C^*[\{\sqrt{C}\} \cup C + S_2 \vdash \textit{'S}_1 \textit{ won the race'} \cup C + S_2 \vdash \neg \textit{'S}_1 \textit{ won the race'}]_{S_1}$$

Krifka argues that a disjunctive question is modeled as a disjunction of two monopolar questions, rather than a bipolar interpretation. The meanings are constructed by an Act phrase head $[\text{Act}_0 ?]$ that requests the commitment denoted by the complement of the ActP to be performed by the addressee, rather than the speaker. A simple answer yes/no is disallowed as there are two propositional discourse referents:

(47) $[\text{ActP } [\text{ActP } ? \textit{Did} [\text{CmP } \vdash [\text{IP } \textit{I win the race}]]]]]$ or $[\text{ActP } ? \textit{did} [\text{CmP } \vdash [\text{IP } \textit{I not win the race}]]]]$

$\text{IP} \rightarrow \varphi$ $\text{IP} \rightarrow \neg\varphi$

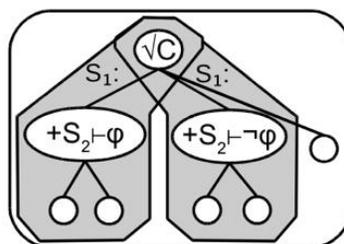


Figure 2.8: Alternative Question Disjunction in CSD

This is how alternative questions are modeled in this framework (see Figure 2.8). The gray shaded areas combining \sqrt{C} and a potential continuation are meant to convey that the mother (\sqrt{C}) is a proper subset of the daughter (potential continuation). Thus the model predicts that licit answers are either the first or the second discourse referent, and any other answer would be a rejection of the proposed update to the commitment space. However, this question uses ‘?’ as the head of ActP, which is meant to represent that final prosody of the phrase; thus it appears that the framework either predicts alternative questions to end with rising prosody (‘?’ head of ActP) or the model fails to use ‘.’ in the ActP position for alternative questions. Possibly the framework is meant to *allow* for both rising and falling intonation differences, and the head of ActP is assigned based on the intonation observed.

In fact, if we look at a disjunctive Y/N question, it is *also* a disjunction of two monopolar questions, such as (48). In this example, the two monopolar questions are unrelated, but the visualization suggests that this type of question is handled the same way as the alternative question (Figure 2.9, taken from Krifka 2016).

(48) Did Ed meet Ann, or did Ed meet Beth? (rising accent on Ann and Beth)

[[[ActP [ActP *Did Ed meet Ann*] or [ActP *Did Ed meet Beth*]]]S1S2

with [[[ActP *Did Ed meet Ann*]]]S1S2 = $\lambda C^*[\{\sqrt{C}\} \cup C + S_2 \vdash \text{Ed met Ann}']_{S_1}$

and $[[[\text{ActP } \textit{Did Ed meet Beth}]]]_{S_1 S_2} = \lambda C^*[\{\sqrt{C}\} \cup C + S_2 \vdash' \textit{Ed met Beth}']_{S_1}$

and $[[\text{or}]]_{S_1 S_2} = \lambda A \lambda A' \lambda C^*[A(C) \cup A'(C)]_{S_1}$, where A, A' : variables over speech acts

$= \lambda C^*[[\{\sqrt{C}\} \cup C + S_2 \vdash' \textit{Ed met Ann}'] \cup [\{\sqrt{C}\} \cup C + S_2 \vdash' \textit{Ed met Beth}']]_{S_1}$

$= \lambda C^*[\{\sqrt{C}\} \cup C + S_2 \vdash' \textit{Ed met Ann}' \cup C + S_2 \vdash' \textit{Ed met Beth}']_{S_1}$

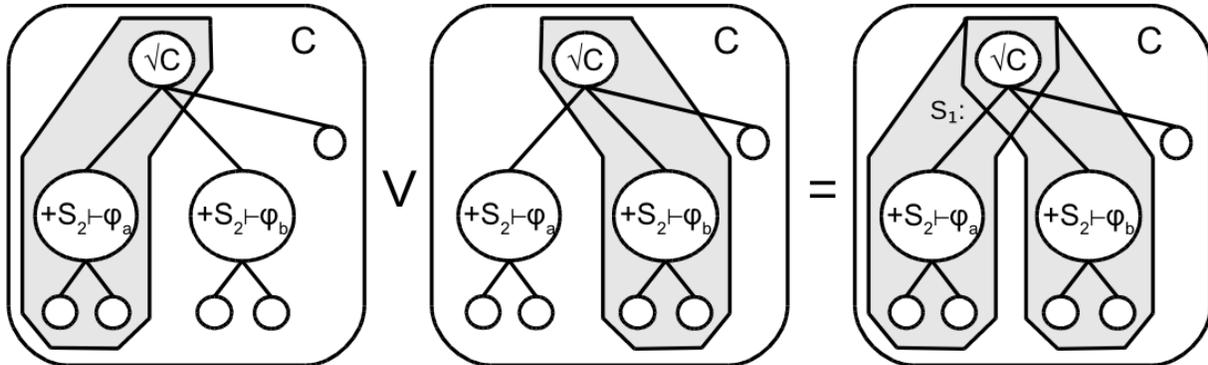


Figure 2.9: Disjunction question as disjunction of two monopolar questions

There are again two propositional discourse referents, as with the alternative question utterance (46). In fact, the model ends up with the same analysis for alternative questions and disjunctive polar questions (c.f. Figure 2.8 and Figure 2.9). In this scenario, the head of the ActP is appropriately specified as ‘?’. Therefore, this framework implies that a disjunctive Y/N question would reject both a simple *Yes* and a simple *No* from its licit responses; however, having *Yes* before the response of *Ed met Ann* or *Ed met Beth* seems to be licensed, mostly because it is not expressly disallowed. It is unclear whether this analysis would allow an answer such as *Yes, Ed met Ann and Beth*; because both propositions are separate, it may be inferred that speaker 2 could accept both proposition a and proposition b (see Figure 2.9) as they are not a negation of each other ($a \neq \neg b$; $b \neq \neg a$).

This point brings up an interesting caveat within the framework; recall that the alternative question analysis disallows simple *yes* and *no* answers. Figure 2.8 also demonstrates that, in this

example, the two discourse referents are a proposition and its negation; therefore a *both* answer is also disallowed. However, many alternative questions have two discourse referents that are not *inherently* opposed; rather it is something the speaker is supposing, based on context. For example, if speaker 1 were to utter (49) as an alternative question, it isn't the case that in all worlds in which Ed, Ann, and Beth exist, Ed could either meet Ann or Beth but not both; rather, in the alternative question interpretation, the speaker is taking for granted that *exactly one* of these alternatives holds.

(49) Did Ed meet Ann, or did Ed meet Beth? (falling accent)

Thus, the framework would have to account for the fact that, in this case, speaker 1 is supposing something similar to an assertion of 'a' implies an assertion of 'not b' ($a \rightarrow \neg b$, where \rightarrow indicates implication). Hamblin (1971) calls this "the presumption of a question", and states it "becomes a commitment of a participant who asks it, perhaps even all participants"(132). This addition would allow the framework to correctly predict that both discourse referent a and discourse referent b cannot be accepted by speaker 2. The alternative question in (50) is updated to account for this imposition by speaker 1 (bold) and also includes '.' as the head of the ActP for the second disjunct, as the second disjunct ought to have falling intonation (Bartels 1999, Pruitt & Roelofsen 2013, etc.)

(50) Did Ed meet Ann, or did Ed meet Beth?

$[[[_{\text{ActP}} [_{\text{ActP}} \textit{Did Ed meet Ann}] \textit{or} [_{\text{ActP}} \textit{Did Ed meet Beth}]]]]_{\text{S1S2}}$

with $[[[_{\text{ActP}} ? [_{\text{ActP}} \textit{Did Ed meet Ann}]]]_{\text{S1S2}} = \lambda C^*[\{\sqrt{C}\} \cup C + S_2 \vdash \textit{Ed met Ann}]_{\text{S1}}$

and $[[[_{\text{ActP}} . [_{\text{ActP}} \textit{Did Ed meet Beth}]]]_{\text{S1S2}} = \lambda C^*[\{\sqrt{C}\} \cup C + S_2 \vdash \textit{Ed met Beth}]_{\text{S1}}$

and $[[\text{or}]]_{S_1 S_2} = \lambda A \lambda A' \lambda C^* [A(C) \cup A'(C)]_{S_1}$, where A, A' : variables over speech acts

$$\begin{aligned} & \mathbf{and} \ [[S_1 \vdash [S_2 \vdash A] \rightarrow \neg [S_2 \vdash A']]] \ \mathbf{and} \ [[S_1 \vdash [S_2 \vdash A'] \rightarrow \neg [S_2 \vdash A]]] \\ & = \lambda C^* [[\{ \sqrt{C} \} \cup C + S_2 \vdash \text{'Ed met Ann'}] \cup [\{ \sqrt{C} \} \cup C + S_2 \vdash \text{'Ed met Beth'}] \cup [\{ \sqrt{C} \} \cup \\ & \quad C + S_2 \vdash \text{'Ed met Ann'} \rightarrow \neg \text{'Ed met Beth'}]]_{S_1} \\ & = \lambda C^* [\{ \sqrt{C} \} \cup C + S_2 \vdash \text{'Ed met Ann'} \cup C + S_2 \vdash \text{'Ed met Beth'} \cup C + S_2 \vdash \text{'Ed met Ann'} \\ & \quad \rightarrow \neg \text{'Ed met Beth'}]_{S_1} \end{aligned}$$

Now the framework can account for alternative questions and disjunctive Y/N questions, although it disallows simple *yes* and *no* answers for both types of questions. If, however, the disjunctive Y/N question were to be analyzed as a long bipolar question (where both parts of the disjunct are simply part of one propositional phrase), then we might imagine that the question would be analyzed as in (51). The comma is removed to indicate that the entire phrase is one propositional phrase. In fact, this type of analysis seems more applicable with narrow scope, as the parentheses indicate. In this scenario, the *yes* and *no* answers are licensed (c.f. (42) and (43)); however, it would not license more specific answers.

$$\begin{aligned} & (51) \text{ Did Ed meet Ann or (did Ed meet) Beth?} \\ & = [[[_{\text{ActP}} \text{Did Ed meet Ann or (did Ed meet) Beth}]]]_{S_1 S_2} \\ & = \lambda C^* [[\{ \sqrt{C} \} \cup C + S_2 \vdash \text{'Ed met Ann or (Ed met) Beth'}] \cup [\{ \sqrt{C} \} \cup C + S_2 \vdash \neg \text{'Ed met} \\ & \quad \text{Ann or (Ed met) Beth'}]]_{S_1} \end{aligned}$$

We would predict that these *yes* and *no* licit responses would only be available if the entire phrase was considered one proposition; the presence of a phrasal break (or comma) in between disjuncts would seem to rule out this analysis in those contexts.

Summary

Commitment Space Discourse offers a framework based on speech acts in order to account for disjunctive questions and their licit responses. The framework as described in Krifka (2015, 2016) required an extension in order to elicit the correct responses for alternative questions. In this framework as well, there were two different ways to classify disjunctive Y/N questions: as two monopolar propositions or as one bipolar proposition. The distinction in these two classifications may lie in the presence (or absence) of a comma or phrasal break between the two disjuncts. Like Inquisitive Semantics, these two classifications of Y/N questions have different answerhood conditions. Both, however, are accounted for in this framework. Table 2.3 (next page) gives a breakdown of the types of disjunction questions and their corresponding licit responses, according to the Commitment Space Discourse framework.

Disjunctive Question	ActP	Type	Licit Responses				
			Yes	No	Ed met Ann	Ed met Beth	Ed met both
Did Ed meet Ann, or did Ed meet Beth?	Rising accent on Ann and Beth	Polar, two monopolar propositions			✓*	✓*	✓*
Did Ed meet Ann, or did Ed meet Beth?	Rising accent on Ann, falling accent on Beth ¹⁰	Alternative, two monopolar propositions			✓	✓	✓ ¹¹
Did Ed meet Ann or did Ed meet Beth?	Rising accent on final disjunct	Bipolar proposition	✓	✓			

Table 2.3: Disjunctive Questions and licit answers according to CSD

An * indicates an optional Yes could occur before these licensed answers.

2.6 Typology

In English, both alternative questions and disjunctive polar questions use the same connective, *or*, to express the disjunction, and intonation is required to disambiguate the two. However, several languages mark the distinction by having a different *or* for the alternative question reading. This *or*, called the ‘interrogative disjunction’ (Haspelmath 2007), occurs only in alternative questions; the *or* that occurs in declarative disjunction or disjunctive Y/N questions is known as the ‘standard disjunction’ (Winans 2012). This relationship is assumed to be a superset (Y/N questions) and subset (Alt. questions) relationship; interrogative disjunctions only occur in interrogatives, while the standard disjunction can occur in any utterance type

¹⁰ The framework seems to allow either type of accent on both disjuncts and will simply assign the head of ActP according to the intonation observed. The intonation itself does not seem to alter the analysis

¹¹ This answer seems licensed, unless the framework adds in an assertion aspect by S₁ in alternative question scenarios that the two discourse referents cannot both be licensed as an answer without a rejection by S₂

(interrogative included). However, this assumption has received criticism from Winans (2012), who argues a complementary relationship via examples in Egyptian Arabic. A cross-linguistic examination is necessary to determine the exact nature of this relationship. Using language conditions found in natural language, an artificial language experiment may deduce the cognitive delineation of these two question types, with the idea that participants would more easily learn language conditions that adhere to their own mental demarcation between the two interpretations. Thus the assumption concerning the relationship between the two interpretations may be supported or refuted via empirical data.

Mandarin Chinese (Li and Thompson, 1981) is one such language that has two distinct lexical items for the two interpretations; *háishì* is restricted to alternative questions¹² (52), while *huòzhe* delivers a yes/no question reading (53).

(52) nǐ yào wǒ bāng nǐ háishì yào zìjǐ zuò?

you want I help you **ALTi**¹³ want self do

‘Do you want me to help you, or do you want to do it yourself?’

(Li and Thompson, 1981, 653)

(53) wǒmen zài zhèlǐ chī huòzhe chī fàndiàn dōu xíng

we at here eat **ALTS** eat restaurant all OK

‘We can either eat here or eat out?’¹⁴

(Li and Thompson, 1981, 532)

¹² *Háishì* is restricted to Alternative questions but does not occur in all alternative questions; there is a special type of alternative question (A-not-A) in Chinese, that has the restriction that the disjunction cannot scope out of the immediate clause (whereas using *háishì* allows scoping out). A-not-A questions are not within the scope of this experiment.

¹³ **ALTS** indicates the standard disjunction; **ALTi** indicates the interrogative disjunction

¹⁴ Carl Pollard (p.c.) has argued that this translation is incorrect and should be a declarative sentence; in both cases, the standard disjunction *huòzhe* would be used.

Finnish is another example of this dual representation; *vai* is an interrogative disjunction (54,55), while *tai* expresses a yes/no question (and is not restricted to questions, 56).

(54) Mattiko näki sinut vai Maija?

Matti-Q see-IMP-(3SG) you-ACC **ALTi** Maija?

‘Did Matti see you or was it Maija?’ (Sulkala and Karjalainen, 1992, 11)

(55) a. Otakko kahvia vai teetä?

you-take-? coffee **ALTi** tea

‘Do you want coffee, or tea?’ (Vainikka, 1987, 164)

b. Kahvia / Teetä

coffee / tea

‘Coffee. / Tea.’

(56) a. Otatko kahvia tai teetä?

you-take-? coffee **ALTS** tea

‘Do you want (some) coffee or tea?’ (Vainikka, 1987, 164)

b. Otan. / En.

yes / no

‘Yes. / No.’

Other languages use a standard disjunctive connective that can only occur in a bisyndetic construction (repeated before each disjunct). Somali is an example of such a language, using *ama* as a standard disjunctive and *misé* conveys the interrogative relation (57).

(57) Somali, Cushitic, Afro-Asiatic (Saeed 1993, 275)

d. **amá** wuu kéeni doonaa **amá** wuu sóo.díri doonaa

ALTS 3SG bring that **ALTS** 3SG send that

‘Either he will bring it or he will send it.’

e. *ma tégaysaa misé waad jóogaysaa?*

INT go:2SG **ALT**i here stay:2SG

‘Are you going or are you staying?’

Mauri (2008) found that if a language uses a disjunctive connective in interrogative disjunction (e.g. an Alt. question), it will also use a disjunctive connective in standard disjunction (e.g. a Y/N question or in declaratives), as opposed to juxtaposition or dubitative markers. In fact, the presence of an interrogative disjunction connective *implies* the presence of a standard disjunctive connective. However, some languages don’t have an interrogative disjunctive, simply using juxtaposition in interrogatives; but they still use a disjunctive connective in standard disjunction. Korean juxtaposes two interrogative sentences (with appropriate interrogative markers) to make an alternative sentence¹⁵ (58), while the standard disjunction is realized by either *-kena* or *-tunci*¹⁶ (59).

(58) *Wuli ka ka-l kka-yo*↑ *Mia lul ponay-l kka-yo?*↓

we NM go-PRS whether-POL Mia AC send-PRS whether-POL

‘Shall we go, or shall we send Mia?’ (Sohn 2001, 307)

(59) *Kiho ka w-ass-kena/tunci Nami ka w-ass-e*

Kiho NM come-PST- **ALTS** Nami NM come-PST-INT

‘Either Kiho or Namie came.’ (Sohn 2001, 305)

¹⁵ Interestingly, Korean also relies on “rising intonation on the first disjunct, and falling intonation in the second” (Sohn 1994, 306)

¹⁶ These two words are synonymous, except the former is considered more formal

Such languages often make use of interrogative or dubitative markers. Some languages, in fact, use such markers in both interrogative and standard disjunction, having no overt disjunctive connective. The example from Mangarayi (60) gives an example of the ‘perhaps’ dubitative adverb, which is used in all considerations of alternatives; it is not considered a true disjunctive connective because it also occurs when only one alternative is present (Mauri 2008); however, it is necessary to convey an alternative relation- in fact, there is no other way to do so. This is the only disjunctive strategy available in the language; there is no equivalent of ‘or’.

(60) Mangarayi, Gunwinguan, Australian (Merlan 1982, 39)

manaya ja- ϕ -nina-n *manaya dayi*

perhaps 3-3SG-come-PRES *perhaps* NEG

‘Perhaps he’ll come, perhaps not.’ (i.e. ‘he may or may not come’)

This type of language, which lacks a true disjunctive connective and instead uses interrogative, dubitative, etc. markers (‘irrealis markers’) for both types of disjunction, is infrequent in the world’s languages (Mauri 2008). No language, however, has been found where the standard disjunction can *only* be conveyed by juxtaposition of overtly irrealis clauses, while the interrogative disjunction is expressed by means of a connective.

These naturally occurring conditions of the disjunctive question relationship are mimicked in an artificial language experiment in order to determine the nature of the correlation.

2.7 Overview of Terminology

Across the two semantic theories and typological discussion, authors use various terms to describe disjunctive questions. Table 2.4 indicates the breakdown of terminology for Alt. and Y/N questions. Alt questions are known as closed interrogatives in IS or as two monopolar propositions in CSD; they use the interrogative disjunctive connective according to typology.

Both IS and CSD have two different terms for Y/N questions, which are determined by the licit responses for the question (Figure 2.10). Y/N questions use the standard disjunctive connective across languages (Haspelmath 2007, Mauri 2008).

Disjunctive Question	IS	CSD	Typology
Alt question	Closed interrogative	Two monopolar propositions	Uses Interrogative connective
Y/N question	Open interrogative Block interrogative	Two monopolar propositions Bipolar proposition	Uses Standard connective

Table 2.4 Summary of terminology of disjunctive questions for IS, CSD, and Typology

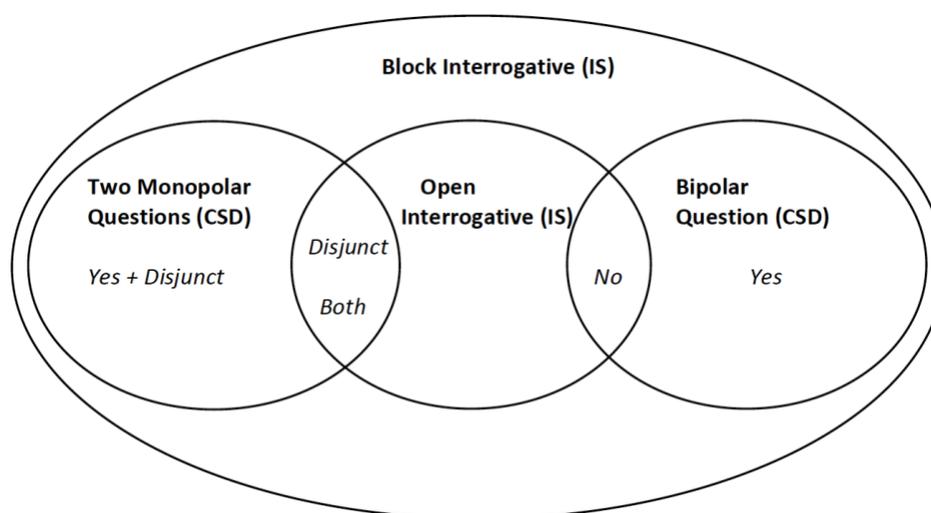


Figure 2.10: Relationship of licit responses for different Y/N questions across IS and CSD. Licit responses taken from Tables 2.2 and 2.3. Question terminology is in bold, while responses are italicized.

Of particular note is that CSD's two Y/N questions have a complementary relationship in terms of licit responses: A Y/N question as two monopolar questions allows *Yes + Disjunct*, a disjunct and *Both*, while the bipolar question interpretation allows *Yes* and *No*. The open interrogative has a subset relationship with the Block interrogative in IS; the open interrogative

does not allow *Yes* as a bare particle or with disjuncts, while the Block interrogative allows those responses, as well as *No*, a disjunct, and *Both*.

2.8 Summary

Empirical work on disjunctive questions has led to an incomplete descriptive stance on how to demarcate alternative questions and polar questions. The rise of the first disjunct and fall of the second disjunct seem to be important, although not quantified as of yet. The production experiment (Chapter 3) was conducted in order to change the assumption regarding intonational contours to tunes that have experimental support.

The insertion of *either* into disjunctive questions has historically been limited to the Y/N interpretation; this is yet another disjunctive question assumption tested, via the text (Chapter 4) and perception (Chapter 5) experiments. Those two experiments also test variations on a typical disjunct answer, by including a cleft answer (in both text and perception experiments) and a disjunct answer with an uncertainty contour (in the perception experiment). Current work on disjunctive questions allows a disjunct response for an Alt. question without restriction; responses to Y/N questions are even less restricted. These two experiments reveal that both structure and intonation must be taken into account when allowing a disjunct response.

Theoretical work incorporates both intonational predictions and answerhood conditions in two different frameworks- Inquisitive Semantics and Commitment Space Discourse. Both frameworks offer specific predictions concerning disjunctive questions and their licensed responses, which will be explored in the methodology of the production, text, and perception experiments. These frameworks are built around the same assumptions the experiments are measured against, and therefore provide predictions on the validity of those assumptions.

A typological look at languages reveals different methods to encode the alternative question/polar question relationship: only through intonation (English); two separate connectives (Finnish, Chinese); two connectives, but the standard disjunction occurs bisyndetically (Somali); a standard connective without an interrogative connective, where alternative questions are realized via juxtaposition of interrogative clauses (Korean); or without any connective, where both types of questions are realized via a dubitative marker (Mangarayi). These language examples will be used as conditions in the artificial language experiment (Chapter 6) in order to explore the relationship between the Alt. interpretation and the Y/N interpretation. Their connection is traditionally assumed to be a superset/subset relationship (with Alt. questions being the subset); the artificial language experiment will offer insight to test this claim.

Chapter 3: Production Experiment

3.1 Introduction

In the previous literature, the canonical intonational contours for Alt. questions and Y/N questions have been given via author introspection, and licit or illicit tunes have been specified according to these intuitions. The production experiment presented in this chapter was conducted in order to elicit Alt. and Y/N questions in minimal pair form, to determine the range of contours that are licit for each interpretation type. The aim is to determine how well the assumptions underlying disjunction question intonation contours are borne out by data from naïve speakers, and, if necessary, replace those assumptions with contours derived from experimental evidence. The production experiment is unique in that it contains no sound itself, nor any utterance directed text; rather, participants use images in order to determine context. Participants were successfully able to disambiguate between the two interpretations in this design, indicating the salient contextual difference between an Alt. question and Y/N question.

3.2. Methodology

3.2.1 Design and Materials

The images for the production experiment were constructed to bias the participant to ask an Alt. question (Figure 3.1) or a Y/N question (Figure 3.2). Each stimulus token was divided into three image slides: a question slide, a response slide, and a result slide. In the Alt. question slide, Jordi stood in the middle of two gates. Each gate had one sign, to indicate the animal inside. After the participant recorded their question, Jordi would always respond with one of the two animals listed on the signs (response slide). In this way, the participant may learn that this scenario is an alternative question scenario, if they didn't initially give the Alt. question tune. The animal with which Jordi would respond was counterbalanced between the left sign and right

sign to appear with equal frequency. The result slide would reveal Jordi by the same gates, but now with the animal that Jordi requested to see. In the Alt. question scenario, Jordi would always end up seeing an animal.



Figure 3.1: Alternative Question scenario example, question slide

In the Y/N question slide (Figure 3.2), Jordi would stand to the left or right (counterbalanced) of a gate, which had two animals listed beside it. Importantly, the zoo rules stipulate that only one animal will emerge from a gate and visitors do not get to choose which animal that may be. In this case, the difference between the disjuncts is not salient, and the context ought to bias participants to the Y/N interpretation; i.e. Jordi can choose to open the gate (“Yes”) or leave it closed (“No”), but he cannot choose which animal will emerge, if the gate opens. The response slide for this context was always a “Yes!” or a “No!”. If Jordi responded with a “Yes!”, then the final slide would reveal Jordi with one of the two animals listed (counterbalanced between top and bottom animal throughout experiment). If Jordi responded with a “No!”, then the participant would see the same slide as they initially saw (e.g. Figure 3.2).

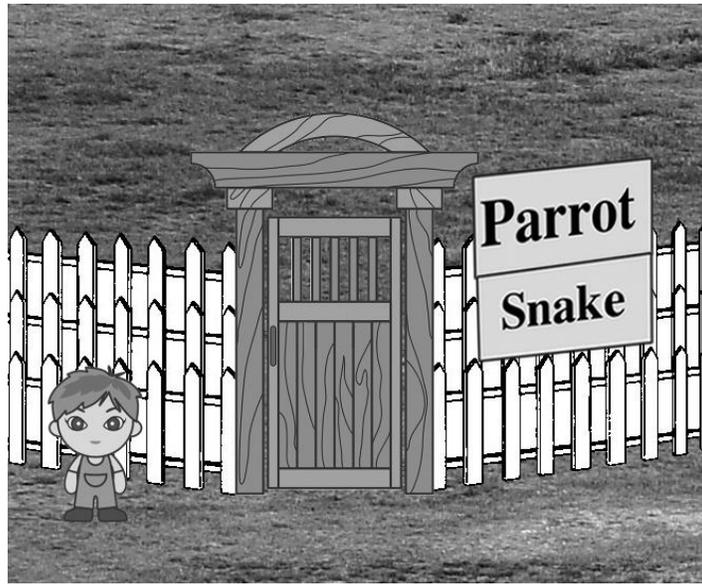


Figure 3.2: Y/N question scenario example

The experiment also contained single gate/single animal slides (Figure 3.3), which were included as a baseline simple polar question. The side of the gate that Jordi stood on was also counterbalanced in this context. Jordi also responded to this image context with either “Yes!” or “No!”, and his response dictated whether or not an animal was with Jordi in the result slide.

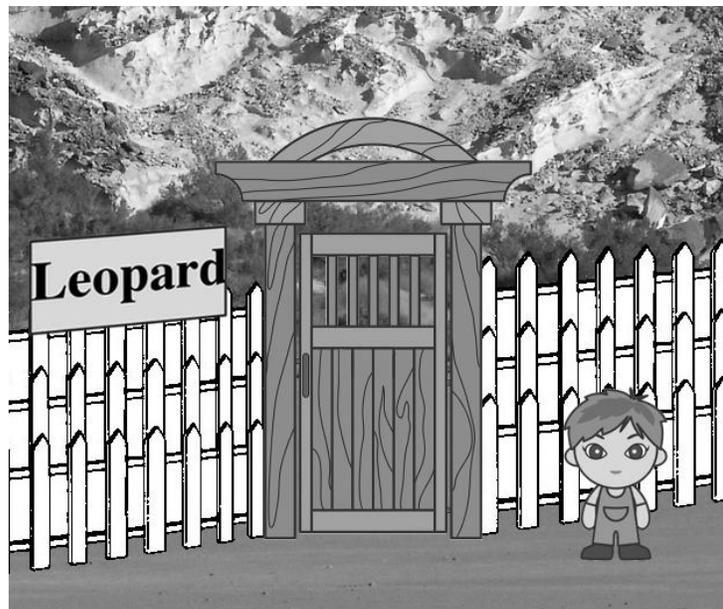


Figure 3.3: Baseline example (simple polar question)

The experiment was divided into four zoo regions (Arctic, Africa, Asia, and Australia), with nine tokens in each region. The nine tokens consisted of three Alt. questions, three Y/N questions, and three simple polar questions in each region. Therefore, the whole experiment contained twenty-four stimuli (twelve Alt, twelve Y/N) and twelve fillers (simple polar questions). There were two lists of the experiment, in order to counterbalance the Alt and Y/N questions (list 1 would have tokens 1-3 as Alt, 4-6 as Y/N, while list 2 would have tokens 1-3 as Y/N, 4-6 as Alt. etc). Both lists had the same simple polar questions. There were also sixteen different region order combinations, making for 32 versions total.

3.2.2 Procedure

The production experiment was designed as a children's game, in which the participant was narrating a child's experience through a petting zoo. In this scenario, the participant was expected to ask the child (Jordi) a question, and Jordi would give a response to the question. The important aspects of the instructions are as follows:

- Jordi is allowed to wander the zoo and open gates to see animals
- Jordi can only see one type of animal at a time
- If more than one animal is inside a gate, only one type of animal will emerge
- Jordi will appear by one or more signs, which indicate a path or a gate. Participants must use one sentence to ask Jordi if he wants to see the animals listed (if he is by gates) or go to the regions (if he is by paths)
- Participants must ask about all regions or animals listed, but only use one sentence to do so.
- Jordi will respond to the question and his response will dictate what happens next

In order to get to the regions, the participant first had three training slides (two Y/N questions, one Alt. question) to get to their first region; there was one simple polar question between regions 1 and 2 and regions 3 and 4. Halfway through the experiment (between regions 2 and 3) there was an additional Y/N question and Alt. question. These training slides served to move Jordi through the zoo (e.g. “Do you want to go to Africa or Australia?”) so that each participant visited each of the four regions (but in a pre-determined order, depending on their version). Once in a region, the participant would see all the slides associated with that region before moving onto the next region. The animals within each region were seen in a random order.

For each question, the participant was required to hit the “record” button to record their question. The experiment allotted ten seconds for the recording, and then would cut off. Participants could also stop the recording at any time with the “stop” button. Participants could re-record as many times as they wanted before hitting “next” to see Jordi’s response to their question. A participant could not push “next” until they had hit the “record” button. Participants were unable to hear their own recordings.

3.2.3 Participants

95 participants were recruited from Amazon Mechanical Turk (AMT), 52 from list 1 ordering and 43 from the list 2 ordering. Each version (of the 32 versions) had at least two participants. The experiment ran from 7-3-18 until 1-11-19. The entire experiment took 10-30 minutes. Participants also filled out a brief survey which asked their gender and native language(s). They were paid \$3 for their participation.

3.3 Predictions by the Semantic Theories

The production experiment controlled the answer portion for all the stimuli; no matter what the participant recorded, Jordi would always answer with a licit response, according to the type of question that ought to have been recorded. In addition, the image scenario was also controlled during the experiment. Because the participant was told that Jordi could only see one animal at a time, the ‘exactly one’ presumption holds for the two signs, two gates scenario. And because Jordi could not choose which type of animal emerged from a gate, a single gate, multiple sign image ought to be produced with a Y/N tune, rather than an alternative tune. The two theories offer predictions on whether a phrasal break will occur, whether the entire phrase of each disjunct will be focused, and the final contour of the utterance.

The phrasal break predictions for Inquisitive Semantics are based on the number of lists in the utterance; utterances with two lists have a phrasal break, while single list utterances do not. Similarly, in Commitment Space Discourse, utterances that are made up of two monopolar questions will contain a phrasal break; this is exemplified in the text examples by a comma (Krifka 2015, 2016). Both theories claim that an utterance with a phrasal break will have focus¹⁷ on both disjuncts (i.e. narrow focus on each disjunct), while one without a phrasal break will have focus on the whole phrase (i.e. broad focus; Roelofsen and van Gool 2010, Krifka 2016). Both theories also allow Y/N questions to be uttered with and without a phrasal break; however, the two different Y/N productions (one with a phrasal break; one without) permit different answerhood conditions. Predictions are based on the idea that participants will attempt to fit their recordings (i.e. intonation) into the given context (Heidenreich 2014b), and thus the

¹⁷ Focus here is realized as an emphasis on the acoustic signal, i.e. a pitch accent (Cutler et. al. 1996).

types of responses seen are also used in formulating predictions. For example, Inquisitive Semantics allows different Y/N tune variations that permit a *No* response; however, only a Y/N question without a phrasal break allows a *Yes* response. The following tables illustrate the predictions as outlined by previous experimental results (Heidenreich 2014b) (Table 3.1), Inquisitive Semantics (Table 3.2), and Commitment Space Discourse (Table 3.3).

Jordi's Answer	Image scenario	Previous Experimental Results			
		Phrasal break	Final contour	Focus	Type
"Yes"	Two signs, one gate	Present or not present	Rising	Whole phrase or each disjunct	Y/N
"No"	Two signs, one gate	Present or not present	Rising	Whole phrase or each disjunct	Y/N
A disjunct	Two signs, two gates	Present or not present (first disjunct ends high)	Falling	Each disjunct	ALT

Table 3.1: Predictions of tune based on previous experimental results

The predictions are formulated based on the context gathered by the image scenario, as previous studies have not been done on the relationship between disjunctive questions and their answers.

Answer	Image scenario	IS predictions			
		Phrasal break	Final contour	Focus	Type
"Yes"	Two signs, one gate	No	Rising	Whole phrase	Y/N (Block)
"No"	Two signs, one gate	Yes or no	Rising	Whole phrase or both disjuncts	Y/N (Open or Block)
A disjunct	Two signs, two gates	Yes	Falling	Both disjuncts	ALT (Closed)

Table 3.2: Predictions of tune based on Inquisitive Semantics theory¹⁸

¹⁸ Predictions are formulated based on image scenario context, as well as the relationship between context and given response

Answer	Image scenario	CSD predictions			
		Phrasal break	Final contour	Focus	Type
“Yes”	Two signs, one gate	No	Rising	Whole phrase	Y/N (Bipolar) ¹⁹
“No”	Two signs, one gate	No	Rising	Whole phrase	Y/N (Bipolar)
A disjunct	Two signs, two gates	Yes	Falling	Both disjuncts	ALT ₂₀ (Two monopolar)

Table 3.3: Predictions of tune based on Commitment Space Discourse theory^{19,20}

Note that while the final contour and the ultimate type of question predicted by each theory is the same, the focus structure and presence/absence of a phrasal break differs for each theory. In particular, previous empirical results (Heidenreich 2014b) suggest that it isn't the presence of a phrasal break between disjuncts that determines an alternative question, but rather the pitch at the end of the first disjunct (must be high); whether that is achieved through a H- or, for example, a L*+H, is unimportant. In addition, empirical results suggest that both disjuncts of Y/N questions could be focused (have a pitch accent); this is not sufficient to make the utterance an alternative question (see Bartels 1999 as well).

¹⁹ Only the bipolar Y/N question allows *yes* and *no* as licit answers; the disjunctive of two monopolar questions requires uttering a specific disjunct as there are two discourse referents (Krifka 2015)

²⁰ This is the only predicted type for this scenario, as the participant is told that if Jordi chooses A, then he cannot also choose B ($A \rightarrow \neg B$); the propositions are in negation to each other

3.4. Results

The twenty-four stimuli (12 Y/N and 12 Alt) from each participant were given a ‘correct’ score²¹ as well as an Error factor. A correct answer was coded as “1” for an Error Type.

3.4.1 Error Analysis

Six errors were deemed instructional errors (the participant didn’t understand the instructions correctly or use the experiment buttons correctly, which resulted in something other than a disjunctive question); those errors were a question that was cut-off (“cut-off”), a question that asked about only one animal (“one”) or no animals (“none”), a recording without sound (“0”), a question that blended the two animals listed on the signs into one animal (“blended”), e.g. *bat-crane* instead of *bat or crane*, and an utterance that was clearly not a question (“nonquestion”). The instructional errors combined accounted for less than 10% of the data, indicating that participants were able to understand the instructions.

The remaining three errors were considered interpretation errors: the participant understood the instructions and correctly produced both animals in a question, but it wasn’t the right type of question. If the question was supposed to have an Alt. tune but was produced with a Y/N tune, then the error was “Y/N”; similarly, if the question was supposed to have a Y/N tune but was produced with an Alt. tune, then the error was “ALT”. In a fairly large number of instances, participants produced the Y/N tune correctly but used the conjunct *and* instead of the disjunct *or*²²; this error was coded as “AND”. The full breakdown of errors is listed in Table 3.4.

²¹ This was determined by listening to each recording without knowing the intended type and deciding what types of answers the question would allow. If the question would only allow one disjunct as an answer, it was deemed an Alt question. Otherwise, it was deemed a Y/N question.

²² This error was noted only if the type of question produced was a Y/N question (as context requested). If the participant produced a Y/N tune with “and” for an Alt. context, the error was counted as “Y/N”.

Error Type	Count	% of Total
1 (Correct)	1369	60.04%
ALT	177	7.76%
Y/N	178	7.81%
and	352	15.44%
0	1	0.04%
blended	46	2.02%
Cut-off	22	0.96%
none	1	0.04%
nonquestion	65	2.85%
one	69	3.03%
TOTAL	2280	100%

Table 3.4: Breakdown of Errors in Production Data

Additionally, a factor of “Half” was given, where each stimuli was given either a “1” to indicate it occurred during the first half of the experiment and “2” to indicated it occurred during the second half of the experiment. This factor was constructed for three reasons:

1. A production experiment that elicits 40+ recordings may induce fatigue, which could mean participants perform worse as the experiment continues.
2. A new type of production experiment that attempts to elicit tunes based solely on image context may require a longer training period than the 3 training slides before the first stimuli, which could mean that participants perform better as the experiment continues. Additionally, recall that participants receive two additional training slides at the halfway point, which may also influence performance.
3. Both the Inquisitive Semantic predictions and the Commitment Space Discourse predictions are based on answerhood condition predictions, which may require the

participant to see a certain amount of licit answers in order to understand the intonational pattern required.

3.4.2 Standard Error

For the visualizations used in the data analyses that follow, the error bars typically represent the overall standard error of the experiment (c.f. Figure 3.4, 3.5, 3.6, 3.7, etc., i.e. all error bars will be the same length). However, there are two instances where this is not the case: if a single factor is being examined, and the factor contained more than one level, the error bars represent the standard error of the difference of means (c.f. Figure 4.1, 4.2 in Chapter 4, Figure 5.1, 5.2 in Chapter 5, etc.). If the visualization has two factors represented, and one of the factors has more than two levels, then the error bars represent the standard error of the data within that cross-factor combination (c.f. Figure 4.3 in Chapter 4, Figure 5.3 in Chapter 5, etc.).

3.4.3 Stimuli Analysis

A logistic regression model was constructed using the ‘glmer’ function implemented in the R package ‘lme4’ (Bates et. al., 2015). The dependent variable was the binary “Correct” value, while the factors “Type” and “Half” were independent variables. Logistic regression models have been shown to offer many advantages over ANOVA in linguistic data such as forced-choice variables, question-answer accuracy, and even choice in production (Jaeger 2008). All data, including errors, were used in this analysis. Each subject sees both Type A and Type B (Alt and Y/N stimuli), so a random slope and intercept for participant (pID) was added. This was the model of best fit (Table 3.5), after pruning out random slopes and intercepts for stimuli number (Number), as described below.

Each stimuli number had both Type A and Type B recordings, so a random slope and intercept for Number were initially added to the model. This model did not converge, so two

models were constructed: the first, without a random slope for Number, and the second without a random intercept for Number. Using likelihood ratio tests to examine the significance of these random effects components (Levy 2007a), it was determined that both of these models were inferior to a model that simply had random slopes and intercepts for participants. The likelihood ratio tests operate on a Chi square value, which was significant when the probability (p) was less than 0.05 (Levy 2007a, b). Since the random slope for Number was not significant, it may be inferred that the difference between the amount of correct Alt. questions and amount of correct Y/N questions was comparable across all items. The random intercept for Number not being significant means that none of the items were biased to one Type of disjunctive question over the other.

Model comparison using likelihood ratio tests (Levy 2007a, b) revealed that both random slopes and intercepts for the participant random factor were necessary in the model of best fit (i.e. more pruning was not necessary). Additionally, both “Type” and “Half” were found to be significant (a model with the interaction included was tested and found not to be significant ($p=0.2517$); furthermore, including the interaction in a model produced a worse model fit (model comparison $p=0.2539$)).

	Estimate	Std. Error	z value	Pr(> z)
Intercept	2.5502	0.4106	6.211	5.27e-10
TypeB	-3.4137	0.5308	-6.432	1.26e-10
Half2	0.7285	0.1716	4.244	2.20e-05

Table 3.5: Best fit model parameters for the production data.

The R code that produced this model: correct~type+half+(1+type|PID)

Table 3.5 reveals a significant positive intercept ($\beta = 2.55$, $SE = 0.41$, $z = 6.2$, $p < 0.0001$), which demonstrates that participants had more correct answers than incorrect answers,

and therefore were able to successfully do (or learn) the experiment. A significant value for Type with a negative estimate ($\beta = -3.41$, $SE = 0.53$, $z = -6.4$, $p < 0.0001$) indicates that participants performed worse on the Y/N trials vs. the ALT. trials, which is also shown in Figure 3.4. The figure averages the Correct value across the experiment, so that the mean (between 0 and 1) is indicative of the percent correct in the experiment, grouped by disjunct type (Alt or Y/N). Finally, a significant value for Half with a positive estimate ($\beta = 0.7285$, $SE = 0.17$, $z = 4.2$, $p < 0.0001$) indicates that participants performed *better* in the second half of the experiment in comparison with the first half, as seen in Figure 3.5 (next page). That is, participants were already using the tunes properly in the first half of the experiment but became more consistent as they progressed through the experiment. Because the image stimuli remained the same throughout the experiment (e.g. Y/N stimuli always had one gate or path with two signs), it can be inferred that participants were able to successfully learn the correct tune to use based on Jordi's *response* to their recordings. For example, because Jordi only responded with "Yes" or

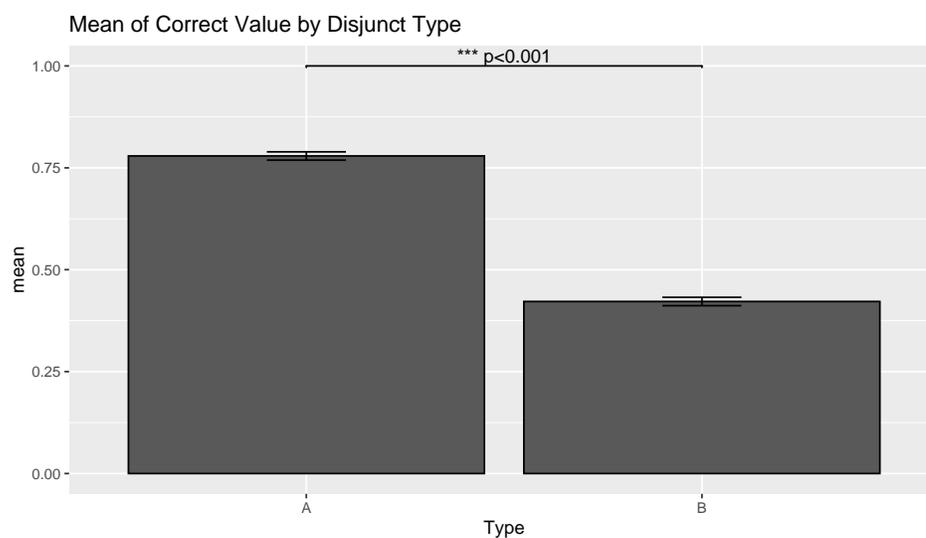


Figure 3.4: Mean of Correct value grouped by disjunct Type (A= Alt, B= Y/N).

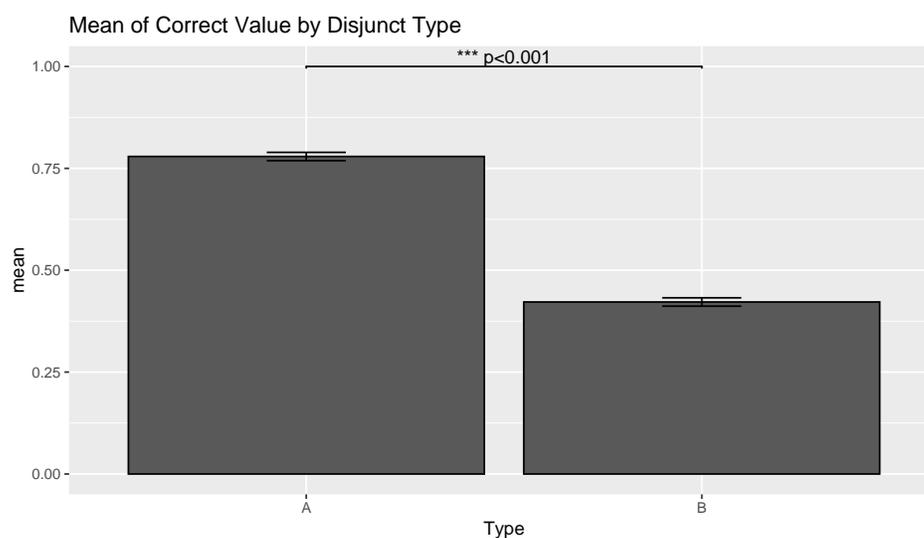


Figure 3.5: Mean of Correct value grouped by Half

(1=1st half of experiment; 2 = 2nd half of experiment).

“No” in Y/N image situations, the participants learned to ask a Y/N question tune. The fact that the interaction between Type and Half was not significant indicates that participants improved about equally well for both Types from the first half to the second half.

3.4.4 Prosodic Annotation

Examples for prosodic coding were chosen based on accuracy of recording the correct ‘type’ (Alt. vs. Y/N question) and also version number. Since the Half variable was found to be significant, one participant was taken from each region ordering for each version (32 in total). The participant had to have scored at least 15 of the 24 recordings correct, without discounting any errors (cut-off, ‘and’ instead of ‘or’, nonquestion, etc.) One combination of region ordering and version did not have a participant fit into this restriction, only because the chosen participant (P25) had enough correct (22) but the sound quality was too poor to attempt an analysis. In total,

31 participants' data were analyzed and ToBI annotated by myself solely²³, and from each recording, the following was extracted:

1. Pitch accent on first disjunct (ToBI value, e.g. H*, and pitch)
2. Pitch accent on second disjunct (ToBI value and pitch)
3. Phrase accent after first disjunct (ToBI value and pitch)
4. Phrase accent after second disjunct (ToBI value and pitch)
5. Phrase boundary tone (ToBI value and pitch)
6. Duration (in seconds) of first disjunct
7. Duration (in seconds) of second disjunct
8. Disjunct used by participant (or/and)
9. Sentence duration
10. 1st and 2nd animal names, and full utterance
11. Independent factors were also recorded for each stimuli: intended type (Type), version (1 or 2), region (Africa, Arctic, Asia, Australia), item number (number), item order (order), half (1 or 2), # of recordings the participant attempted for the trial, and gender.
12. 'Correct' factor in terms of ALT or Y/N; Error Type ('and', 'cut-off', 'nonquestion', 'one', 'ALT', 'Y/N'; a 'correct' utterance received a '1' in this factor); and if the first and second animals were said in the reverse order from the majority ('reverse'). This 'reverse' was not calculated as an error (17 examples, approximately 2.28% of responses). Table 3.6 shows the

²³ When annotating a participant, all annotations were done in alphabetical order by region and number (e.g. Africa_1) where number was not order seen by participant but a static variable. The type of question (Alt. vs. Y/N) the recording was supposed to be was unknown when annotating the recordings.

breakdown of errors for the 31 participants whose recordings were analyzed in this manner.

Instructional errors account for less than 3% of the data (2.55%).

Error Type	Count	% of Total
Correct	605	81.32%
ALT	50	6.72%
Y/N	33	4.44%
and	37	4.97%
Cut-off	5	0.67%
nonquestion	3	0.40%
one	11	1.48%
TOTAL:	744	100.00%

Table 3.6: Breakdown of Errors from 31 participants

The data were first analyzed in the same way as the entire dataset: correct as the dependent variable, type and half as independent variables, with a random intercept for participant and random slope for participant by type.²⁴

	Estimate	Std. Error	z value	Pr(> z)
Intercept	2.9754	0.5599	5.314	1.07e-07
TypeB	-2.1906	0.6594	-3.322	0.000894
Half2	0.7358	0.2193	3.355	0.000793

Table 3.7: Best fit model for filtered production data.

The R code that produced this model: correct~type+half+(1+type|participant)

Again, the intercept ($\beta = 2.97$, $SE = 0.56$, $z = 5.3$, $p < 0.0001$) as well as the independent factors of Type ($\beta = -2.19$, $SE = 0.66$, $z = -3.3$, $p < 0.001$) and Half ($\beta = 0.74$, $SE = 0.22$, $z = 3.36$, $p < 0.001$) were significant. This was the best fit model after pruning. The model that contained the interaction between type and half was not significant and a worse fit for the data ($p = 0.7495$).

For this data subset, the full order of the stimuli (1-24) was also considered; that variable was

²⁴ Again, the model that had a random slope and intercept for item(number) did not converge; those with either just a random slope or just a random intercept were inferior to a model that included neither. The remaining analyses in this experiment will only contain random intercepts for participants and random slopes for participant by type, unless otherwise noted.

switched with the ‘half’ variable to determine if ‘order’ was a better fit. Surprisingly, ‘half’ was the variable in the model with the better fit ($p < 2.2e-16$). Recall that halfway through the experiment, participants were given two training slides, one Y/N and one ALT; perhaps this additional training in the middle of the experiment contributed to the dichotomy between the first and second parts of the experiment. The additional independent factor of gender was entered into the best-fit model and found to be not significant ($p=0.617572$, model comparison $p=0.6172$). Men and women performed equally in the task.

Therefore, Half and Type are significant when it comes to ‘correct’. Type again had a negative estimate ($\beta = -2.19$, $SE = 0.66$, $z = -3.3$, $p < 0.001$), indicating that participants performed worse for Y/N questions than ALT questions (Figure 3.6). Half again had a positive estimate ($\beta = 0.74$, $SE = 0.22$, $z = 3.36$, $p < 0.001$), indicating that performance improved in the second part of the experiment in comparison to the first (Figure 3.7). Comparing Figures 3.4 and 3.5 to figures 3.6 and 3.7, respectively, reveal that the subset of participant data had similar characteristics to the entire dataset.

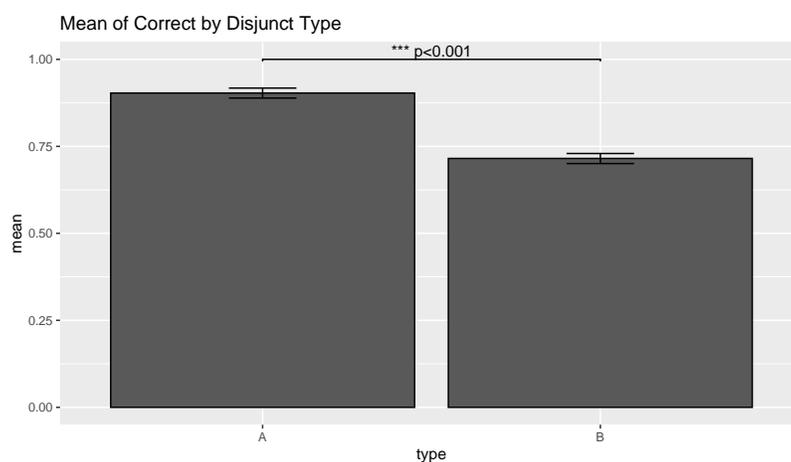


Figure 3.6: Mean of Correct value grouped by disjunct Type (A= Alt, B= Y/N)

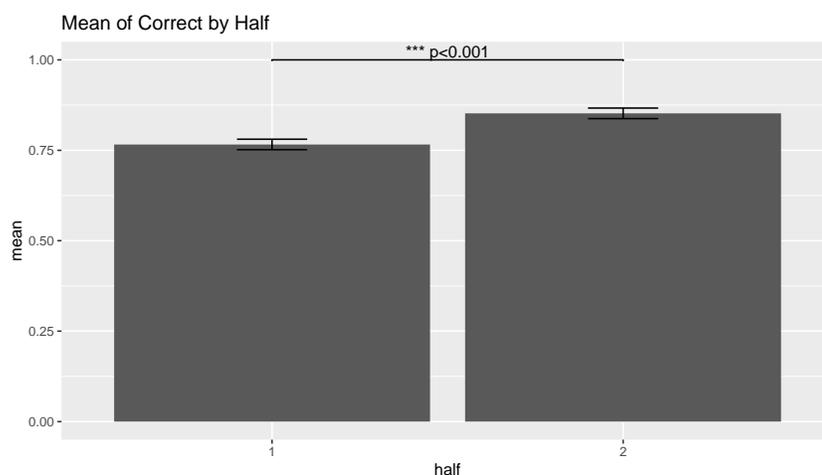


Figure 3.7: Mean of Correct value grouped by Half of the experiment

3.4.5 Final Contour

3.4.5.1 Phrasal Accent: Phonological Properties

Next, let us explore the final contour, as the theories are united in their predictions on how the final contour ought to be realized for each type of disjunction. Nearly all of the recordings contained the disjuncts in utterance-final position; only 13 utterances (5 alternative questions, 8 Y/N questions) had additional words after the disjuncts. We will assume that an analysis of either the final intonational phrase accent (ip) or intonational boundary tone (IP) factor will result in near identical results; therefore, we will specifically look at the phrasal accent after the second disjunct (D2ip), since previous empirical results (Heidenreich 2014b) and some theories (Bartels 1999) have identified the phrasal accent as the important ‘final fall’ in an alternative question. A logistic-mixed effects regression with the phrasal accent after the second disjunct as the dependent variable (either H- or L-) reveals that type of question²⁵ is significant in determining which type of phrasal accent is present (Table 3.8). The negative estimate of Type

²⁵ Half was just below significance in the model ($p=0.0559$), and thus was dropped.

($\beta = -5.17$, $SE = 0.56$, $z = -9.2$, $p < 0.0001$) indicates that the Y/N questions were correlated with H- vs. L-. The positive intercept ($\beta = 3.12$, $SE = 0.52$, $z = 5.99$, $p < 0.0001$) reveals that there were more L- than H- in the analysis.

	Estimate	Std. Error	z value	Pr(> z)
Intercept	3.1182	0.5204	5.992	2.08e-09
TypeB	-5.1703	0.5649	-9.152	< 2e-16

Table 3.8: *Glmer with D2ip as dependent variable. R code that produced this model: D2ip ~ type + (1 + type | participant)*

A full breakdown of the final phrasal accent across all 31 participants (Table 3.9) shows that a H- was produced in intended Y/N questions approximately 84% of the time; intended Alt. questions had L- approximately 89% of the time. These percentages are even higher when observing correct productions alone (over 97% for Y/N; over 98% for Alt.). The type of accent was not significantly different when comparing male to female, indicating that both genders were consistently producing the same phonology.

	All Data ²⁶		Correct Only	
	H-	L-	H-	L-
Alt.	10.3% (38)	89.7% (331)	1.8% (6)	98.2% (330)
Y/N	84.1% (296)	15.9% (56)	97.4% (259)	2.6% (7)

Table 3.9: *Final phrasal accent across question Type, including and excluding errors. Token count is included in parentheses.*

A clear distinction is made in the choice of phrasal accent, with a near complementary distribution between the H- for Y/N questions and the L- for Alt. questions. However, the theories specifically state that it isn't the phrasal accent per say, but the final contour, that

²⁶ Instructional errors were removed from the analysis, as productions with only one disjunct (“one”), cut-off (“cut-off”), or not interrogative in nature (“nonquestion”) were missing the relevant data or would not inform the analysis.

contributes to interpretation; specifically, whether the final contour is ‘rising’ or ‘falling’.

Therefore, the pitch of the second disjunct must also be taken into account.

3.4.5.2 D2 Pitch Accent: Phonological Properties

Where the phrasal accent for the two types of questions were in near complementary distribution (H- corresponding to Y/N questions and L- corresponding to Alt. questions), the pitch accent had the most variation in Alt. questions, while Y/N questions contained a L* most of the time (Table 3.10).

	!H*	H*	L*	L*+H	L+H*
Alt.	2.44% (9)	38.75% (143)	53.12% (196)	5.69% (21)	0% (0)
Y/N	1.14% (4)	12.22% (43)	83.24% (293)	3.13% (11)	0.28% (1)

Table 3.10: Second Disjunct pitch accent distribution across question type. Token count is included in parentheses.

Even more interesting, the twenty-one examples of a L*+H pitch on the second disjunct of an intended Alt. question were all produced as an Alt. question (Table 3.11, next page).

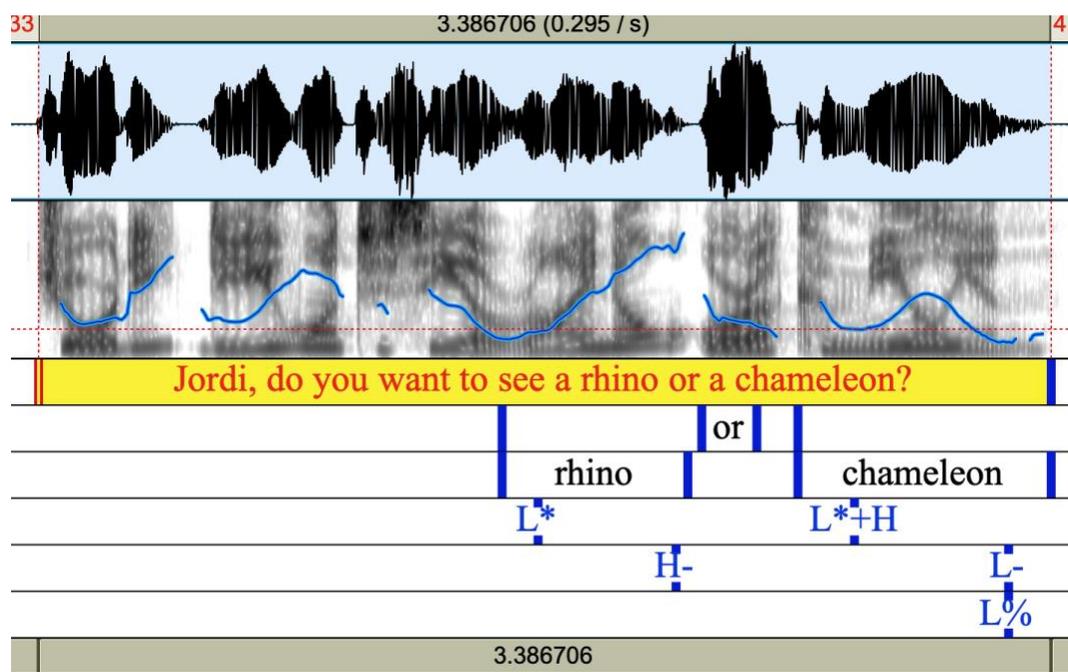


Figure 3.8: L*+H pitch accent on second disjunct (chameleon) of Alt. question.

	!H*	H*	L*	L*+H
Alt.	2.38% (8)	41.67% (140)	49.70% (167)	6.25% (21)
Y/N	1.13% (3)	9.40% (25)	86.84% (231)	2.63% (7)

Table 3.11: Second Disjunct accent placement percentages across question type, excluding errors. Token count is included in parentheses.

The productions were from twelve different participants (7 females, 5 males), indicating that the pitch accent choice was not individually biased. Of these twenty-one productions, a third of them (7) occurred when the second disjunct was “chameleon” (Figure 3.8). Perhaps the polysyllabic stress structure of this particular word contributed to the accent choice, although the accent also occurred on many bisyllabic words (6 occurrences on 5 words), trisyllabic words (5 occurrences on 3 words), and even two separate monosyllabic words (3 occurrences; two on “seal” and one on “snake”).

Perhaps the L*+H choice was taken in order to accentuate the ‘final fall’ that Alt. questions are supposed to exhibit. Regardless, nearly 50% of Alt. questions were produced with a L* on the second disjunct, an interesting pitch accent choice when the utterance ought to end with a fall. It is possible that relative L*’s on Alt. questions were higher than their L* counterparts on Y/N questions, which would allow Alt. questions to still exhibit a final fall. Phonetic data is used in order to gauge the perceptual salience of this final contour.

3.4.5.3 Final Contour: Phonetic Properties

A cursory look at the phonetic data from the pitch accent of the second disjunct and the final phrasal accent reveals that female pitch (fundamental frequency, or F0 frequency) range (whether rising or falling) is larger than their male counterparts. This is to be expected, as female speakers have been found to repeatedly produce a wider pitch range than male speakers

(Simpson 2009). However, this wider pitch range doesn't necessarily translate to a wider range in perception, as listeners perceive sound in a non-linear manner. Various acoustic scales have been developed in order to scale F0 frequency accurately (mel, bark, etc.), though the semitone scale seems to be the best choice when plotting intonation contour changes in which men and women are to be compared (Nolan 2003). Therefore, the change in F0 frequency from the second disjunct to the phrasal accent is used in scaling the final contour to semitones.

When looking at the final contour (scaled to ST) as the dependent variable, the intercept, type, and interaction between type and half were significant (Table 3.12). Type had a negative estimate ($\beta = -9.36$, $SE = 0.91$, $z = -10.34$, $p < 0.0001$), reflecting the negative slope of a Y/N question (the phrasal accent F0 frequency was higher than the F0 frequency on the second disjunct) and the positive slope of the Alt. question (the phrasal accent F0 frequency was lower than the F0 frequency on the second disjunct). Thus, despite the fact that the Alt. questions had a L* on the second disjunct half the time, the F0 frequency drop was still present, and significantly different from a Y/N final contour. Although half was not significant ($\beta = 0.16$, $SE = 0.55$, $z = 0.295$, $p > 0.05$), the interaction between half and type was significant ($\beta = -1.81$, $SE = 0.79$, $z = -2.30$, $p < 0.05$) and was a better model fit than a model without half as a factor (model comparison $p=0.01404$).

	Estimate	Std. Error	t value	Pr(> z)
Intercept	3.6229	0.5810	6.236	8.91e-08
TypeB	-9.3625	0.9057	-10.338	1.12e-13
Half2	0.1618	0.5487	0.295	0.768
TypeB:Half2	-1.8075	0.7872	-2.296	0.022

Table 3.12: Lmer with final contour (scaled in ST) as dependent variable. R code that produced this model: $FinalFallST \sim type * half + (1 + type | participant)$

Figure 3.9 reveals that this significance occurs due to participants increasing the F0 frequency range of the final rise of Y/N questions in the second half of the experiment (the final contour change is significantly steeper in comparison to the first half). Additionally, gender was pruned from this analysis after revealing that the scaling was successful; there was no significant difference between male and female final F0 contour change (model comparison $p=0.4203$).

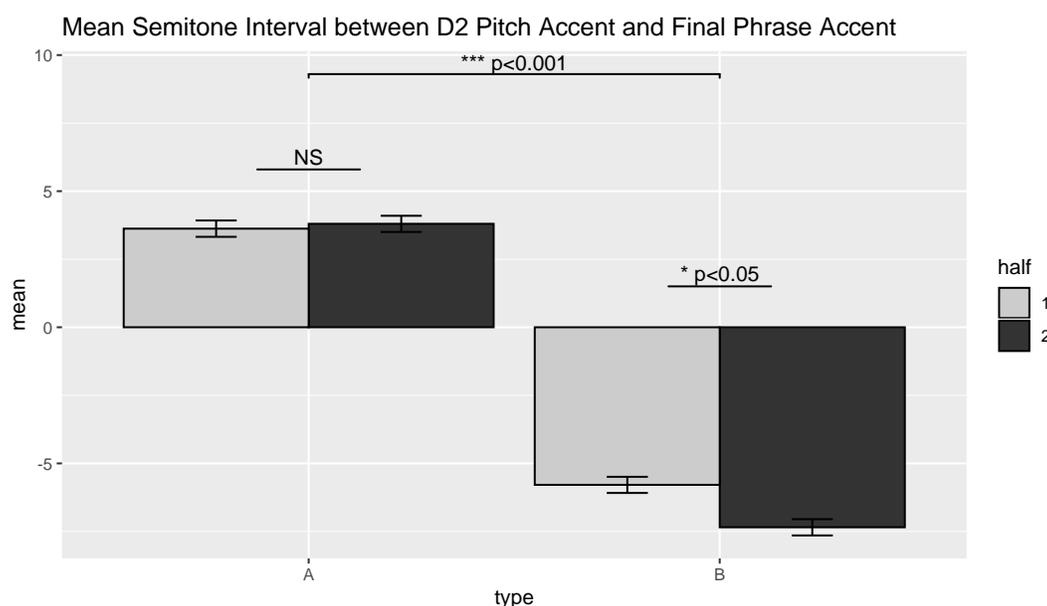


Figure 3.9: Average F0 change (in semitones) of the Final Fall. The Final Fall was measured from the second disjunct (D2Pitch) to the final phrase accent (D2ipPitch); sorted by both half of experiment (1 vs. 2) and disjunct Type (A (Alt) vs. B (Y/N)).

Thus, these results support the predictions that a final ‘fall’ (high D2 pitch, low D2 phrasal accent) occurs for alternative questions, while a final ‘rise’ (low D2 pitch, high D2 phrasal accent) occurs for Y/N questions. In fact, the dichotomy between the final phrase accent pitch is so large that most final phrase accent pitches are higher for Y/N questions than alternative questions, *regardless of gender*. That is, even though men have lower pitch ranges

than women, their Y/N questions still had higher phrase accent pitches than women's alternative question phrase accent pitch.

A linear discriminant analysis (LDA) was performed in R using the `lda` function in the MAAS package (Venables and Ripley, 2002). LDA is a standard pattern recognition technique often used with acoustic variables to classify cases (Adank & Van Hout, 2004). The final contour (the pitch change from the second disjunct to the final phrase accent, scaled to semitones) was the independent variable. The LDA uses this variable to demarcate each token into one of the two interpretations. Table 3.13 includes an analysis over both the entire dataset and only correct answers; the analyses reveal that the final contour *alone* accurately delineates the two interpretation types over 96% of the time (86% over the entire dataset). This indicates a robust perceptual effect, but not a sufficient one for 100% interpretational accuracy.

Data Set	Type A % Correct	Type B % Correct	Total % Correct
All Data (incl. errors)	89.4%	82.4%	86.0%
Correct Only	97.9%	95.1%	96.7%

Table 3.13: Linear Discriminant Analysis based on change from *D2Pitch* to *D2ipPitch* (in semitones). R code: `type~finalfallST`

3.4.5.4 Summary

Thus, the semantic theories accurately predict that utterance-final contours will vary depending on the speaker's intended question type. Additionally, the LDA analyses show that this final contour is necessary to the correct semantic interpretation, but not sufficient in determining the correct interpretation 100% of the time. Table 3.14 (next page) shows the theories examined, their predictions, and the matching observations. Overall, the assumption of

a rising contour for Y/N questions and falling contour for Alt. questions holds, although the final contour produced did not vary monotonically with question type.

Theory	Alternative Question			Y/N question		
	Prediction	Observation	Conclusion	Prediction	Observation	Conclusion
Previous empirical work	Final Fall	Final Fall	✓	Final Rise	Final Rise	✓
Inquisitive Semantics	Final Fall		✓	Final Rise		✓
Commitment Space Discourse	Final Fall		✓	Final Rise		✓

Table 3.14: Predictions and Observations for the Final Contour of Disjunctive Questions

3.4.6 Prosodic Focus

It is known that the prosodic phrasing of an utterance affects the interpretation of focus (Schafer 1997), and that focus is often conveyed via pitch accent (Cutler 1976). While the empirical research suggests that presence/absence of focus on each disjunct for Y/N questions is unimportant, it does predict prosodic focus on both disjuncts for alternative questions. Due to the nature of the experiment, all participants had a pitch accent on every animal, regardless of question type²⁷. For the second disjunct, this is understandable, as most participants had the disjunct at the end of the sentence and so the nuclear stress of the sentence occurred on this disjunct. Only 13 productions did not have the second disjunct utterance-final; however on these disjuncts a prenuclear pitch accent still occurred. The type of pitch accent varied, whether the

²⁷ This includes errors, excepting instructional errors like “cut-off” or “one”, in which the second disjunct was absent altogether or was cut-off preceding the pitch accent.

accent was a prenuclear pitch accent or nuclear pitch accent. Table 3.15 reveals the breakdown of pitch accents on the second disjunct.

Disjunct 2 Accent	!H*	H*	L*	L*+H
Alt Question	2.38% (8)	41.76% (140, 3 prenuclear)	49.70% (167, 2 prenuclear)	6.25% (21)
Y/N Question	1.13% (3, 2 prenuclear)	9.40% (25, 1 prenuclear)	86.84% (231, 3 prenuclear)	2.63% (7, 2 prenuclear)

Table 3.15: Occurrence of pitch accent types on the second disjunct, by question type (excluding errors). Percentages are percent of pitch accent as part of total for that question type. Token counts are included in parentheses.

Overall, the Y/N question interpretation was less variable, with over 86% of the pitch accents occurring as a L*. This is the only pitch accent (of the four different types produced) that ends low; as the final rise of the Y/N tune has been shown to be quite robust, this accent choice makes the most sense. The L* pitch accent also occurred the most on Alt. questions, despite the H* accent being the canonical pitch accent assigned to the second disjunct. This was striking, as the final fall for Alt. questions was also shown in the previous section to be extremely robust. A quick comparison of the L* pitch across question types (Table 3.16, next page) reveals that the average L* pitch for Alt. questions was higher than its Y/N question counterpart²⁸, which would therefore help facilitate that final fall.

²⁸ The accent type (prenuclear vs. nuclear) was combined in the table because prenuclear accents accounted for approximately 1% of the data (5 of 398 tokens); it was true that of these five tokens (2 Alt. and 3 Y/N), the Y/N pitch was higher than the Alt pitch, which is the opposite of what occurs in nuclear accent position. This is further evidence that perhaps the realized L* in nuclear position is lower for Y/N than Alt. questions specifically to facilitate the final contour. However, with only 5 tokens (produced by two females) with prenuclear pitch accents, this interesting trend cannot be supported without more data.

	Male	Female
Alt. questions	2.587	10.254
Y/N questions	2.041	9.098

Table 3.16: Average F0 (in semitones, base 100) across L* productions on the second disjunct

Many of the pitch accents on the first disjunct for Y/N questions were prenuclear accents, appearing for phonological reasons (e.g. discourse-new nouns, Frazier et. al. 2006). However, there were many Y/N questions (nearly 30% of total Y/N questions) that indeed had two intonational phrases, with a (nuclear) pitch accent on each disjunct, indicating focus. Table 3.17 shows the breakdown of pitch accents, divided into prenuclear vs. nuclear classification, across all correct productions. The presence of nuclear accents on the first disjunct in Y/N questions ties into the next section, concerning phrasal breaks. Furthermore, there were nearly 25% of Alt. questions that had prenuclear pitch accents on the first disjunct, rather than nuclear accents, indicating a lack of focus on the first disjunct. That is, “to produce an utterance with double foci explicitly on two different words, it is necessary to make two intermediate phrases, with an intervening intermediate phrase break marked by a phrase accent” (Beckman 1996, 35).

Accent	Question Type	!H*	H*	L*	L*+H	% of Total Type
Prenuclear Accent	Alt.	1.27% (1)	7.59% (6)	3.8% (3)	87.34% (69)	23.51% (79)
	Y/N	1.06% (2)	5.85% (11)	34.57% (65)	58.51% (110)	70.68% (188)
Nuclear Accent	Alt.	0.39% (1)	17.12% (44)	77.82% (200)	4.67% (12)	76.49% (257)
	Y/N	0% (0)	15.38% (12)	79.49% (69)	5.13% (4)	29.32% (78)

Table 3.17: The percent of first disjunct pitch accent distribution across question type, divided into prenuclear vs. nuclear accent. The total column indicates the % of total productions of Alt. vs. Y/N questions. Token counts are included in parentheses.

These Alt. questions did not have a phrasal break between the disjuncts. The implications of Alt. questions not having a phrasal break will be discussed in the next section.

When the first disjunct had a nuclear accent, it was most often realized as a L* across both question types. However, when the first disjunct was prenuclear in nature, it was most often realized as a L*+H across both question types. For Alt. questions, it is possible that this accent may have been intended to be a L*H-, as it can be difficult to differentiate an intermediate phrase when two accents are close together (Beckman 1996). While this accounts for most of the prenuclear accents, it does not account for all of them, and thus introduces the possibility that an Alt. question can be produced without focus on both disjuncts.

The choice of L*+H is also an interesting one, as this accent type has been associated with speaker uncertainty (Ward & Hirschberg 1985). However, Pierrehumbert & Hirschberg (1990) argue that speakers choose L*+H to “convey [a] lack of predication and to evoke a scale. Together these can convey the impression of lack of speaker commitment” (296). The interrogative nature of questions may account for this lack of speaker commitment: the speaker is requesting commitment by the hearer. Furthermore, the scale evoked is one of a set; both questions introduce two choices, while the Alt. question additionally stipulates ‘exactly one’.

Regardless, none of the theories specify the types of pitch accents that must occur, or must not occur, for each question type. The previous experimental results which stated that the first disjunct of an alternative question must end high will be addressed in the next section, phrasal breaks. Table 3.18 (next page) lists the theories and their predictions of focus placement on the disjuncts. Most of the Alt. productions had focus on both disjuncts; however, nearly 25% of them did not. This result is not supported by any theory, which stipulates that focus on both

disjuncts is necessary. It certainly remains typical to have focus on both disjuncts, but it does not appear *necessary* for an Alt. interpretation.

Similarly, most Y/N productions did not have focus on both disjuncts; however, nearly 30% of them did. Previous experimental work supported the conclusion that focus structure neither aided nor prevented an Y/N interpretation. Furthermore, the addition of prosodic focus to both disjuncts may be a result of the speaker attempting to prioritize the pertinent information; the presence of an accent leads the listener to give primary attention to the acoustic/phonetic properties of the word (Terken & Nootboom 1987). That is, observing more accents than predicted has happened before (c.f. Pierrehumbert 1994, German et. al. 2006). Accented words may also be recognized faster than unaccented words (Cutler 1976).

Theory	Alternative Question			Y/N question		
	Prediction	Observation	Conclusion	Prediction	Observation	Conclusion
Previous empirical work	Both disjuncts	Focus occurred on whole phrase or	X	Whole phrase or both disjuncts	Focus occurred on whole phrase or	✓
Inquisitive Semantics	Both disjuncts	both disjuncts	X	Whole phrase or both disjuncts	both disjuncts	✓*
Commitment Space Discourse	Both disjuncts		X	Whole phrase		X

Table 3.18: Theories and predictions on prosodic focus occurrence

Over question type, in comparison to observations

	Focus	Licit Responses		
		Yes	No	A disjunct
Inquisitive Semantics	Both disjuncts	X	✓	✓
	Whole phrase	✓	✓	✓
Commitment Space Discourse	Both disjuncts	X	X	✓
	Whole phrase	✓	✓	X

Table 3.19: Treatment of Y/N questions under IS and CSD

While both Inquisitive Semantics and Commitment Space Discourse allow Y/N questions to have focus on both disjuncts, these types of Y/N questions also simultaneously do not allow *yes*, or *yes* and *no*, respectively, as answers. It is possible that participants disregarded the responses in the experiment when forming the tunes to their questions; however, based on the fact that participants did seem to learn which question type contour to use as the experiment went on suggests that participants were attuned to the responses and how the responses impacted the type of tune they may use; therefore, neither theory predicts the experimental data, although the data is consistent with the previous experimental research (Heidenreich 2014b).

3.4.7 Phrasal Breaks

Previous experimental results differ from both Inquisitive Semantics and Commitment Space Discourse in predicting phrasal breaks between disjuncts for each question type. Again, both Inquisitive Semantics and Commitment Space Discourse account for Y/N questions that do contain phrasal breaks; however, these questions do not allow all the responses (*yes* and *no*) that participants saw in the experiment. Both theories, however, do not account for an alternative question that does not contain a phrasal break (Table 3.20, next page).

Theory	Alternative Question		Y/N question	
	Prediction	Observation	Prediction	Observation
Previous empirical work	Ends H (phrasal break not necessary)		Yes or no	
Inquisitive Semantics	Phrasal Break		Yes or no ²⁹	
Commitment Space Discourse	Phrasal Break		No	

Table 3.20: Predictions of phrasal breaks across theories by question type

Table 3.21 shows the breakdown of phrasal breaks in the production data, both including and excluding errors (in case errors are accounting for the discrepancy between observed phrasal break patterns and predicted phrasal break patterns).

Question	Presence of Phrasal Break		Absence of Phrasal Break	
	Correct only	Total	Correct only	Total
Alternative	257	274	79	98
Y/N	78	130	188	242

Table 3.21: Phrasal Break patterns for both types of disjunctive questions

The table indicates that while it is more common to not have a phrasal break for a Y/N question, there were many Y/N productions with a phrasal break between disjuncts (about 30% of correct Y/N productions had a phrasal break); similarly, while it is more common to have a phrasal break for an alternative question, there were many alternative productions without a phrasal break (about 23.5% of correct alternative productions did not have a phrasal break).

²⁹ A *no* answer is licensed for either one or two lists, whereas a *yes* answer is only licensed for a single list (no phrasal break).

There is a trend, however, to produce most alternative questions with a phrase break, while most Y/N question are produced without.

To investigate further, a logistic-mixed effects regression was performed, where “number of lists” (1= no phrase break, 2= phrase break between disjuncts) was the dependent factor, and the independent factor was type, as well as gender and half (to check the other factors that were significant or close to significant in the other models). Random intercepts for participants and random slopes for participant by disjunct type were also included as random effects in the model³⁰. Both type ($\beta = -2.35$, $SE = 0.355$, $z = -6.61$, $p < 0.0001$) and gender ($\beta = -1.50$, $SE = 0.512$, $z = -2.94$, $p < 0.01$) were significant; half was not significant and was discarded from the model ($p=0.445$). Additionally, an interaction between gender and type was tested and found not to be significant ($p=0.3504$). Table 3.22 details the output of the best fit logistic regression model.

	Estimate	Std. Error	z value	Pr(> z)
Intercept	2.2590	0.4105	5.504	3.72e-08
TypeB	-2.3481	0.3552	-6.610	3.84e-12
GenderM	-1.5037	0.5123	-2.935	0.00333

Table 3.22: Logistic model with number of lists as dependent factor. Type and gender are independent factors. R code to produce model: `listNo~type+gender+(1+type|pID)`

Figures 3.10 and 3.11 show the counts of recordings broken into number of phrases. The graphs reveal the significant difference ($\beta = -2.35$, $SE = 0.355$, $z = -6.61$, $p < 0.0001$) between disjunct Type, as well as the significant difference ($\beta = -1.50$, $SE = 0.512$, $z = -2.94$, $p < 0.01$) between male and female productions. The model estimates and graphs reveal that Y/N

³⁰ This model was the best fit: a model with random intercepts for items was not significantly better ($p=0.6419$) and a model with random intercepts and slopes did not converge.

questions are more likely than Alt questions to have one list vs. two (negative estimate for type B) and men had more utterances in a single list format across both question types than women (negative estimate for gender M). It can be said that, *canonically speaking*, alternative questions have a phrase break while Y/N questions do not; however, based on the results in Table 3.22 and Figures 3.10 and 3.11, it seems that both alternative questions and Y/N questions *can* be produced with or without phrasal breaks.

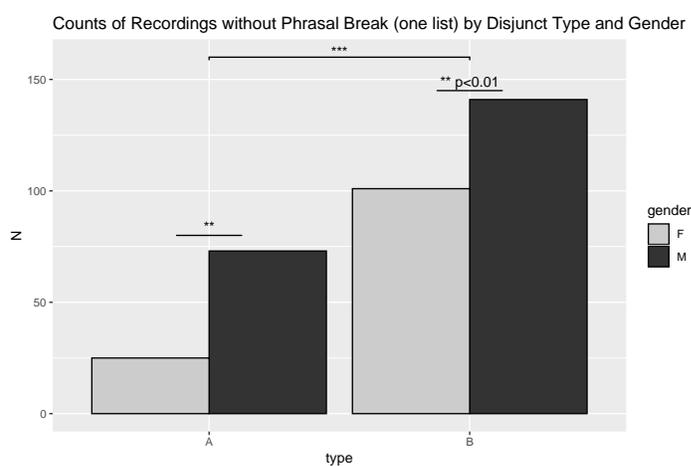


Figure 3.10: Counts of Recordings without a phrasal break, by disjunct type and gender.

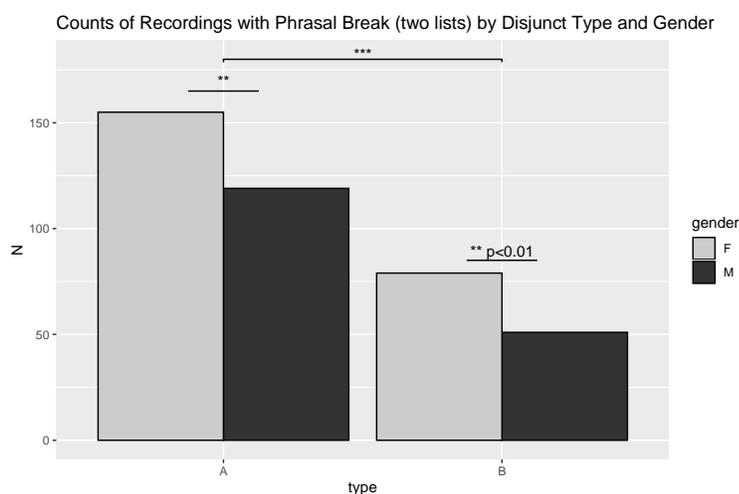


Figure 3.11: Counts of Recordings with a phrasal break, by disjunct type and gender

These results can be supported with phonetic data: phrasal boundaries have been known to exhibit durational differences, such that the existence of a phrasal boundary will result in a longer production of the preceding word (Lehiste 1973). Therefore, the duration of the first disjunct can be used as a dependent variable in a linear mixed-effects regression model that examines whether this duration is significantly determined by disjunct type. The model of best fit (Table 3.23) revealed that disjunct type is significant ($\beta = -0.048$, $SE = 0.008$, $z = -6.09$, $p < 0.0001$), with a negative estimate that indicates Y/N questions are produced with a shorter duration of the first disjunct than their Alt question counterparts (Figure 3.12).

	Estimate	Std. Error	z value	Pr(> z)
Intercept	0.6071	0.0296	20.502	<2e-16
TypeB	-0.0475	0.0078	-6.092	1.9e-09

Table 3.23: Linear model with duration of first disjunct as dependent factor. Type is an independent factor. R code to produce model: `D1length~type+(1|pID)+(1+gender|number)`

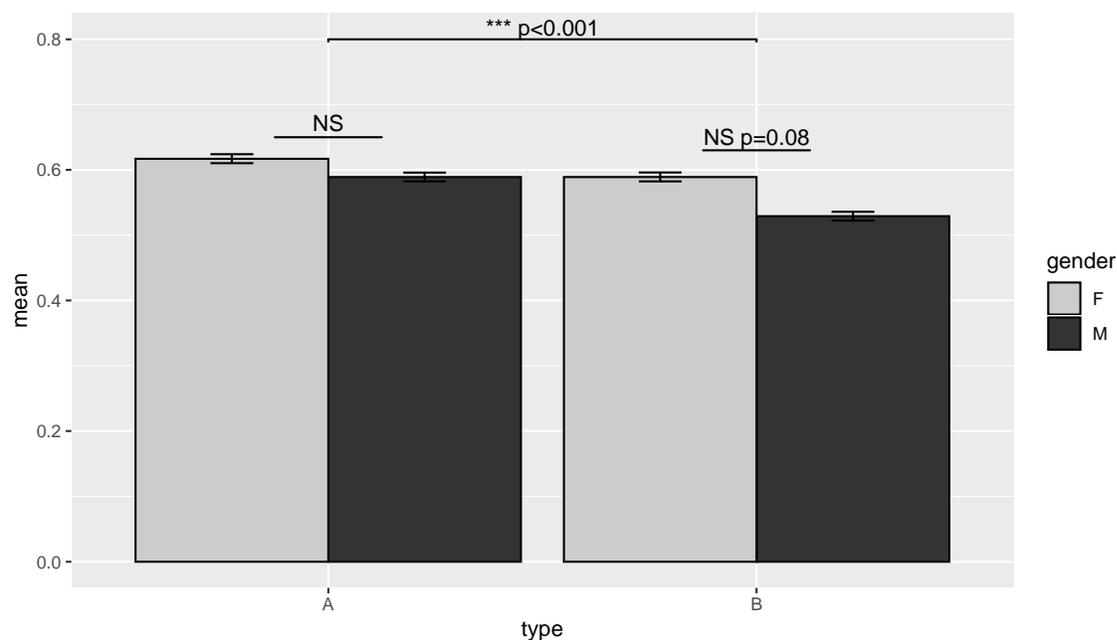


Figure 3.12: Mean length of first disjunct (D1length) by type and gender. Disjunct Type was significant in the model, while gender was not. The model that had the interaction between gender and disjunct type revealed near significance ($p=0.08$) for the interaction.

Gender is no longer a fixed effect: rather the best fit model had a random effect of random slopes of items by gender. A model that included gender as a fixed effect with an interaction with type was nearly significantly better ($p=0.07$). Nevertheless, the first disjunct duration model supports the “list number” model in that the first disjunct in a Y/N question often does not have a phrase break (realized by a shorter duration of the first disjunct), while the first disjunct in an Alt question often does have a phrase break (realized by a longer duration of the first disjunct). The “list number” model also reveals that both types of disjunctive questions can occur with and without phrasal breaks. Table 3.20 has been updated (below) to reflect the observations.

Theory	Alternative Question			Y/N question		
	Prediction	Observation	Conclusion	Prediction	Observation	Conclusion
Previous empirical work	Ends H (phrasal break not necessary)	Yes or no	?	Yes or no	Yes or no	✓
Inquisitive Semantics	Phrasal Break		X	Yes or no		✓*
Commitment Space Discourse	Phrasal Break		X	No		✓*

Table 3.20 (updated): Predictions and observations of phrasal breaks. Table is updated with conclusions across theories by question type.

Both Inquisitive Semantics and Commitment Space Discourse have allowances for Y/N with phrasal breaks; therefore the observations coincide with the predictions *if we assume that participants did not take Jordi's responses into account*, as both theories discount *yes* as a plausible answer for this type of Y/N production (and Commitment Space Discourse discounts *no* as well). However, neither theory can account for an alternative question without a phrasal

break; Inquisitive Semantics does not account for closed interrogatives with a single list item when that item contains a disjunction (single list items are defined as something like “Do you want to see the giraffe?”, i.e. no disjunction). Commitment Space Discourse defines alternatives as two monopolar questions combined; this combination allows for the licit responses of each disjunct rather than a *yes* or *no* answer. There is no methodology for accounting for an alternative question that would be a ‘single question’ (i.e. without a phrase break). Only previous empirical research had suggested that the phrase break itself may not be vital for alternative question interpretation (Heidenreich 2014b).

That same study, however, suggested that alternative questions may need the first disjunct to end with a high pitch, be it a pitch accent or phrasal accent. Table 3.24 shows the breakdown of utterances without a phrasal break.

Question Type	Pitch accent on first disjunct			
	!H*	H*	L*+H	L*
Alternative	3	7	79	9
Y/N	2	19	143	78

Table 3.24: Breakdown of first disjunct pitch accents without a phrasal break

Of the 98 occurrences of alternative questions without a phrasal break, only nine of them did not have a high pitch accent or end with a high pitch on the first disjunct. Even further, of these nine productions, six were produced like canonical Y/N questions and were thus considered errors. The three remaining productions were comprised of a wide scope disjunctive question (61), an incomplete phrase (62), and a mix between a disjunctive question and wh-word question (63).

(61) Would you like to see the hornbill or would you like to see the gorilla?

(62) How about a musk ox or a seal?

(63) If you could choose a muskrat or a camel, which would you choose?

Utterance (63) seems like it falls into a different type of disjunctive question type, while utterance (62) seems to require additional context as an incomplete phrase. Only the wide-scope disjunctive question would seem to fit into the current context of disjunctive questions. Both Inquisitive Semantics and Commitment Space Discourse treat wide-scope and narrow-scope disjunction in the same way (they both are confined to the same allowable tunes and answerhood conditions); however it is possible that wide scope disjunctive questions may have different conditions necessary in order to be categorized as alternative questions. Previous experimental studies (Pruitt & Roelofsen 2013, Heidenreich 2014b), as well as this one, contain nearly all narrow-scope disjunctive questions. In any regard, the three utterances described above account for less than 1% of the correct alternative question utterances, which supports the notion that alternative questions ought to have the first disjunct end with a high pitch (comparative to speaker range). Table 3.20 (reprinted below) is updated to reflect this final observation.

Theory	Alternative Question			Y/N question		
	Prediction	Observation	Conclusion	Prediction	Observation	Conclusion
Previous empirical work	Ends H (phrasal break not necessary)	Yes or no, but first disjunct ends high	✓	Yes or no	Yes or no	✓
Inquisitive Semantics	Phrasal Break		X			Yes or no
Commitment Space Discourse	Phrasal Break		X	No		✓*

Table 3.20 (updated): Predictions and observations of phrasal breaks. The table is updated across theories by question type, including whether the observation matched the prediction.

3.5 Discussion

Participants were successful in producing the intended tune, based on image context and previously seen licit responses. Rather than becoming fatigued during the experiment, participants performed better as the experiment progressed, indicating that the ‘responses’ to their recordings were taken into account for future recordings of similar image context. It is true that participants often produced the assumed typical representations of the respective question types: for Alt. questions, this meant focus on both disjuncts, with a phrase break in between, and a final fall. For Y/N questions, it was broad focus on the utterance (no phrase break) with a final rise. However, the data showed that there are other, equally licit productions that do not follow these exact patterns.

The production data supports various aspects of Inquisitive Semantics and Commitment Space Discourse. Alternative questions end with a final fall, which is represented typically by a higher pitch on the second disjunct and a low phrasal accent after the disjunct. Y/N questions typically have a final rise, represented by a lower pitch on the second disjunct and a high phrasal accent. The recordings were elicited by image context alone, which resulted in many disjunct-final utterances, thereby conflating final phrasal accent (ip) and the intonational phrase boundary (IP). The phrasal accent was primarily used in statistical analysis due to its lower deviance in the models. A linear discriminant analysis on this final contour revealed that the final contour alone could correctly classify 86% of the data (and over 96% of the correct data), revealing a robust effect. Thus both theories, as well as previous experimental research (Pruitt & Roelofsen 2013, Heidenreich 2014b), correctly predicted the final contour as being significant in the disambiguation of alternative questions and Y/N questions (Table 3.25, next page).

		Predictions			Observations	Conclusion
	Question Type	Inquisitive Semantics	Commitment Space Discourse	Previous Empirical Results	Current Study	Theories supported
Final Contour	Alt	Final Fall	Final Fall	Final Fall	Final Fall	All
	Y/N	Final Rise	Final Rise	Final Rise	Final Rise	All
Focus	Alt	Both disjuncts	Both disjuncts	Both disjuncts	Both disjuncts or whole phrase	None
	Y/N	Both disjuncts or whole phrase	Both disjuncts ₁ or whole phrase	Both disjuncts ₁ or whole phrase	Both disjuncts or whole phrase	All ₁
Phrase Break	Alt	Phrasal Break	Phrasal Break	Yes or no, D1 ends high	Yes or no, D1 ends high	Previous results only
	Y/N	Yes ₂ or no	Yes ₂ or no	Yes or no	Yes or no	All ₂

Table 3.25: Theory predictions, observations, and conclusions from the production data.

*1*The theory supports focus on both disjuncts, but then does not license the responses seen in the experiment. Participants would have had to be disregarding responses.

*2*The theory supports phrase breaks in Y/N questions, but then does not license the responses seen in the experiment. Participants would have had to be disregarding responses.

The focus and phrasal break predictions, however, did not follow the prediction of either theory. Utterances had pitch accents on all the disjuncts present, though they were divided between nuclear pitch accents and prenuclear pitch accents. However, both Commitment Space Discourse and Inquisitive Semantics only posit focus on both disjuncts when the Y/N question does not allow *yes* or *no* (in the case of Commitment Space Discourse) or *yes* (in the case of Inquisitive Semantics) as a plausible response. Statistical analysis revealed that participants did

improve as the experiment went on; however, this focus structure did not change. Either participants did not alter their focus structure in regard to the type of responses they saw during the experiment, or these types of focus structures must be allowed for Y/N questions that permit *yes* and *no* as answers.

Furthermore, there was no prediction, either empirical or theoretical, that an Alt. question might not have focus on both disjuncts (no phrasal break), although this type of production occurred nearly 25% of the time; however, the first disjunct still had a prenuclear accent. Both theories also predicted the absence of a phrasal break between disjuncts for Y/N questions, based on the response seen; however, both theories do account for Y/N questions with a phrasal break. Again, we must infer that either participants did not alter their phrasing intonation based on seen responses, or that the theories must permit these answers with a Y/N questions of “two lists”. Neither theory, however, accounts for an alternative question without a phrase break, although this occurred in over 26% of the data (23.5% of correct alternative questions). Therefore, neither theory accurately models the distribution of alternative questions and Y/N questions based on the presumed significant factors: final contour, focus presence on the disjuncts, and phrase break between disjuncts. In fact, ending pitch of the first disjunct, regardless of whether there was a phrasal accent or not, appears to be significant in disambiguating, but neither theory accounts for this.

In order to gauge whether these factors (and only these factors) contribute to the delineating of alternative questions and Y/N questions, two logistic mixed-effect models were designed, with “type” as the dependent variable. The idea was to determine which of the observed dependent factors were influenced by the disjunct type of the question; if a factor proved significant in a reverse analysis (where the independent factor of “type” becomes the

dependent factor, and the various dependent factors become independent factors), then we might conclude that participants were manipulating this factor based on the context provided. The first model had intonational data (pitch accents, phrasal accents, IP boundary) as the independent variables, while the second model had phonetic data (D1pitch, D1length, D1ipPitch, etc.) as the independent variables. The length of the first disjunct, in particular, was included in the second model due to the fact that phrasal boundaries have been known to exhibit durational differences (Lehiste 1973) whereby a longer duration for a first disjunct (within the same item) would indicate a phrasal break, whereas a shorter duration would indicate the absence of a break. All the measurements from the utterances were used as independent factors to determine which factors, exactly, contributed to one interpretation over the other. Both models were performed with all errors included in the 31 participants. After slimming the models (by eliminating non-significant factor each iteration), the best-fit model for the intonational data included the significant factors of the second disjunct phrasal accent (D2ip), the intonational phrase boundary tone (IP), the first disjunct phrasal accent (D1ip) and one level of the first disjunct accent type (D1accent). Table 3.26 lists the significant factors and their estimates, errors, etc. of the best fit model. Using the correct data in a linear discriminant analysis with these factors as the independent variables, the model was able to correctly classify the question type by these factors alone over 98% of the time.

Interestingly, the model found both the phrasal accent after the second disjunct (D2ip) and the intonational phrase boundary (IP) significant; the best fit model contained both of these and was significantly better than a model with only one of these factors (Figures 3.13 and 3.14). Both of these factors contained two levels, low (L) and high (H). Both D2ip and IP had negative estimates (D2ip: $\beta = -2.23$, $SE = 0.54$, $z = -4.49$, $p < 0.0001$; IP: $\beta = -2.13$, $SE = 0.54$, $z = -3.97$,

$p < 0.0001$), indicating that both L- and L% occurred more often with the Alt question contexts than the Y/N question contexts. The first disjunct phrasal accent (D1ip) was significant when it did not occur (D1ipNA) in comparison to when it ended low (L-)³¹ or high (H-). The positive estimate of D1ipNA ($\beta = 1.38$, $SE = 0.41$, $z = 3.35$, $p < 0.001$) indicates that it occurred significantly more often with the Y/N question contexts than Alt question contexts, c.f. Figure 3.15; thus, while a phrasal break doesn't seem necessary for an alternative question, it may contribute to swaying the interpretation in that direction.

	Estimate	Std. Error	z value	Pr(> z)
Intercept	-1.5512	1.2182	-1.273	0.202888
D2ip L-	-2.4290	0.5405	-4.494	6.99e-06
IP L%	-2.1294	0.5366	-3.969	7.23e-05
D1ip L-	16.6361	362.0453	0.046	0.963350
D1ip NA	1.3751	0.4100	3.354	0.000798
D1accent !H*	-3.61426	1.21846	-2.966	0.003015
D1accent H*	0.06416	0.40627	0.158	0.874515
D1accent L*+H	-0.56185	0.41351	-1.359	0.174226

Table 3.26: Logistic mixed effect model with dependent variable as question type. Independent variables are phonological values. R code for model: $\text{type} \sim \text{D2ip} + \text{IP} + \text{D1ip} + \text{D1accent} + (1 / \text{participant})$

The last significant factor was the accent on the first disjunct (Figure 3.16). The baseline level for this factor was L*, as the remaining accent types (!H*, H*, and L*-H) all end high in comparison. Only !H* was significant in the analysis³² ($\beta = -3.61$, $SE = 1.22$, $z = -2.97$, $p <$

³¹ Only four instances of L- at D1ip occurred, all within Y/N disjunct type; removing these from the model did not change the significance of the other variables.

³² Removing the six productions with !H* as the first accent type did eliminate the significance of the D1accent factor; comparing models with and without this factor (with !H* removed) revealed $p=0.3475$, indicating the more complex model was not significantly better. However, there is no clear motivation on why !H* ought to be left out of the analysis, other than its low

0.01), possibly because of the few number of !H* (token count: 6) observed. A !H* was more likely to occur in an Alt question, while the remaining three accent types appeared at about the same frequency across disjunctive question type.

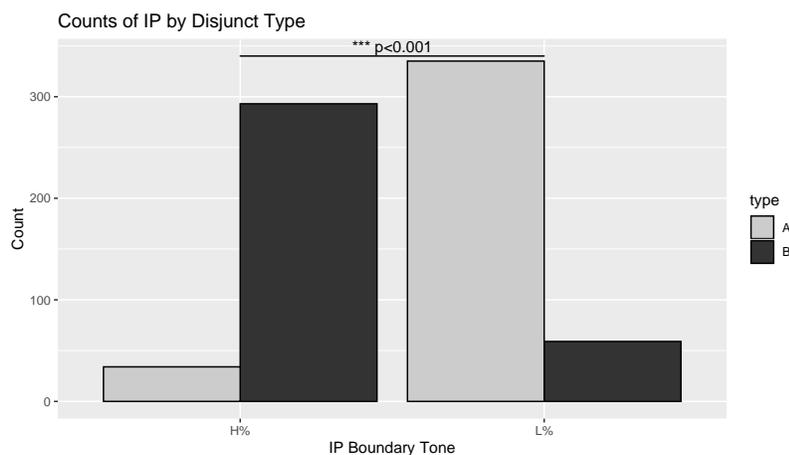


Figure 3.13: Counts of boundary tones (“IP Boundary Tone”), separated by disjunct type.

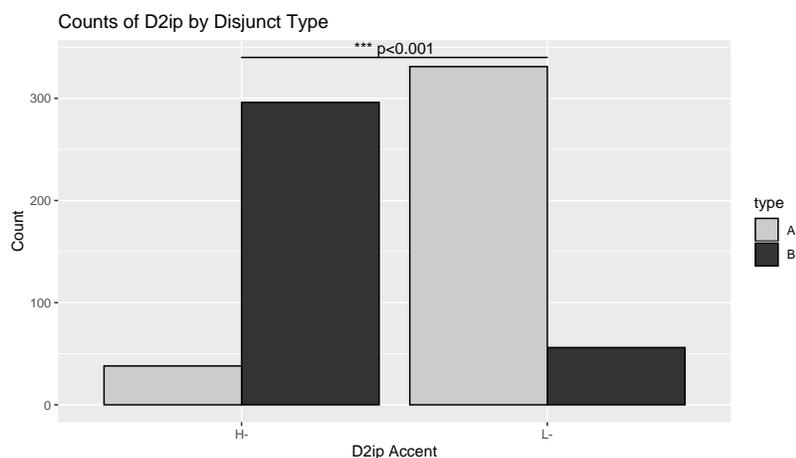


Figure 3.14: Counts of final phrase accent (“D2ip Accent”), separated by disjunct type.

frequency. Indeed, using !H* as the baseline example resulted in the remaining three levels all showing significant difference ($p < 0.05$) from it.

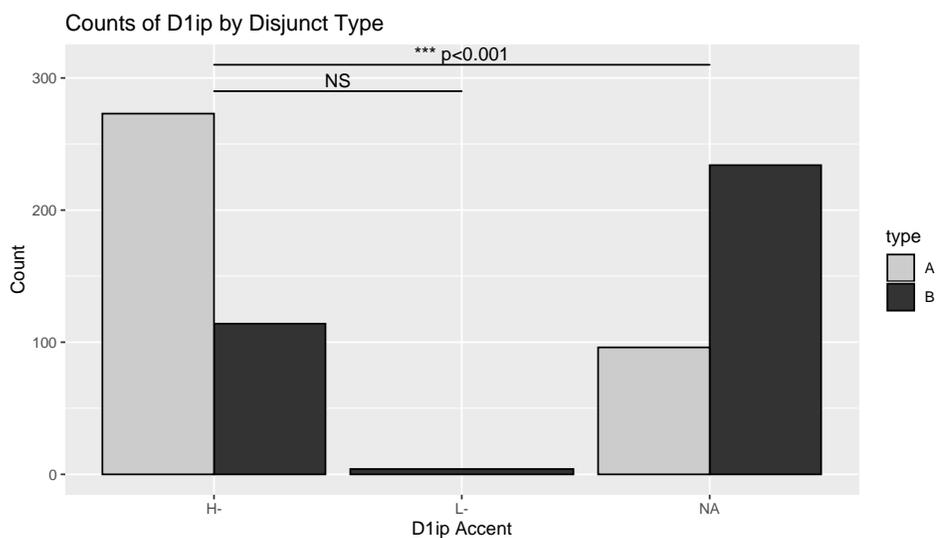


Figure 3.15: Counts of phrase break between disjuncts (“D1ip Accent”), separated by disjunct type. Four instances of a low phrase break (L-) between disjuncts occurred in Y/N questions; this level was not significant. There was a significant difference between a H- and the absence of any phrase break.

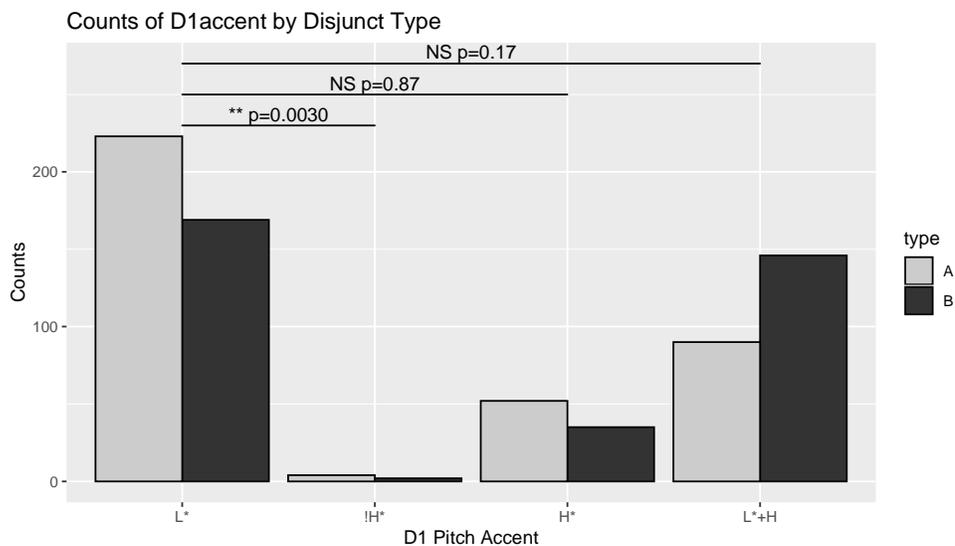


Figure 3.16: Counts of accent on first disjunct (“D1 Pitch Accent”), separated by disjunct type. The only significant level for D1accent was !H*.

The second model, using (scaled)³³ phonetic data from the productions, had the pitch of the second disjunct and final phrase accent pitch significant, as well as the pitch at the end of the first disjunct (D1ipPitch)³⁴. As seen in the final contour analysis, the D2 pitch and D2ipPitch both significantly contribute to the final contour of the question; thus it is unsurprising that these two factors were significant (Table 3.27 and Figures 3.17 and 3.18).

	Estimate	Std. Error	z value	Pr(> z)
Intercept	0.28835	0.31071	0.928	0.35338
D2pitch	-0.08991	0.03149	-2.855	0.00431
D1ipPitch	-0.18409	0.02914	-6.317	2.66e-10
D2ipPitch	0.030023	0.02354	12.755	<2e-16

Table 3.27: Logistic mixed effect model scaled in ST (base 100 Hz). Model had dependent variable as question type and independent variables were phonetic measurements. R code for model: `type~D2pitch+D1ipPitch+D2ipPitch+(1|PID)`

Additionally, the pitch at the end of the first disjunct was also significant; alternative questions not only end high on the first disjunct, but this pitch height is higher than Y/N questions' first disjunct (Figure 3.19).

³³ The model scaled each pitch point by a base of 100. When raw pitch values were used, the log length of the first disjunct was also significant in the analysis ($p < 0.05$); however, using the scaled pitch values reduced the significance of this factor to $p = 0.069$, just below significance. It is probable that the length of the first disjunct correlated with an effect of gender, which was then accounted for when the pitch data was scaled. Recall in the analysis of phrasal breaks that the length of the first disjunct was modeled by random slopes for gender by item, as well as a near significant interaction between gender and disjunct type (with men producing shorter words). This was correlated with the fact that in the list number analysis, men also produced fewer phrasal breaks, overall, than women.

³⁴ The variable is the pitch at which a phrasal accent would occur, but doesn't necessarily occur; thus there is a pitch reference for each recording, not just those with a phrasal accent after the first disjunct.

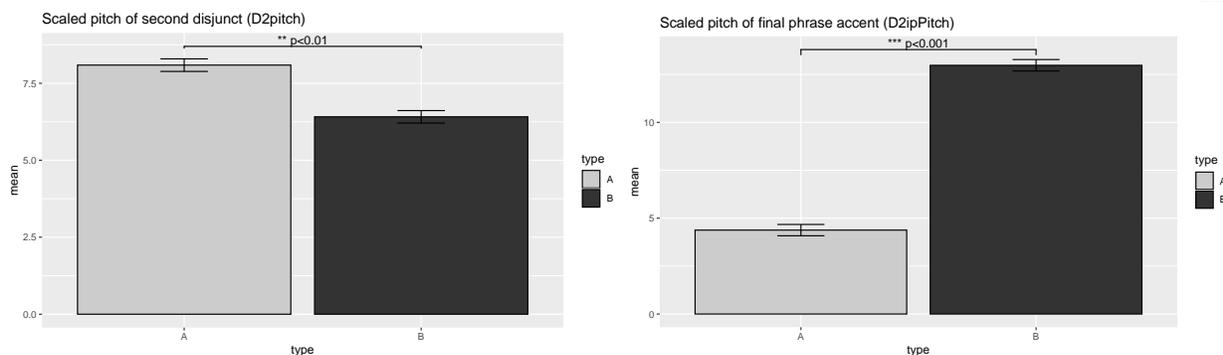


Figure 3.17 and 3.18: Mean scaled pitch of second disjunct (“D2pitch”, left) and final phrase accent (“D2ipPitch”, right), grouped by disjunct type. Alt questions have a significantly higher pitch on the second disjunct, but a significantly lower pitch at the final phrase accent.

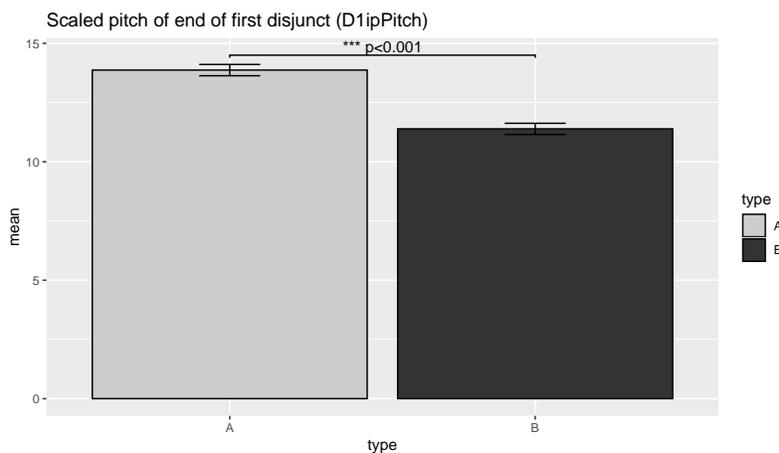


Figure 3.19: Mean scaled pitch at the end of the first disjunct (“D1ipPitch”), grouped by disjunct type. Alt questions had significantly higher pitch values at the end of this first disjunct, in comparison to Y/N questions.

In this model, the pitch of the first disjunct is not significant, though recall the D1 accent was a significant factor in the intonational model (only the !H* level). An ANOVA comparison of both models reveals that the intonational model has a lower deviance (499.59) compared to the phonetic model (581.08), despite four additional degrees of freedom (9 vs. 5) from the D1accent factor. Furthermore, a linear discriminant analysis was performed with the significant

phonetic factors as independent factors, and the resulting model correctly classified 97% of the productions, compared to 98% with the model with intonational factors.

Thus we have a better understanding of the intonational cues that truly distinguish the production of Alt. vs. Y/N questions: the final contour (as delineated by the phrasal accent after the second disjunct), and the pitch between the first and second disjunct (whether this is on a phrasal accent, or simply the pitch accent on the first disjunct). Figure 3.20 shows the average pitch at various points in the utterance, separated by question type and gender.

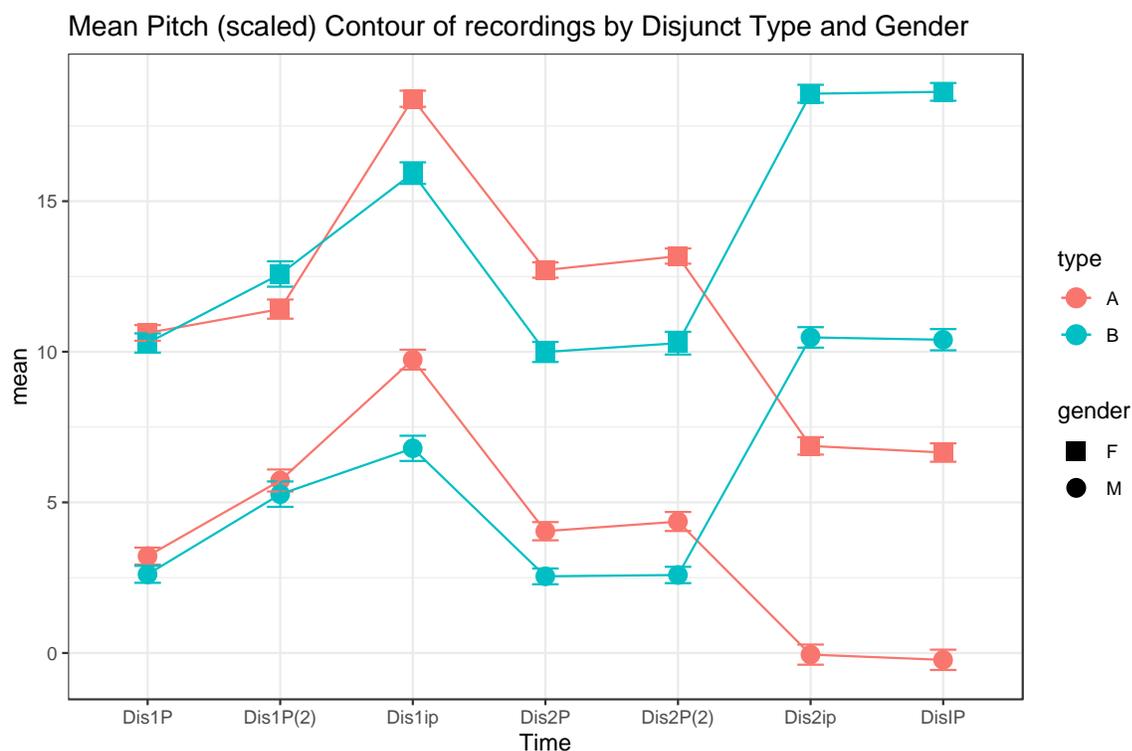


Figure 3.20: Average pitch across full utterance. Average pitch taken at first disjunct (Dis1P), a second point at disjunct 1 if a L^*+H occurred (Dis1P(2)), the end of the first disjunct (Dis1ip), the second disjunct (Dis2P), a second point at disjunct 2 if a L^*+H occurred (Dis2P(2)), the phrasal accent after the second disjunct (Dis2ip) and the boundary tone (DisIP). Disjunct Type is represented by red (Alt) vs. blue (Y/N); Gender is represented by squares (female) vs. circles (male). Note the contours are nearly identical within the same color, indicating that men and women produce the same contours.

Across genders, we see the same contour: a higher pitch at the end of the first disjunct (Dis lip) for alternative questions vs Y/N questions, a higher pitch on the second disjunct for alternative questions than Y/N questions, and then a rise for Y/N questions vs. a fall for alternative questions. The scaled pitch data reveal that the range for male and female speakers is equivalent.

3.6 Conclusion

The production experiment was conducted to test the previous assumptions concerning canonical productions of Alt. questions and Y/N questions. In particular, Alt. questions are assumed to have a final fall, with focus on both disjuncts and a phrase break between disjuncts. Y/N questions are assumed to have a final rise, without focus on the disjuncts and no phrase break between the two of them. The semantics theories that propose to model these types of questions assume these basic intonational cues. However, production data has revealed that, although these cues occur typically, they are not necessary for each question interpretation.

Alternative questions were confirmed to have a robust cut of a final fall, with the final contour being able to successfully disambiguate all data (even errors) over 89% of the time in an LDA. This final fall was a combination of the pitch of the second disjunct and the phrase accent pitch (as most productions had the second disjunct occur utterance finally). Similarly, the Y/N canonical rise was also very robust, with the LDA model successful in classification over 81% of the time. Scaled pitch data revealed that even across genders, Alt. questions end low and Y/N questions end high (female speakers usually end Alt. questions lower than male speakers end Y/N questions). Thus, the assumptions concerning the final contours of the two interpretations were supported.

The remaining assumptions concerning focus and phrase breaks, however, revealed a greater variation in Alt. and Y/N productions than predicted by the semantic theories. While Alt.

questions typically contained a phrase break, with focus on both disjuncts, participants freely produced Alt. questions with a single intermediate phrase (no phrase break, no focus on the first disjunct) nearly a fourth of the time. The semantic theories that claim to model disjunctive questions predicted such a production as illicit. However, previous empirical evidence predicted that it was simply the pitch of the first disjunct that was important in disambiguation (the first disjunct pitch in Alt. questions must be high); in over 99% of the correct Alt. question productions, this proved true. Thus, semantic theories would do better to ensure semantic interpretation relies on that intonational cue, rather than focus or phrasal breaks. The Discussion chapter (Chapter 7) explores how the theories might be altered in order to model this data.

Both theories concede that Y/N questions could be produced with or without phrasal breaks (recall in Inquisitive Semantics, these two interpretations are called ‘Open’ and ‘Block’ structures, respectively; in Commitment Space Discourse, it is a monopolar vs. bipolar interpretation). However, the theories stipulate that the interpretation with a phrase break cannot have *Yes* as a licit response, although participants were given this response in half of the Y/N questions. Although it was shown that participants’ performance improved over the course of the experiment, it still cannot be determined if participants were using the previously seen responses in order to inform their future recordings. The perception experiment (Chapter 5) will offer further insight, when participants must rate Y/N questions with a phrase break against responses like *Yes* or *No*.

Participants were very successful, overall, in producing the intended interpretation, based on image context alone. With the experiment revealing that participant performance improved over time, it might be concluded that a simple disjunct answer ought to be licit for an Alt. question, while *Yes* and *No* ought to be licit for (at least some kind of) Y/N question. The

answerhood conditions for each type of question, however, has not been studied. The next experiment (Text Experiment, Chapter 4) aims to investigate the validity of these assumptions by collecting experimental evidence of what type of response is licit for each question.

Chapter 4: Text Experiment

4.1 Introduction

This experiment was conducted to test several assumptions concerning disjunctive questions: the assumption that the text form of a disjunctive question is ambiguous, or completely unbiased; the assumption that the insertion of *either* before the first disjunct of a disjunctive question rules out an Alt. interpretation, leaving only the Y/N interpretation available; the implicit assumption that any licit answer for an Alt. question (i.e. a disjunct response) is also a licit answer for a Y/N question. While the scope of this experiment is confined to text only (no audio), the data reveal serious implications for these assumptions. In particular, it seems that the text form a disjunctive question contains inherent bias to the Y/N interpretation, even when the type frequency of each type of question is controlled in the experiment. Because of this implicit bias, it is unclear whether *either* rules out an Alt. interpretation; however, it can be stated that *either* does not further bias the text to a Y/N interpretation. Furthermore, cleft answers are significantly worse than simple disjunct responses in the results, indicating at least a less acceptable response in comparison to other answer types (Yes + disjunct, simple disjunct, Yes + both). In general, the idea that responses are either ‘licit’ or ‘illicit’ seems erroneous, as the acceptable answers fell more on a spectrum, or continuous scale, of acceptability. In general, all responses studied seemed acceptable to at least a marginal degree, indicating a degree of flexibility on the interpretation of disjunctive questions.

4.2 Methodology

4.2.1 Design and Materials

The text experiment was designed as a set of question-answer pairs. There were two types of question conditions: a baseline condition (A), and an *either* condition (B):

Question conditions:

- A. Disjunctive question w/ out either. (“Did Travis bring drinks or dessert?”)
- B. Disjunctive questions w/ either. (“Did Travis bring either drinks or dessert?”)

The first question condition (A) is a typical disjunctive question that is “ambiguous” in text- only contexts. Therefore, participants may interpret this as an alternative question or a Y/N question, as only the text will be given (no accompanying audio). The second question condition inserts *either* directly before the first disjunct (B). It is assumed that inserting *either* into a disjunctive question ought to bias it to the Y/N interpretation. Participant interpretation will be determined by the acceptability of the question-answer pair, as some answers are assumed only licit for Y/N questions, others for only alternative questions, and some are permitted for both types. There were six different answer conditions:

Answer Conditions:

1. Disjunct answer (“He brought dessert.”)
2. Yes + disjunct (“Yes, he brought dessert.”)
3. Cleft disjunct (“It was dessert.”)
4. Yes (“Yes.”)
5. No (“No.”)
6. Yes, both (“Yes, he brought both.”)

An example stimulus question with the six answer conditions is given in (64). Only one answer is presented to each participant, and which answer is counterbalanced across lists.

- (64) Did Melanie or Emily go to the party tonight?
- a. Melanie went.
 - b. Yes, Melanie went.
 - c. It was Melanie.
 - d. Yes.
 - e. No.
 - f. Yes, they both went.

In the example above, the disjunct in the question occurred at the start of the utterance. Other questions had the disjunct location occur in the middle (65) or end (66) of the utterance. The disjunct location was counterbalanced across stimuli, so each participant saw 16 instances of all disjunct locations.

(65) *Middle location for disjunct in question*

Did Amanda see Ariel or Miriam at a city council meeting?

(66) *End location for disjunct in question*

Did Travis bring drinks or dessert?

There were twelve question-answer combinations (six for the baseline (64), six for the question with *either*). Thus, there were twelve different lists; each list contained only one question-answer combination per stimuli item. Each participant saw four examples from each question-answer combination. Disjunct location was counterbalanced as much as possible within these factors: each participant saw sixteen questions for each disjunct location, broken down into eight baseline questions and eight questions with *either*. Furthermore, each question-answer

combination had at least one question for each type of disjunct location. Participants only observed one list.

There were also sixteen filler stimuli (the same for each list), for a total of 64 questions. The fillers were divided into infelicitous question-answer pairs (12) and felicitous ones (4)³⁵. The fillers were formed such that they were similar in appearance to the stimuli (some had conjunctions in them), so that the participants might not discern which stimuli were part of the experiment and which were fillers. They were meant to anchor the question-answer pair acceptability within the full range of the spectrum. Appendix A lists all the stimuli.

A quick caveat: if both interpretations are open to the baseline condition (64), then all six responses ought to be acceptable. If only the Y/N interpretation is available in the *either* condition, then at least half of the responses (64a, 64b, 64f) ought to be acceptable (see section 4.3 for more on the predictions that these assumptions make via the semantic theories). Therefore, the stimuli are predicted to motivate a large number of acceptable question-answer pairs. Thus, the fillers had more infelicitous question-answer pairs than felicitous ones, in order to ensure that participants were using the full spectrum of acceptability, as well as to better balance the predicted number of unacceptable vs. acceptable question-answer pairs.

4.2.2 Procedure

Participants were told they would see (text only) a question and an answer to the question; they were to rate the acceptability of the answer to that question on a scale 1-7. A low number was less acceptable; a high number was more acceptable. The participant was exposed

³⁵ Felicitous pairs were further broken down into ‘good’ (or direct answer to the question) pairs (3) vs. one pair that was deemed ‘OK’: “Is Petra going to file her taxes and pay the phone bill?” “Petra is broke”. This may be taken as unacceptable due to the conversational implicature of *no* rather than the outright response of *no*.

to 48 question-answer stimuli pairs, as well as 16 question-answer filler pairs, for a total of 64 questions.

4.2.3 Participants

96 participants completed the experiment on Amazon Mechanical Turk (AMT). The experiment ran from 2-12-18 until 5-18-18. The entire experiment took 5-15 minutes. Participants also filled out a brief survey which asked their age, gender, ethnicity, and native language(s). They were paid \$2 for their participation.

4.3. Theoretical Predictions

Inquisitive Semantics and Commitment Space Discourse offer different answerhood groupings for Alt. and Y/N questions. While both Inquisitive Semantics and Commitment Space Discourse do not offer explicit predictions for cleft disjuncts, we may infer that both theories would allow such an answer for alternative questions; for Inquisitive Semantics, an alternative question highlights two possibilities; in the cleft disjunct answer, the antecedent for *it* is the possibility chosen by the responder (67), since *exactly one* answer must be chosen.

(67) Does Ann[↑] or Bill[↓] play the piano?

- a. $[[\alpha]]_P$ proposes three possibilities: Ann plays, Bill plays, or neither
- b. $[[\alpha]]_H$ highlights two possibilities: Ann plays or Bill plays
- c. $[[\alpha]]_S$ removes the overlap between the highlighting suggestions, resulting in

two distinct possibilities. These are the correct responses (according to IS).

- d. “It’s Ann”, *it* refers to the possibility ‘Ann plays’, one of the two possibilities

Commitment Space Discourse is similar in how it would parse a cleft answer; an alternative question is actually two monopolar propositions, so once again *it* is the antecedent for the

proposition chosen by the responder³⁶. Thus, both theories offer the same predictions for alternative questions (Table 4.1).

Theory	Responses involved in text experiment					
	Disjunct	Yes + disjunct	Cleft disjunct	Yes	No	Yes, both
Inquisitive Semantics	✓		✓			
Commitment Space Discourse	✓		✓			

Table 4.1: Permissible answers to alternative question interpretation, by theory

Table 4.2 marks permitted answers for a Y/N question. Note that both Inquisitive Semantics and Commitment Space Discourse have multiple possible interpretations for a Y/N question (see Chapter 2); all interpretations are explored in Table 4.2. Once again, no explicit predictions are offered for ‘cleft’ answers; however, it may be inferred that a cleft answer would not work for any type of Y/N question, based on the exhaustivity implication that clefts generate when seen as an answer to a question (Pollard & Yasavul 2015, DeVeugh-Geiss et. al. 2018). For the open complex question in Inquisitive Semantics, although more than one possibility is highlighted (indicating more than one proposition, or discourse referent), *both* possibilities could be true; therefore *it* cannot refer to a specific proposition. That is, a question like (68) could accept various other answers to signal ‘only one’, but the signaling is done in the response rather than the question.

(68) Does Ann[↑] or Bill[↑] play the piano?

³⁶ This prediction is valid if Commitment Space Discourse updates alternative questions so that any set of propositions that are in alternative question context (e.g. tune) are in an ‘exactly one’ scenario.

- a. #It's Ann.
- b. Only Ann.
- c. Just Ann plays.

Commitment Space Discourse's two monopolar propositions come to the same conclusion: although there are two propositions available, the question doesn't suppose that *only* one of these should hold, and therefore *it* does not have an appropriate referent. The remaining types of Y/N questions all are a single phrase, and therefore *it* would have to refer to the entire phrase itself, rather than one disjunct in the question, and thus would not be an acceptable answer.

Theory	Responses involved in text experiment					
	Disjunct	Yes + disjunct	Cleft disjunct	Yes	No	Yes, both
Inquisitive Semantics: Block	✓	✓		✓	✓	✓
Inquisitive Semantics: Open	✓				✓	
Commitment Space Discourse: Two monopolar	✓	✓				✓
Commitment Space Discourse: Bipolar				✓	✓	

Table 4.2: Permissible answers to Y/N interpretation, by theory

For both Inquisitive Semantics and Commitment Space Discourse, *No* cannot be used in response to an alternative question (in order to deny some part of the question); rather, a weaker form of negation (e.g. *Actually*) must be used to reject the presumption of the question. Both theories also account for an alternative question denying the use of *Yes* in its answer. Thus, while the theories are united in the two permissible responses for alternative questions, the four

different Y/N interpretations offer four different subsets of licit answers. An “Open” utterance in Inquisitive Semantics has the same focus and final contour as a Commitment Space Discourse utterance comprised of two monopolar propositions (69). However, their licit responses differ based on the machinery within the framework. Specifically, IS rules out *Yes* as an option before disjuncts, whereas CSD does not explicitly rule it out. Next, a “Block” intonation pattern in Inquisitive Semantics, comprised of only one “list” (does not contain a phrasal break), allows the largest variety of answers (70a). Finally, Commitment Space Discourse’s bipolar interpretation of a disjunctive question, also comprised of only one list, allows only the bare particles *Yes* and *No* as licit answers (70b).

(69) Did Ed meet Ann[↑], or (did Ed meet) Beth[↑]?

a. Inquisitive Semantics: Open

Licit Responses: Ed met Ann; Ed met Beth; No; Ed met both

b. Commitment Space Discourse: Two monopolar propositions

Licit Responses: (Yes,) Ed met Ann; (Yes,) Ed met Beth; Ed met both

(70) Did Ed meet Ann-or-Beth[↑]?

a. Inquisitive Semantics: Block

Licit Responses: (Yes,) Ed met Ann; (Yes,) Ed met Beth; Yes; No; Yes, Ed met both (Ann and Beth)

b. Commitment Space Discourse: Bipolar proposition

Licit Responses: Yes; No

These responses are licensed when the question is a typical disjunctive question that uses intonation to disambiguate. But what happens when *either* is added to the question? That is, does (71) offer a paraphrase like (a), or like (b)?

(71) Did Ed meet either Ann or Beth?

- a. Did Ed meet *any* of these people: Ann, Beth?
- b. Ed met *which one* of these people: Ann or Beth?

Paraphrase (a) would occur if (71) were interpreted as a Y/N question, while paraphrase (b) would occur if (71) were interpreted as an alternative question. It has been assumed (Huddleston 1994, Haspelmath 1997) that only interpretation (a), or a Y/N interpretation, is available. However, the data from the production experiment may lend some insight into this: *both* interpretations seem to be available, although the Y/N interpretation (a) occurred more often. That is, participants could put either an alternative question tune or a Y/N question tune on utterances like (71). When looking at the 31 participants used in the production analysis, 26 instances occurred when participant used *either* in their utterance; 5 were alternative question interpretations, 21 were Y/N interpretations.³⁷ Interestingly, if a participant used *either* in one question type, they did *not* use it in the other question type. Only two participants used *either* in an alternative tune; seven participants used *either* in a Y/N tune. All of the instances were narrow scope.

We might predict, therefore, that *either* could sway an ambiguous utterance (72) to a Y/N interpretation, as it seems to be more prevalent for that interpretation.

(72) Did Travis bring (either) drinks or dessert?

Thus, one aim of this experiment is to observe whether adding *either* into a disjunctive question successfully biases the interrogative phrase to a Y/N interpretation. Additionally, the

³⁷ Two additional instances occurred in errors: one as a nonquestion, and one as an (still!) ambiguous question where the interpretation was not apparent.

predictions on answerhood conditions made by Inquisitive Semantics and Commitment Space Discourse will be compared to participants' responses.

4.4 Results

4.4.1 Filler Analysis

Two participants' data had to be thrown out because of server error (while they had the correct number of responses, it was because one or more responses were counted twice while others were completely missing). We used the remaining filler data to score each participant. The idea was to filter out any participant who may not have been completing the task at hand (i.e. rating acceptable sentences as unacceptable and vice versa). Table 4.3 details the scoring system used in the filler analysis.

	+2 points	+1 point	+0 points
“Bad” Filler (12)	1-2 rating	3 rating	4-7 rating
“Good” Filler (3)	6-7 rating	5 rating	1-4 rating
“OK” Filler (1)	3-5 rating	2, 6 rating	1, 7 rating

Table 4.3: Filler Scoring

The scoring was also meant to ensure that participants were using the entire scale to score the stimuli, rather than simply one end of it. To make sure the fillers were being scored how we expected, we averaged together the score of the participants for each filler question. The fillers were expected to have a score of at least 0.8 (people were more often scoring the filler as we expected). One filler question scored a 0.62 and was therefore thrown out³⁸; the rest of the

³⁸ The question was “Do you want ketchup and mustard on your hot dog”; the response was “I hate ketchup”. While this does not fill the traditional role of an adequate answer in terms of addressing both conjuncts, it is possible many took this as a conventional implicature to mean that the speaker is choosing only mustard and not ketchup.

scores were 1.3 or above (up to 2.0), indicating that most participants were scoring 1 or 2 on the question, which was on the correct end of the spectrum. Thus, of the 15 remaining fillers, participants were expected to score at least a 15 out of possible 30. 92 participants scored a 20 or above; The remaining two had to be thrown out because the filler data suggested that the participants were not actually completing the task at hand (scoring a 5 and 6, respectively). The remaining 92 participants' data were used in the analysis.

4.4.2 Stimuli Analysis

An ordinal mixed-effects regression (Christensen 2019) was performed on the data, with independent factors of question type (baseline or *either*), answer type (answers 1-6), and disjunct location (beginning, middle, end)³⁹, and the participant's rating was the dependent factor. Answer type had random slopes by participant, and participants each had a random intercept. In addition, the trial number had a random intercept⁴⁰. While each answer type was found to be significant (at least $p < 0.05$ for each answer type), question type was not significant ($p > 0.1$), nor was any interaction between question type and answer type ($p > 0.1$). Table 4.4 shows the average rating across items for each question type and answer type. While the ratings are comparable between both question types within an answer type, the ratings differ greatly from one answer type to another.

³⁹ The base level for the answer type was answer condition 1, which was the only answer condition permissible for both types of disjunctive questions. The base level for disjunct location was end, because canonical discussion of Alt. and Y/N questions always have the disjunct in this position, and it is unclear whether a different position would affect acceptability.

⁴⁰ The model failed to converge with random slopes for the trial number. However, random slopes aren't needed for trial numbers when the lists have been counterbalanced (as in this experiment). Additionally, a model with random slopes by disjunct position was a worse fit than one without ($p = 0.2513$ model comparison)

Average of entered Response							
	Disjunct	Yes, disjunct	Cleft	Yes	No	Yes, both	Grand Total
Baseline	6.64	6.45	5.52	4.18	5.71	6.56	5.84
Either	6.70	6.50	5.54	4.35	5.70	6.58	5.89
Grand Total	6.67	6.48	5.53	4.26	5.71	6.57	5.87

Table 4.4: Average rating by question type and answer type. Question type is not significant (see grand total in right column for each question type), while answer type is significant (see grand total in bottom row for each answer type).

The best fit model, in fact, did not have ‘question Type’ as a factor at all. Each answer in the answer type factor was found to be significant, as well as the random slope by participant and random intercepts for participant and trial number⁴¹. The answer type factor was compared to the baseline condition of the plain disjunct; all other answer types had negative estimates in the model (Yes + disjunct: $\beta = -1.47$, $SE = 0.299$, $z = -4.915$, $p < 0.0001$, Cleft: $\beta = -3.58$, $SE = 0.316$, $z = -11.33$, $p < 0.0001$, Yes: $\beta = -5.97$, $SE = 0.347$, $z = -17.22$, $p < 0.0001$, No: $\beta = -3.41$, $SE = 0.34$, $z = -10.03$, $p < 0.0001$, Yes, Both: $\beta = -1.26$, $SE = 0.31$, $z = -3.495$, $p < 0.0001$), indicating that those answer conditions were rated lower than the condition expected to be licit for both disjunctive question types. Additionally, where the disjunct occurred in the sentence (start, middle, end) was also significant; both the middle ($\beta = -1.08$, $SE = 0.31$, $z = -4.04$, $p < 0.001$) and the start ($\beta = -1.02$, $SE = 0.31$, $z = -3.32$, $p < 0.001$) positions for disjunct location had a negative estimate, indicating they were rated less acceptable than the questions with the disjunct at the end of the utterance. The disjunct location also had a significant interaction with

⁴¹ Random slopes for answer type by trial number was attempted but the model did not converge.

answer type⁴² across all levels. Table 4.5 shows the specifics of the best fit model. Figures 4.1 and 4.2 show the significant factors of answer condition and disjunct location. Figure 4.3 shows the significant interactions between these two factors. See section 3.4.2 for the description of how the error bars were produced.

	Estimate	Std. Error	z value	Pr(> z)
answerType2	-1.4705	0.2992	-4.915	8.89e-07
answerType3	-3.5795	0.3158	-11.334	<2e-16
answerType4	-5.968	0.3466	-17.219	<2e-16
answerType5	-3.4079	0.3396	-10.026	<2e-16
answerType6	-1.257	0.3089	-4.069	4.72E-05
disLocMiddle	-1.0842	0.3103	-3.495	0.000475
disLocStart	-1.0245	0.3089	-3.317	0.00091
answerType2:disLocMiddle	0.846	0.3358	2.519	0.01176
answerType3:disLocMiddle	0.7149	0.315	2.27	0.023226
answerType4:disLocMiddle	1.9133	0.3124	6.125	9.05E-10
answerType5:disLocMiddle	2.0076	0.3292	6.099	1.07E-09
answerType6:disLocMiddle	1.2636	0.3479	3.633	0.000281
answerType2:disLocStart	1.073	0.3384	3.171	0.001521
answerType3:disLocStart	1.3658	0.3146	4.342	1.41E-05
answerType4:disLocStart	2.0338	0.3103	6.554	5.59E-11
answerType5:disLocStart	1.6618	0.3225	5.152	2.57E-07
answerType6:disLocStart	1.0108	0.3445	2.934	0.003341

Table 4.5: Model of best fit for ordinal data: all answer conditions and disjunct location conditions were significant, as well as their interactions. R code for model: $enteredResponse \sim answerType * disLoc + (1 + answerType | participantID) + (1 | trialNumber)$

⁴² Disjunct Position (DisPOS) was tested as a different factor (subject vs object); this model also had significant interaction between DisPOS and answer type, but the model with disjunct location was a better fit ($p=1.702e-12$). The model with random slopes for disjunct position by participant was not a better fit than one without ($p=0.6643$).

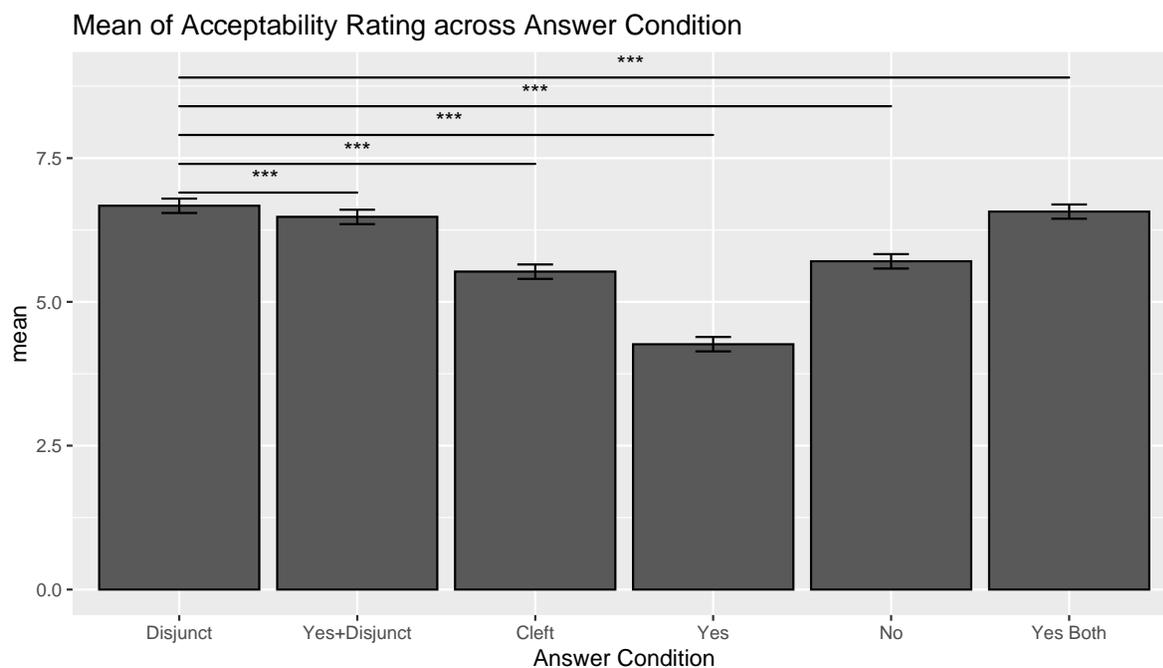


Figure 4.1: Mean of acceptability rating by answer condition.

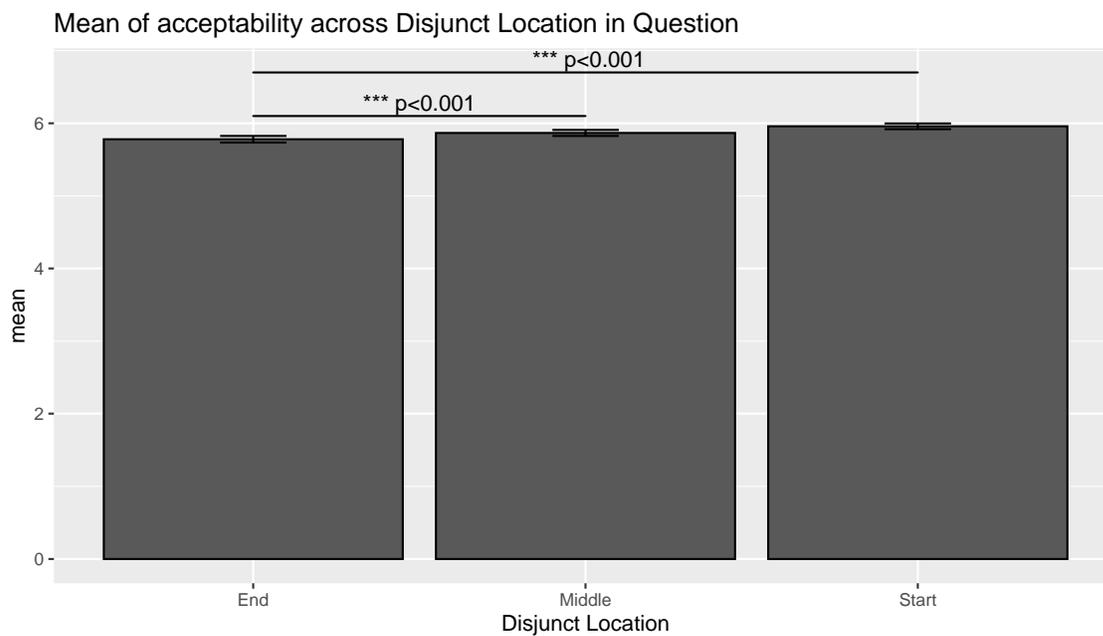


Figure 4.2: Mean of acceptability rating by disjunct location in question. Means are End: 5.78; Middle: 5.87; Start: 5.96.

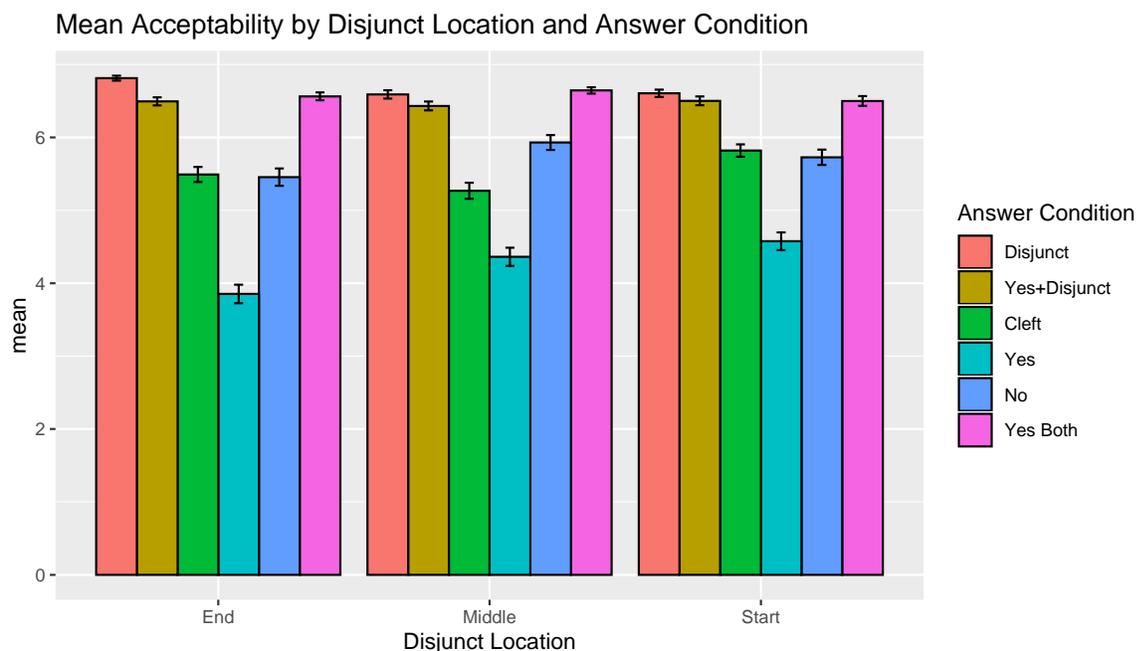


Figure 4.3: Mean of acceptability rating by answer condition (color) and disjunct location (x-axis grouping). All levels within the factors were significant ($p < 0.05$).

The further the disjunct occurred from the end of the sentence, the better the overall rating, combining all answer types (Figure 4.2). This does not hold for each specific answer type, however. In particular, the cleft answer and bare *Yes* answer were considered quite better when the disjunct occurred at the start of the sentence vs. the middle or end of a sentence (Figure 4.3, cleft and *Yes* answer conditions). That is, participants rated those answers higher to questions like (73) than those like (74) or (75).

(73) [Start] Did Cameron or Jillian feed the dog yesterday?

(Cleft: It was Cameron.)

(74) [Middle] Is Martha videotaping Lillian or Dorian in the school play?

(Cleft: It's Lillian)

(75) [End] Does Amy use Hulu or Netflix?

(Cleft: It's Netflix)

No and *Yes, both* had the highest ratings when the disjunct was in the middle of the sentence, and finally a simple disjunct answer and “Yes + disjunct” answer had the best ratings when occurring at the end of a sentence. This is an interesting result that doesn’t seem to be predicted by either theory; neither Inquisitive Semantics nor Commitment Space Discourse offers differing answerhood conditions based on the *location* of the disjunct in the question. Both theories operate on an assumption that a response will remain licit or illicit, regardless of the location of the disjunct in the question.

4.5 Discussion

Participants did not score sentences with *either* vs. sentences without *either* differently in the text experiment. This can have one of two interpretations: *either* does not successfully bias the text to a Y/N interpretation, or the text is already biased to a Y/N interpretation (and thus at most we can claim that *either* does not bias the utterance to an alternative interpretation and is neutral at best).

If the text is already biased to a Y/N interpretation, then the predictions for acceptable answers for Y/N questions ought to match the observed ratings in the experiment. Recall that both theories offered multiple Y/N answerhood predictions, depending on the focus and phrase break structure of the question. Table 4.6 lists the average rating across answer type, collapsed over question type (since this was not a significant factor) and disjunct location (since the theories offer the same predictions regardless of location).

	Answer Types						
	Disjunct	Yes, disjunct	Cleft	Yes	No	Yes, both	Grand Total
Rating Average	6.67	6.48	5.53	4.26	5.71	6.57	5.87

Table 4.6: Average Rating across answer types

The answer with the lowest acceptability was the bare particle *Yes* at 4.26 (which is still on the side of more acceptable, however). Next, the cleft answer and *No* both scored about a full point to point-and-a-half higher; finally, the *Yes, both* and “Yes + disjunct” scored almost 6.5 each, while the answer with the best rating was the simple disjunct answer with 6.67. Thus the six answers, while each having significantly different ratings (as the ordinal regression model revealed), fall into three distinct tiers of acceptability: borderline acceptable (rating 4-5), quite acceptable (rating 5-6), and very acceptable (rating 6-7).

Table 4.7 reveals that the ‘very acceptable’ tier matches Commitment Space Discourse’s predictions on which answers are acceptable for a Y/N question comprised of two monopolar questions (refer back to Table 4.2). This type of interpretation requires a rising accent on both disjuncts (Krifka 2016). None of the Inquisitive Semantics predictions match this tier; the acceptable answers for a Y/N interpretation with rising accents on both disjuncts allows the bare disjunct answer and *No*, as IS specifically rules out allowing *Yes* before one or both disjuncts (Roelofsen 2013a).

Tier	Answer					
	Disjunct	Yes, disjunct	Cleft	Yes	No	Yes, both
Borderline Acceptable (4-5)				✓		
Quite Acceptable (5-6)			✓		✓	
Very Acceptable (6-7)	✓	✓				✓

Table 4.7: Answer types categorized into tiers by average observed rating

While Commitment Space Discourse, therefore, does account for the best answer ratings, it does not explain the acceptable ratings for the remaining answer types. Possibly, this type of

interpretation of the disjunctive question is the *default* when reading the text; however, it is known that participants will try to fit a question together with its context (Heidenreich 2014b), so when the default interpretation does not produce an acceptable rating, other interpretations may be considered. This could also explain the reduced, though not necessarily bad, ratings; if participants are aware of an interpretation of the question that does *not* fit with the answer, they may lower the rating, even after thinking of an acceptable interpretation.

The second tier of ratings contains the cleft answer and bare *No*. Inquisitive Semantics predicted that bare *yes* is infelicitous while bare *no* is perfectly acceptable (Roelofson & van Gool 2010); it is the only theory to offer a prediction that *no* may be more acceptable than *yes*. In fact, if Inquisitive Semantics would allow *yes* before an answer of one or both disjuncts, it would account for all four answer types discussed so far, as there is evidence that bare particle responses are considerably less felicitous than ones with an explicit prejacent (Farkas and Roelofson 2015). This would account for the reduced rating of *no* in comparison to the three answers in the upper tier. Commitment Space Discourse cannot explain why bare *no* may be more acceptable than *yes*; all predictions by this theory have both bare particles as equally acceptable or unacceptable.

The cleft answer is acceptable for an alternative question interpretation under both theories; it is here than we may hypothesize that the alternative question interpretation is an exception when reading text, and that participants had to ‘search’ for a fitting context in order to make the question and answer fit together. The reduced rating may be due to the fact that the first Y/N context would not accept this answer (and thus a context exists where this question-answer pair does not fit together). This result suggests, first, that the assumption that a disjunct answer in any form is acceptable as a Y/N question response; the cleft answer is decidedly less

felicitous than the simple disjunct response. However, since the results suggest that the ambiguous text actually has implicit bias to a Y/N interpretation, it is unclear whether the cleft answer is more acceptable for an Alt. interpretation. The perception experiment (Chapter 5) aims to respond to the assumption that any acceptable answer for an Alt. question is also acceptable for a Y/N question. The hypothesis concerning *why* this result occurs is also dependent on a specific type of relationship between a Y/N question and an Alt. question, where the Y/N question acts as the ‘default’ and the Alt. question acts as the ‘exception’; we might expect this kind of relationship to occur in a superset/subset relationship. The relationship between the two interpretations is further explored in the artificial language experiment (Chapter 6).

Finally, the bare *Yes* answer has the lowest acceptability, though still on the end of ‘acceptable’. Both theories offer predictions for when bare *yes* is acceptable: it would occur after a Y/N question that is comprised of a single phrase (i.e. no phrase breaks; a one-list question in Inquisitive Semantics, a bipolar question in Commitment Space Discourse). It is possible that this type of Y/N interpretation is possible, but not considered immediately; that is, participants first consider a Y/N question with raising accents on both disjuncts (usually contains a phrase break); then, attempting to exhaust all possibilities, they consider a Y/N interpretation comprised of only a single phrase (or an alternative question interpretation). This type of Y/N interpretation occurs when the distinction between the disjuncts is not salient. For example, consider if the speaker of (76) were attempting to bake a cake for Mary.

(76) Is Mary allergic to dairy or soy?

Suppose that cake used both dairy and soy, and the speaker wants to ensure that Mary isn’t allergic to dairy or soy. The distinction between the disjuncts would not be important; the simple

yes response would be sufficient enough for Mary to know that she would need to consider other options.

This necessity to consider other interpretations might be observed in the reaction time of the participants. Unfortunately, Amazon Mechanical Turk might not offer reaction time data that would either confirm or deny this claim: there were many responses that took 30+ seconds, and as participants were not instructed to complete the experiment in one sitting without breaks, or even to complete the experiment as quickly as possible, it is likely that reaction times will not be an accurate indicator of how long the participant considered the question-answer pair.

Therefore, fifteen participants were chosen to complete the reaction time analysis. There had to be at least one participant from each list, and participants were chosen based on their fastest overall average reaction time, while also keeping standard deviation low (below 3 seconds). The low standard deviation hopefully translated to continued focus on the experiment, rather than taking breaks. Table 4.8 reveals the participant with the lowest (minimum) average reaction time and average standard deviation from the subset (column 1), the participant with the highest (maximum) average reaction time and average standard deviation (column 2), the overall average reaction time and standard deviation of the chosen fifteen participants (column 3), and then the overall reaction time and standard deviation across all participants from the experiment (column 4). Participant 73 (pID 73) has both the maximum average reaction time and maximum average standard deviation, but had to be included because the data were the best from that list group.

	Minimum avg.	Maximum avg.	Average (subset)	Average (total)
Reaction Time (sec)	3.05 (pID 48)	6.41 (pID 73)	4.61	8.00
Standard Deviation	0.86 (pID 26)	2.84 (pID 73)	1.81	19.40

Table 4.8: Metrics of fifteen participants used in RT analysis, in comparison to full dataset

Overall, these participants performed more consistently and faster than the average participant. A linear-mixed effects regression was modeled, with the dependent variable of reaction time (in seconds) and the independent variables that were found to be significant in the ordinal regression analysis: answer type and disjunct location (and their interaction) with random slopes for answer type by participant and random intercepts for participants and trial numbers. Question type was added back in as a factor, as the reaction time was measured from when the page loaded to when the participant entered a rating; therefore, the addition of reading *either* may contribute to the reaction time. Other random effects (random slopes for question type by participant, random slopes for disjunct location by participant) were tested until the model with the best fit was found. Surprisingly, neither question type nor answer type were significant in predicting reaction time; however, disjunct location did have an effect: namely, reaction times were significantly longer when the disjunct occurred in the middle of the sentence ($\beta = 0.57$, SE = 0.15, $z = 3.87$, $p < 0.001$) vs. utterance-initial or utterance-final (Table 4.9 and Figure 4.4).

	Estimate	Std. Error	z value	Pr(> z)
Intercept	4.423158	0.24276	18.242	1.09e-13
disLocMiddle	0.573200	0.148147	3.869	0.000334
disLocStart	-0.009983	0.148147	-0.067	0.946558

Table 4.9: Linear mixed-regression model of best fit for reaction data of 15 participants. Only the middle disjunct location was significant. R code for model:

reactionT~disLoc+(1/participantID)+(1/trialNumber)

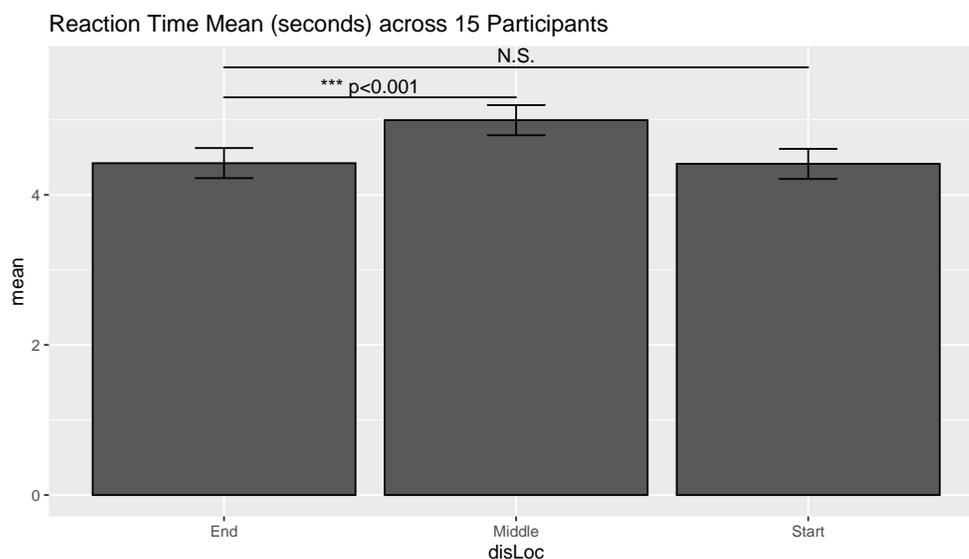


Figure 4.4: Reaction Time mean in seconds, grouped by location of disjunct in the question. RT average for end, middle and final location was 4.423, 4.996, and 4.413 seconds, respectively.

Participants were significantly delayed when the disjunct occurred near the center of the utterance; however, this did not equate to a necessarily lower rating (recall that both *No* and *Yes*, both had their highest ratings when the disjunct occurred in the middle, c.f. Figure 4.3); additionally, questions with disjuncts in the middle had higher overall ratings than those with the disjunct utterance-final, c.f. Figure 4.2). To ensure this wasn't a coincidental finding of the subset used, Table 4.10 reveals the average reaction time by disjunct location for the entire dataset.

	Disjunct Location in Question		
	Utterance-initial	Middle of utterance	Utterance-Final
Reaction Time (average in sec)	7.54	10.43	7.56

Table 4.10: Average reaction time (in seconds) for stimuli by location of the disjunct in the question, whole dataset

A linear mixed-effects regression over the entire dataset confirms the significance finding of disjunct location and the middle disjunct location alone ($\beta = -1.33$, $SE = 0.39$, $z = 3.43$, $p < 0.01$, c.f. Table 4.11). Thus Figure 4.4 is an accurate representation of the entire dataset.

	Estimate	Std. Error	z value	Pr(> z)
Intercept	6.8743	0.3910	17.583	<2e-16
disLocMiddle	1.3319	0.3873	3.439	0.00123
disLocStart	0.2437	0.3873	0.629	0.53228

Table 4.11: Linear mixed-regression model of best fit for reaction data of all participants. Only the middle disjunct location was significant. R code for model:

reactionT~disLoc+(1/participantID)+(1/trialNumber)

As the reaction time data records from when the page loads to when the participant chooses a rating, other factors such as reading speed may have affected the data; therefore the hypothesis that participants may be initially thinking of a specific Y/N tune when entering their ratings cannot be confirmed or denied by the results. However, clearly the location of the disjunct affected either reading times or acceptability judgments.

As a final note, it appears that the Centering theory (Brennan 1995) may account for why the cleft answer was preferred with the disjunct location at the start of the utterance, rather than the middle or the end: participants prefer to pronominalize the syntactically prominent discourse entity, or ‘center’, which is typically the subject. However, the theory does not account for the remaining differences in acceptability between disjunct location and answer condition.

Furthermore, the theory does not explicitly handle conversation between two speakers, but rather details how a single speaker may continue discourse via pronominalization. This is an intriguing area for future research.

4.6 Conclusion

These observations support the notion that *either* does not bias the text because it is already biased to a Y/N interpretation; this is contrary to the assumption that disjunctive questions, without intonation, are ambiguous and unbiased. The results can neither confirm nor deny that *either* can only occur within the Y/N interpretation, as the text was shown to be inherently biased to that interpretation (without *either*). However, the perception experiment (Chapter 5) offers more data on this assumption. Additionally, the production experiment data revealed that some participants inserted *either* in their Alt. question productions, already indicating that this assumption concerning the grammaticalization of disjunctive questions may prove false. That is, the assumption that the insertion of *either* can occur grammatically in Y/N questions but not Alt. questions is not supported by the production data, where *either* was freely inserted into both contexts.

The remaining assumption concerns the acceptability of answers to specific disjunctive question interpretations. Overall, the empirical data supported canonically acceptable answers, based on the predictions by the semantic theories. Commitment Space Discourse accurately predicts the highest tier of acceptable answers; Inquisitive Semantics correctly predicts *No* being more felicitous than *Yes*. Additionally, the reason that the bare *No* answer is seen as less acceptable, despite also being an ‘adequate’ Y/N answer, is accounted for by previous evidence that bare particle responses are considerably less felicitous (Farkas and Roelofsen 2015). One theory is that participants might have been reading the text and ultimately applying the “rule” (specific Y/N interpretation), and if that did not fit, then applying specific “exception” contexts (Alternative interpretation, other Y/N interpretations). If the answer did not fit with the default context, a lower rating was given (even if it fit well with another interpretation). Of course, this

analysis depends on the current assumption that Alt. questions are a subset of Y/N questions, with Y/N questions being a ‘default’ interpretation, and Alt. questions being an ‘exception’ interpretation. This type of relationship is tested in the artificial language experiment (Chapter 6).

Finally, neither theory predicts why licit answers may be more or less *acceptable* depending on where the disjunct occurs in the question. Both theories (and, in fact, all discussion concerning disjunctive questions in general) operate on the assumption that an answer is licit or illicit, regardless of where the disjunct occurs in the question. The experimental data refute this assumption. Furthermore, it is not the case that all answers are more acceptable when the disjunct occurs in a certain place; rather, some answers were more acceptable when the disjunct occurred utterance initially, while other answers received higher ratings when the disjunct occurred at the end of the utterance; some even received their highest ratings when the disjunct occurred in the middle of the sentence. It is possible that the theories may be altered so that they can account for such a difference; this will be explored in the Discussion chapter (Chapter 7).

Chapter 5: Perception Experiment

5.1 Introduction

The text experiment (Chapter 4) revealed an apparent bias of disjunctive questions in text form to the Y/N interpretation. It was therefore unclear whether *either* could be inserted into a disjunctive question and have an Alt. question interpretation. The perception experiment was conducted in order to shed light on the assumption that *either* cannot coexist with an Alt. interpretation (Huddleston 1994, Haspelmath 2007). The production experiment already revealed that some participants voluntarily placed *either* in their Alt. productions, placing doubt on this previously held belief.

This experiment was also conducted to shed light on variable intonational contours and syntactic variations on the typical ‘disjunct’ response. It is assumed that a disjunct (and only a disjunct) answer is licit for an Alt. question (Aloni & van Rooy, 2002); yet the intonational contour and syntactic nature of that disjunct is unspecified. Furthermore, it is assumed that any response that is licit for an Alt. question would also be licit for its Y/N question counterpart (Aloni & van Rooy 2002, Biezma & Rawlins 2012). These assumptions are tested in a question-answer acceptability task.

5.2 Methodology

5.2.1 Design and Materials

The stimuli questions were the same as used in the text-only experiment. Each question stimulus had two different tunes:

- (A) An alternative question tune: this was not a specific ToBI tune, but rather had the restrictions that the first disjunct must end high, and the second disjunct must end low

(B) Y/N question tune : again, not a specific ToBI tune, but rather the entire phrase had to end with H-H%

Thus each stimulus had four different question types, crossing the tune factor with the *either* factor (w/ or w/out *either*). Each question type was paired with one of the four⁴³ following answer conditions:

Condition 1: A noun/sentence response with a falling tone (L- L%)

Condition 2: A noun/sentence response with a rising (continuation) tone (ended L*+H L- H%)

Condition 3: *Yes*, followed by the response from condition 1

Condition 4: A cleft response

The first and second answer conditions contain the same response per item, merely differing in intonation. The third answer condition is the typical “Yes + disjunct” response, while the fourth condition is the cleft response. Each stimulus token had four question variations and four answer conditions for sixteen different question-answer combinations (77). The ToBI transcriptions for answer conditions 1 and 2 (which varied only intonationally) are provided.

(77) Question variations:

- A. Did Melanie or Emily go to the party tonight? (Alt. tune, w/ out *either*)
- B. Did Melanie or Emily go to the party tonight? (Y/N tune, w/ out *either*)
- C. Did either Melanie or Emily go to the party tonight? (Y/N tune, w/ *either*)
- D. Did either Melanie or Emily go to the party tonight? (Alt. tune, w/ *either*)

Answer Conditions:

⁴³ The text experiment had more answer conditions: Bare ‘yes’ and ‘no’ as well as ‘Yes, both’. Adding these three answers in this experiment would have given 28 unique question-answer combinations per item; this would have resulted too few unique question-answer pairs per participant. Instead, it is an area for future exploration.

1. Melanie went.

(H* L- L%)

2. Melanie went

(L*+H L- H%)

3. Yes, Melanie went.

4. It was Melanie.

Each participant would hear one combination of the possible sixteen question-answer pairs for each item, with a total of 48 items (the same question stimuli items as the text experiment).

Therefore, sixteen different lists were produced, with each list containing three tokens of each question-answer combination. The three tokens were further broken down to each have a different disjunct location in the question: start (78), middle (79), or end (80).

(78) [Start] Did Cameron or Jillian feed the dog yesterday?

(79) [Middle] Is Martha videotaping Lillian or Dorian in the school play?

(80) [End] Does Amy use Hulu or Netflix?

Thus, the manipulated factors were: tune of question, *either*, answer condition, and location of disjunct in the question.

There were also 12 infelicitous fillers and 12 felicitous fillers (different from the text experiment), which were infelicitous/felicitous based on where the contrastive prosody in the response occurred. Thus, each participant had 72 question-answer pairs.

5.2.2 Procedure

The participant would *only* hear the prompt question and answer (no text), and then were asked to rate the acceptability of the question-answer pair on a scale 1-7. A low number was less acceptable; a high number was more acceptable. Participants could replay the question-answer

prompt as many times as they wanted. They were required to choose a number before continuing to the next stimuli. They heard 72 question-answer pairs: 48 stimuli and 24 fillers.

5.2.3 Participants

93 participants completed the experiment on Amazon Mechanical Turk (AMT), with at least five participants from each list. The experiment ran from 6-13-18 until 6-18-18. The entire experiment took 10-25 minutes. Participants also filled out a brief survey which asked their age, gender, ethnicity, and native language(s). They were paid \$3 for their participation.

5.3 Predictions

Both Inquisitive Semantics and Commitment Space Discourse offer similar predictions for the question-answer combinations above. Table 5.1 shows the predictions for Inquisitive Semantics for each question-answer combination; Table 5.2 shows the predictions for Commitment Space Discourse.

Question Types	Answer Conditions			
	Disjunct↓	Disjunct↑	Yes + disjunct	Cleft
Alternative tune, w/out either	✓	✓		✓
Y/N tune, w/out either	✓	✓	✓*	
Alternative tune, with ‘either’	?	?	?	?
Y/N tune, with ‘either’	✓	✓	✓*	

*Table 5.1: Inquisitive Semantics predictions across question-answer combinations. *This answer is acceptable only if the Y/N tune has one phrase, instead of two.*

Inquisitive Semantics does not take the intonation of a ‘disjunct’ answer into account; therefore, if the framework allows a disjunct as an answer, it predicts both answer conditions 1

and 2 to be felicitous. Experimentally, we would expect different results: answer condition 1 would be the typical disjunct answer and its licensed answerhood, while answer condition 2, which has a continuation rise, would only be licensed by a Y/N question, via its ‘forward reference’ signal (Pierrehumbert & Hirschberg 1990). The continuation rise would indicate incompleteness in answering the question; since one disjunct is already given, logic would dictate that an alternative question (with an ‘exactly one’ stipulation) would not accept a continuation rise on the response disjunct.

Additionally, the acceptability of “Yes + disjunct” for a Y/N tune depends on the specific Y/N tune of the question: a tune with two separate phrases (with a phrasal boundary in between disjuncts) does not allow this as an answer (an optional *Yes* before the disjunct is disallowed); a tune that is one whole phrase (no phrase boundary) does allow this as an answer. As in the text experiment, the predictions both with and without *either* for Y/N questions remain the same.

However, the predictions for an alternative tune with *either* are unknown. The production experiment revealed that *either* could occur in both alternative questions and Y/N questions, though it occurred more often in Y/N questions and thus may sway interpretation that way. The text experiment did not indicate a bias in *either*, only because the text (without context or intonation) seemed to be already biased to a Y/N interpretation. The fact that *either* did not significantly alter ratings in the text experiment, however, indicates that *either* did not sway interpretation to an alternative interpretation and therefore *may* be biased to a Y/N interpretation. Therefore, if participants rely more on aural cues (the tune of the question), then the answer conditions that correspond with an alternative question ought to be predicted; if, however, participants rely on assumed lexical cues (the use of *either*), then the answer conditions that correspond with a Y/N question ought to be predicted.

Commitment Space Discourse (Table 5.2) offers similar predictions, with the caveat that the “Yes + disjunct” answer is licensed for Y/N only if the Y/N disjunctive question has a rising accent on both disjuncts.

Question Types	Answer Conditions			
	Disjunct↓	Disjunct↑	Yes + disjunct	Cleft
Alternative tune, w/out either	✓	✓		✓
Y/N tune, w/out either	✓*	✓*	✓*	
Alternative tune, with ‘either’	?	?	?	?
Y/N tune, with ‘either’	✓*	✓*	✓*	

Table 5.2: Commitment Space Discourse Predictions. *This answer is acceptable only if the Y/N tune has rising accents on both disjuncts

The predictions for the alternative tune with *either* again depend on whether the participants rely more on aural cues or lexical cues. Therefore, the question with the alternative tune and *either* (across both theory types) would have the predictions listed in Table 5.3.

	Answer Conditions			
	Disjunct↓	Disjunct↑	Yes + disjunct	Cleft
Lexical cues (<i>either</i>) sway interpretation	✓	✓	✓	
Aural cues (alternative contour) sway interpretation	✓	✓		✓

Table 5.3: Predictions on answer conditions for alternative question tune with *either*.

5.4 Results

5.4.1 Filler Analysis

The fillers were analyzed first in order to filter out any participant who may not have been completing the task at hand (e.g. ignoring/not listening to the sentences and only paying attention to the text). As the fillers were designed to be felicitous or infelicitous based on prosody, it was necessary for the participants to be actively listening to prosody in order to correctly rate the fillers. Table 5.4 details the scoring system used in the filler analysis.

	+2 points	+1 point	+0 points
Infelicitous Filler	1-2 rating	3 rating	4-7 rating
Felicitous Filler	6-7 rating	5 rating	1-4 rating

Table 5.4: Filler Scoring

The scoring was also meant to ensure that participants were using the entire scale to score the stimuli, rather than simply one end of it. To make sure the fillers were being scored how we expected, we averaged together the score of the participants for each filler question. The fillers were expected to have a score of at least 0.8 (people were more often scoring the filler as we expected). Four “infelicitous” filler questions had to be thrown out because of their low scores in the filler analysis, as seen in Table 5.5. While all four sentence/answer pairs were intended to be infelicitous based on the location of the contrastive prosody (marked by capital letters in the answer), the participants scored the fillers as acceptable or neutral. The unexpected high acceptability of question-answer pairs may indicate that participants are willing to overlook intonation discrepancies in order to accommodate a question-answer context.

Filler Score	Avg. Entered Rating	Sentence/Answer Pair
0.032	5.94	Q. How many pairs of shoes does Felicia have? A. Felicia has ten PAIRS of shoes.
0.086	5.78	Q. What color dress is your daughter wearing to the dance this Saturday? A. She is wearing a green DRESS this Saturday.
0.59	4	Q. Does Astrid have on a long dress? A. She has on a LONG skirt.
0.688	3.957	Q. Is your Mom baking an apple pie? A. She is baking an APPLE cake.

Table 5.5: Fillers removed; capital word in sentence/answer pair indicates contrastive prosody.

Two other fillers had borderline scores (0.8065 and 0.8387), but their presence or absence did not affect how many participants' data went on through the stimuli analysis. With four fillers removed, the remaining 20 fillers were used to gauge whether participants were completing the task at hand. A score of at least 20 was therefore required; five participants did not meet this criterion (scores 12, 17, 18, 18, 19) and therefore their data was not used in the following results. One additional participant was missing data and therefore was discarded as well. Therefore, data from 87 participants (31 females, 56 males) were used in the analysis. Each of the sixteen lists had at least five participants.

5.4.2 Stimuli Analysis

The ratings on the Likert scale were used as the dependent factor in an ordinal regression model (Christensen 2019). Independent factors were the tune (alternative vs. polar), presence of *either*, location of the disjunct in the sentence (initial, middle, or final), and answer condition (Disjunct, rising disjunct, yes + disjunct, cleft). Therefore, the stimuli were modeled on a 2 x 2 x 3 x 4 design. The baseline level for disjunct location was once again the end of the utterance; the baseline level for answer condition was the plain disjunct answer (predicted to be acceptable for

both types of disjunctive questions). All four of the factors were in the model of best fit. Table 5.6 provides the details of the regression for the best fit model.

	Estimate	Std. Error	z value	Pr(> z)
answerType2	-3.00754	0.30503	-9.86	2.00E-16
answerType3	-3.65757	0.30602	-11.952	2.00E-16
answerType4	-2.21874	0.33968	-6.532	6.49E-11
AltY/N	-0.55297	0.23448	-2.358	0.01836
eithereither	0.11156	0.22889	0.487	0.625979
disLocMiddle	-0.3793	0.1402	-2.705	0.006824
disLocStart	-0.06672	0.16091	-0.415	0.678432
answerType2:AltY/N	0.92584	0.28261	3.276	0.001053
answerType3:AltY/N	3.28934	0.29666	11.088	2.00E-16
answerType4:AltY/N	-0.96209	0.28785	-3.342	0.000831
answerType2:eithereither	0.02234	0.28403	0.079	0.937296
answerType3:eithereither	0.6809	0.28753	2.368	0.017878
answerType4:eithereither	-0.35972	0.29389	-1.224	0.220953
AltY/N:eithereither	-0.07472	0.31263	-0.239	0.811111
answerType2:AltY/N:eithereither	0.02429	0.39715	0.061	0.951241
answerType3:AltY/N:eithereither	-0.7471	0.41274	-1.81	0.070278
answerType4:AltY/N:eithereither	0.66643	0.4044	1.648	0.099364

*Table 5.6: Model of best fit for ordinal perception data: significant factors were all answer conditions, tune (Alt vs. Y/N), middle disjunct location, as well as the interaction between answer condition and tune and the interaction between answer condition 3 and either. R code for model: enteredResponse ~ answerType * Alt * either + disLoc + (1 + Alt + answerType + disLoc | participantID) + (1 + answerType | trialNumber)*

First, the middle disjunct location was significant ($\beta = -0.38$, $SE = 0.14$, $z = -2.71$, $p < 0.01$); ratings for questions with the disjunct location in the middle were significantly lower than those questions with the location at the start or end of the utterance (Figure 5.1). The middle of the utterance disjunct location had the lowest rating across all answer types for Y/N tunes; it had lowest or middle rating (of the three disjunct locations) for all answer types for Alt tunes.

The answer conditions were significant (Rising disjunct: $\beta = -3.01$, $SE = 0.31$, $z = -9.86$, $p < 0.0001$, Yes + disjunct: $\beta = -3.66$, $SE = 0.31$, $z = -11.95$, $p < 0.0001$, Cleft: $\beta = -2.22$, $SE =$

0.34, $z = -6.53$, $p < 0.0001$), all with negative estimates, indicating that the falling disjunct answer had the highest acceptability rating (Figure 5.2). See section 3.4.2 for the description of how the error bars were produced.

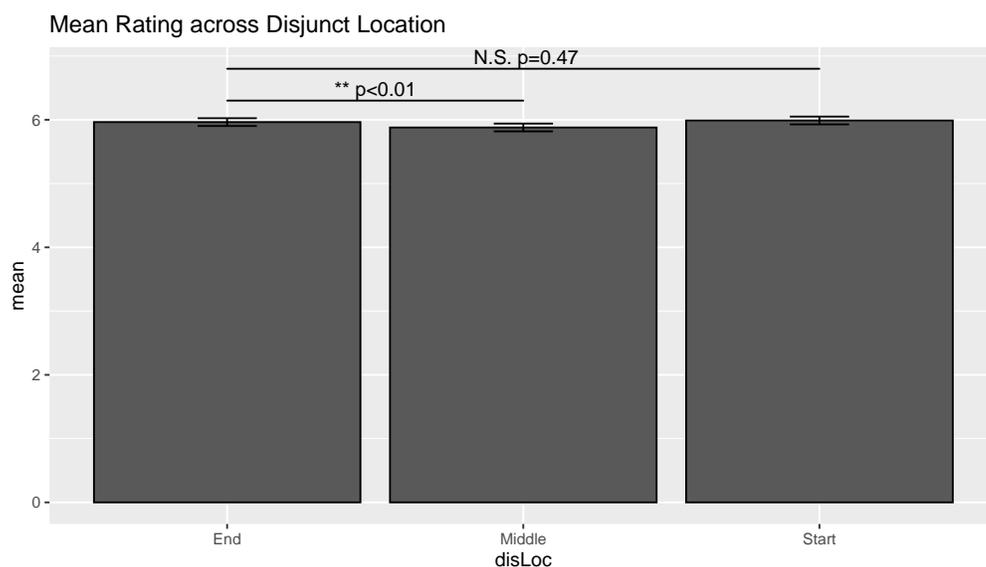


Figure 5.1: Mean Rating collapsed across location of disjunct (*disLoc*) in the question. Mean values are: End: 5.96; Middle: 5.88; Start: 5.98.

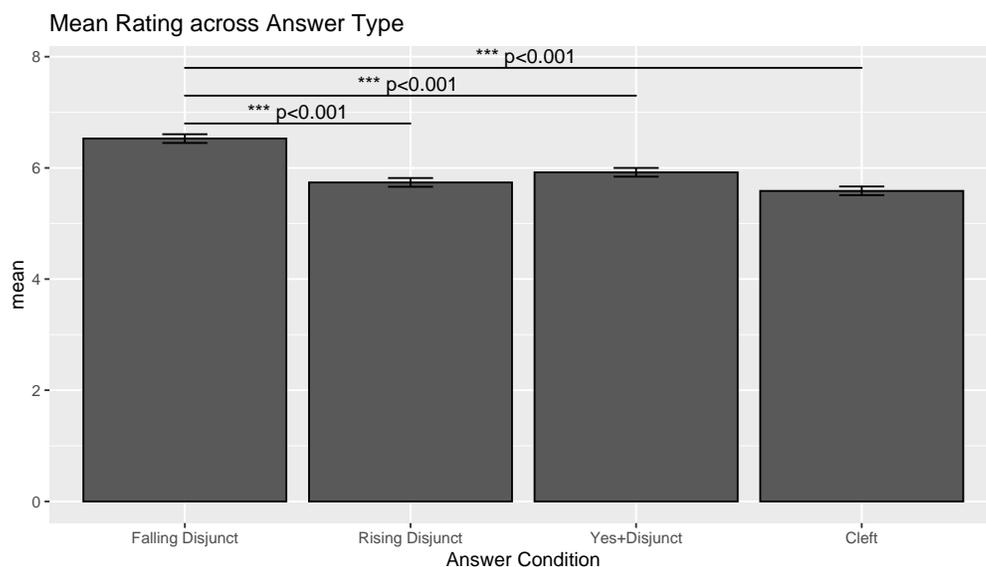


Figure 5.2: Mean Rating collapsed over answer condition. The falling disjunct condition was significantly better than all other answer conditions.

The tune was also significant ($\beta = -0.55$, $SE = 0.23$, $z = -2.56$, $p < 0.05$), revealing that participants scored the Y/N tune overall better than the Alt. tune (mean ratings were Y/N: 5.97; Alt: 5.92). *Either* was not a significant factor alone, but the interaction between *either* and the yes + disjunct answer was significant ($\beta = 0.68$, $SE = 0.29$, $z = -1.22$, $p < 0.05$). The estimate for this interaction was positive, indicating that adding *either* to the question significantly increased acceptability of the question-answer pair when the answer was Yes + disjunct (Figure 5.3).

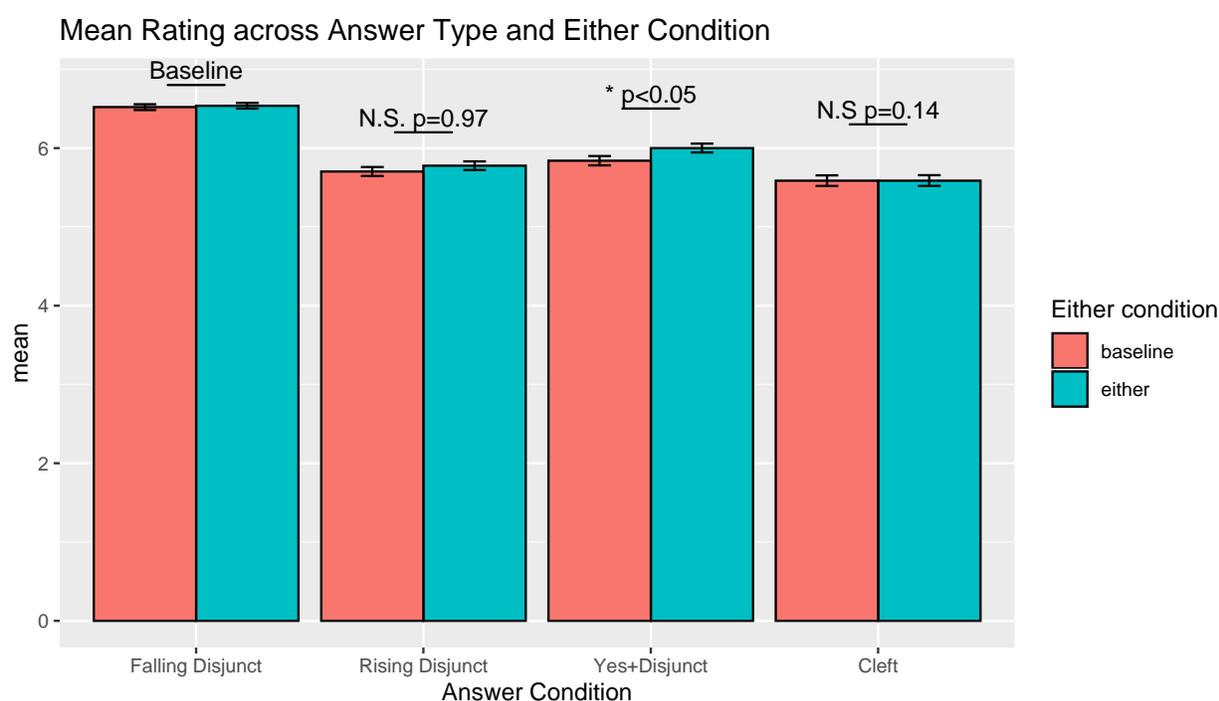


Figure 5.3: Mean rating collapsed over answer condition and either condition. Adding *either* when paired with the Yes + disjunct answer (condition 3) significantly improved the rating.

Mean values listed in Table 5.7

'Either' condition	Answer Condition				Avg. across all answer conditions
	Disjunct ↓	Rising ↑	Yes + Disjunct	Cleft	
baseline	6.52	5.70	5.84	5.59	5.90
w/ 'either'	6.54	5.78	6.00	5.59	5.97

Table 5.7: Average rating over either condition and answer condition; Yes + disjunct interaction with *either* significant, but *either* factor alone not significant.

Whereas most answer condition ratings did not change much when *either* was inserted into the sentence, the “Yes + disjunct” answer condition has a different result: inserting *either* into the sentence significantly improved the rating of the question-answer pair.

Importantly, adding *either* did not decrease acceptability of typical alternative question answers (e.g. the cleft response) when collapsed over tune; however, for the answer condition that was most canonically Y/N, adding *either* did seem to make this answer ‘more acceptable’. It seems likely, therefore, that while *either* does not decrease acceptability for any type of disjunctive question or response, it does have an effect in increasing acceptability for typical Y/N replies when the tune isn’t Y/N: the difference in rating occurs *entirely* in the alternative tune condition (Table 5.8). Adding *either* to an alternative contour increased acceptability of the Yes + disjunct response by .37 points.

‘Either’ condition	Question Tune	Answer Conditions			
		Disjunct↓	Disjunct↑	Yes + Disjunct	Cleft
Baseline (w/ out <i>either</i>)	Alt	6.61	5.66	5.29	5.96
	Y/N	6.43	5.74	6.39	5.21
W/ <i>either</i>	Alt	6.61	5.72	5.66	5.82
	Y/N	6.46	5.83	6.34	5.36

Table 5.8: Average Rating across *either* condition, tune, and answer condition. Note the increase in rating from 5.29 to 5.66 in Alt. Question/Yes + Disjunct response when adding *either*

Also of note is how *either* decreased the acceptability of the cleft answer condition when added into a disjunctive question with an alternative tune. This interaction between the *either* condition, tune contour, and answer condition was nearly significant for the “Yes + disjunct” answer condition ($p=0.070$) and the cleft answer condition ($p=0.099$). Importantly, it seems as though *either* is accepted in both alternative tune context and Y/N question contexts; it can, however, sway acceptability of certain answer contexts. Table 5.8 also demonstrates that for

answer conditions that are licit for both types of questions (e.g. falling disjunct) there is no real difference between acceptability for each tune when adding in *either* (6.61 vs. 6.61 in the alt tune condition; 6.43 vs. 6.46 in the Y/N tune condition). Consistent with the production data discussed in Chapter 4.3, participants accept (and can insert) *either* in both Y/N tunes and alternative tunes.

The final significant interactions included the tune with all answer conditions (Rising Disjunct x Tune: $\beta = 0.93$, $SE = 0.28$, $z = 3.28$, $p < 0.01$; Yes + Disjunct x Tune: $\beta = 3.29$, $SE = 0.30$, $z = 11.09$, $p < 0.0001$; Cleft x Tune: $\beta = -0.96$, $SE = 0.29$, $z = -3.34$, $p < 0.001$). The positive estimates for Rising Disjunct x Tune and Yes + Disjunct x Tune indicate that the ratings were higher for the Y/N question tunes than their Alt counterparts; similarly, the negative estimate for the Cleft x Tune indicates that the Alt. tune received the higher ratings (Figure 5.4). There were also random slopes for tune, answer condition, and disjunct location by participant⁴⁴ and random slopes for answer condition by trial number⁴⁵ and random intercepts for participant and trial.

Table 5.9 shows the average ratings for answer type, collapsed over tune contour. These results are as expected; both tunes have high ratings for a simple disjunct answer (the baseline condition). A Y/N tune scored significantly better with the Yes + disjunct answer in comparison to the Alt tune, while a cleft answer scored significantly better with the Alt tune. The Y/N tune also scored better with the rising disjunct answer, indicating at least a slightly more acceptable answer when paired with the Y/N tune vs. the Alt tune.

⁴⁴ Random slope for *either* by participant did not provide a model of better fit

⁴⁵ Random slopes for trial number would only converge with one factor; the factor of answer condition was in the model of best fit.

Question Tune	Answer Condition			
	Disjunct↓	Disjunct↑	Yes + Disjunct	Cleft
Alternative	6.61	5.69	5.48	5.89
Y/N	6.44	5.79	6.36	5.29

Table 5.8: Average rating over question tune and answer condition; all factors and interactions significant. See Figure 5.4 for visual representation.

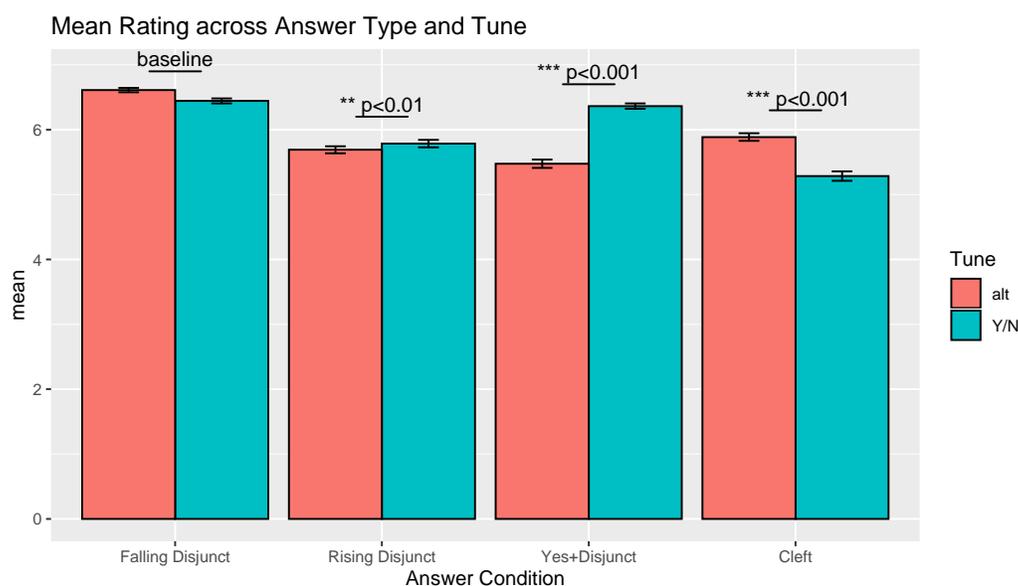


Figure 5.4: Mean Rating over question tune and answer condition; not only is the tune itself significant ($p < 0.05$) and all levels of the answer condition (see Figure 5.2), but the interaction between tune and each answer condition is significant. See Table 5.9 for values.

The disparity in ratings varied across answer condition too; while the yes + disjunct answer scored nearly a full point better with the Y/N tune than alternative tune, the rising disjunct answer was only a tenth of a point better; the cleft answer was nearly six-tenths of a point better for alternative tunes than Y/N, but even the ‘unbiased’ simple disjunct answer scored better for the alternative tune vs. the Y/N tune.

5.5 Discussion

As stated in the text experiment, neither Inquisitive Semantics nor Commitment Space Discourse offer predictions or explanations on disjunctive question acceptability differing based on where in the sentence the disjunct occurs. Therefore, the fact that ‘disjunct location’ was significant (as in the text experiment) merely reveals the necessity for a framework to involve disjunct location in the analysis.

5.5.1 Alt. Questions

Let us move on now to the predictions offered by both theories in regard to alternative questions. Table 5.9 summarizes the predictions of the theories and the observations across *either* condition and answer conditions for the alternative tune. Importantly, the predictions for the alternative tune with *either* are based on participants relying on aural clues (the alternative tune contour) moreso than lexical clues (the presence of *either*, which had previously been assumed to be licit only with Y/N contexts).

‘Either’ condition	Theory	Answer Condition			
		Disjunct↓	Disjunct↑	Yes + Disjunct	Cleft
Baseline (w/ out <i>either</i>)	Inquisitive Semantics	✓	✓		✓
	Commitment Based Discourse	✓	✓		✓
	Observations	6.61	5.66	5.29	5.96
With <i>either</i>	Inquisitive Semantics	✓	✓		✓
	Commitment Based Discourse	✓	✓		✓
	Observations	6.61	5.72	5.66	5.82

Table 5.9: Predictions and observations of Alt tune across *either* condition and answer conditions.

The baseline condition (without *either*) reveals that the observations are in line with the predictions of both theories for the top three of four answer conditions (Figure 5.5); however, even the canonical Y/N reply of “Yes + disjunct” was rated overall acceptable. As all the participants in the analysis had correctly scored infelicitous fillers low, they were shown to be using the entire scale instead of simply rating all question-answer pairs as some form of acceptable.

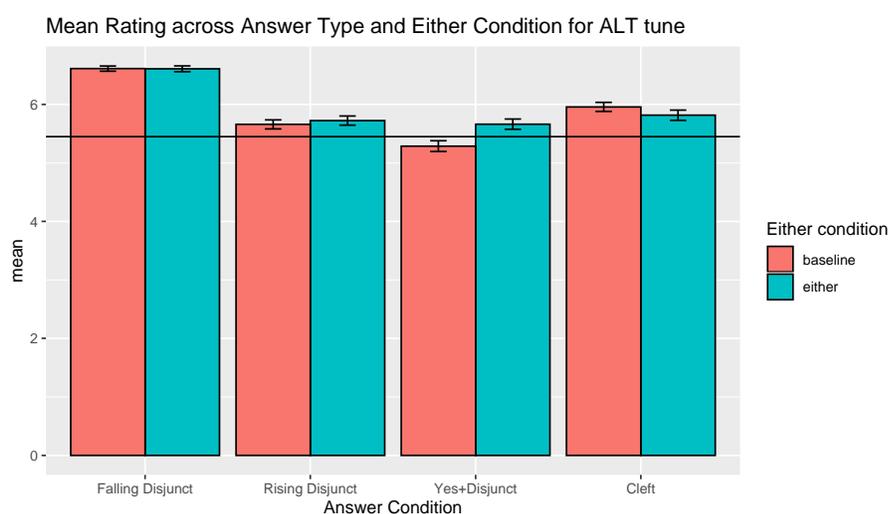


Figure 5.5: Mean rating of question-answer pairs with the Alt tune. Collapsed over answer condition and either condition. The horizontal black line separates predicted licit and illicit answers for the baseline condition.

It is possible that participants were attempting to make the question and answer context fit (Heidenreich 2014b), and therefore took the *Yes* as a confirmation of some sort as to what the questioner was asking (perhaps even confirming the ‘exactly one’ stipulation), and then took the disjunct as the response to the question. This would explain the acceptable, though not very high, rating. Also of note is that the plain disjunct answer is significantly better than the other answers in the experiment; the three bottom answers are separated by .67 points, while the top

answer is separated by .65 points with the second highest-rated answer. The ‘rising disjunct’ tone could account for a lower rating for that condition; however, the cleft answer ought to be rated just as high as a disjunct answer. A cleft answer is still a ‘disjunct’ answer, and even with the exhaustive reading attached (as it is in response to a question, c.f. Pollard & Yasavul 2015, De Veugh-Geiss et. al. 2018), the response would still fit within the ‘exactly one’ stipulation.

When *either* was added to the question, the theories again predicted the top three answers; however, here the plain disjunct answer was significantly better (.79 points) than the second highest-rated answer, and the bottom three answers were all within .16 points of each other, indicating that they were all most likely on the same level of acceptability. This differs from the baseline condition, in which each answer condition seemed to be in a tier of its own (each answer condition had at least a .3 rating separation). Even if *either* were to sway licit answers to Y/N interpretation, the two-tiered observations would not match. Rather, they seem to indicate that the answer condition that goes with *both* interpretations is highest-rated, while the answer conditions that only ought to go with *one* interpretation are rated acceptable, but not as good. This indicates that inserting *either* into an alternative tune may allow typically illicit answers, such as “Yes + disjunct”, to be acceptable (c.f. Figure 5.5, where the horizontal black line seems to separate predicted licit vs. illicit answers for the baseline condition, and that the “Yes + Disjunct” answer w/ *either* is now above that line. This was also a significant interaction in the regression model ($p < 0.05$). Recall as well in Chapter 4.3, via the discussion on *either* in the production experiment, that some participants may allow *either* in only Y/N settings, while other participants may allow *either* in only ALT settings; this could contribute to the lower ratings of answers with a single interpretation. Hence, while *either* does not seem to change the interpretation of the *question* when inserted into an utterance with an alternative tune, it does

alter the range of potential *answers*, allowing for more acceptability for answers that are licit for only Y/N contexts.

5.5.2 Y/N Questions

The predictions for Y/N question-answer acceptability for both theories depends on how the Y/N question itself was phrased: Inquisitive Semantics will allow the Yes + disjunct answer only if the question has one phrase instead of two; Commitment Space Discourse will allow the Yes + disjunct answer, and the plain disjunct and rising disjunct answers only if both disjuncts have rising accents- otherwise, all answers ought to be prohibited. With the addition of *either*, the predictions do not change. Table 5.10 reveals the predictions and observations across *either* condition and answer conditions for Y/N tunes.

'Either' condition	Theory	Answer Condition			
		Disjunct↓	Disjunct↑	Yes + Disjunct	Cleft
Baseline (w/ out <i>either</i>)	Inquisitive Semantics	✓	✓	✓	
	Commitment Based Discourse	✓	✓	✓	
	Observations	6.43	5.74	6.39	5.21
With <i>either</i>	Inquisitive Semantics	✓	✓	✓	
	Commitment Based Discourse	✓	✓	✓	
	Observations	6.46	5.83	6.34	5.36

Table 5.10: Predictions and observations of Y/N tune across *either* condition and answer conditions.

The observations indicate that both the plain disjunct and Yes + disjunct answer conditions were very acceptable to participants, regardless of the presence or absence of *either*. The rising disjunct was significantly less acceptable (over half a point difference in both *either* conditions between this answer condition and the next highest ranked answer condition), though

it was significantly more acceptable than the cleft answer condition (Figure 5.6). Again, however, even answer conditions predicted to be unacceptable were rated ‘slightly acceptable’; perhaps another example of participants attempting to fit the question and answer context together (Heidenreich 2014). If meaning could be derived or recovered, *despite* the current pairing of question tune and answer, then the pair was rated acceptable, but less so. It may be the case that participants are willing to allow *intonational* inconsistencies more than *lexical* ones; recall that four of the fillers from this experiment had to be removed because participants not only performed differently than predicted, but for all instances, they rated the four fillers as more acceptable than they ought to be, based on the location of contrastive prosody (c.f. Table 5.5).

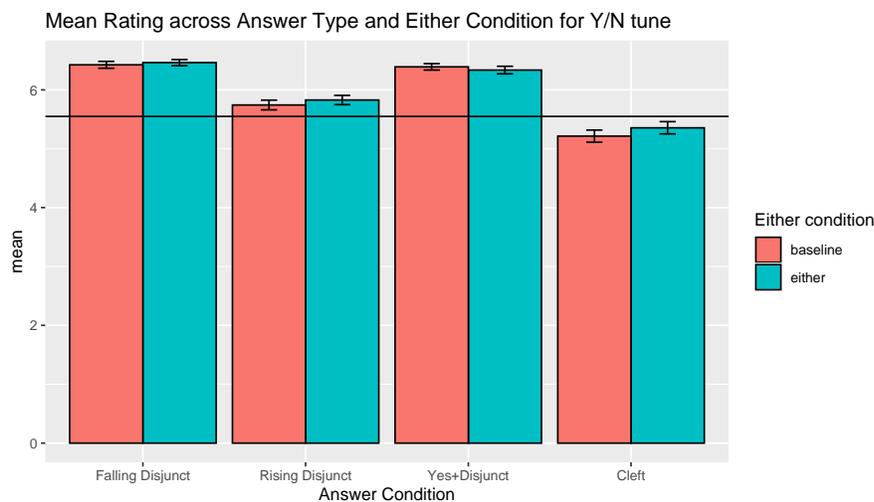


Figure 5.6: Mean rating of question-answer pairs with the Y/N tune. Collapsed over answer condition and either condition. The horizontal black line separates predicted licit and illicit answers for the baseline condition. The line also correctly predicts licit and illicit answers for the either condition.

The cleft answer (predicted to be licit for only Alt. questions) was a slightly more acceptable response when adding *either* into the question. This suggests that *either* does not sway questions nor answer conditions to a Y/N interpretation, but rather than *either* might introduce both types of interpretations when determining licit answers. Perhaps in a way, *either*

introduces some uncertainty itself; the rising disjunct response was the only answer condition that had more acceptable ratings when *either* was present vs. absent across both question tunes.

Thus the observations will fit the predictions, only if the Y/N questions contained only one phrase (as predicted by Inquisitive Semantics) and the Y/N questions had rising accents on both disjuncts (as predicted by Commitment Space Discourse). A post-hoc analysis of the stimuli used in the experiment revealed that only one Y/N question had a phrasal break (trial number 18, Y/N question type with *either*); therefore, the observations fit the predictions for Inquisitive Semantics⁴⁶. However, there were several Y/N questions that did not have a rising accent on both disjuncts (as defined by a L*+H accent on the disjunct or a L* accent on the disjunct followed by a H-).

Using only the Y/N question tune ratings (with and w/out *either*), an ordinal regression was performed with the independent variables of answer type (all four answer conditions) and presence/absence of rising accents on both disjuncts (with binary values 1 and 0, with 1 being rising accents on both disjuncts while 0 indicates the lack of rising accents on both disjuncts). There were at least 46 tokens for each answer type vs. rising accent cross-factor condition⁴⁷. The ordinal regression found the rising accent factor to not be significant ($z = 0.231$, $p = 0.817$, Table 5.11), nor any interaction between this factor and any of the answer conditions (Table 5.12), perhaps because of the disparity in token counts for those questions without rising accents vs. those questions with raising accents. There were only 46-49 tokens without rising accents for each answer condition; however, each answer condition had 473-476 tokens with rising accents.

⁴⁶ There were not enough data points to consider an analysis between phrase breaks (1 vs 2) within the Y/N question tune

⁴⁷ The ratings were collapsed over the *either* condition since Commitment Space Discourse offers the same predictions with or without *either*.

A comparison of these models against a model without the rising accent factor revealed that the more complex models were not significantly better ($p=0.82$).

	Estimate	Std. Error	z value	Pr(> z)
answerType2	-2.09788	0.301	-6.97	3.18e-12
answerType3	-0.37943	0.27435	-1.383	0.167
answerType4	-3.08488	0.3083	-10.006	<2e-16
risingDis1	0.05365	0.23181	0.231	0.817

Table 5.11: Regression model statistics for Y/N data. R code: `enteredResponse~answerType + risingDis + (1+answerType | participantID) + (1 + answerType | trialNumber)`.

	Estimate	Std. Error	z value	Pr(> z)
answerType2	-2.77635	0.59418	-4.673	2.98E-06
answerType3	-0.5407	0.62714	-0.862	0.389
answerType4	-3.07489	0.65429	-4.7	2.61E-06
risingDis1	-0.18631	0.46026	-0.405	0.686
answerType2:risingDis1	0.74699	0.56621	1.319	0.187
answerType3:risingDis1	0.18359	0.62022	0.296	0.767
answerType4:risingDis1	-0.01424	0.63773	-0.022	0.982

Table 5.12: Regression model statistics for Y/N data with interaction. R code: `enteredResponse~answerType * risingDis + (1+answerType | participantID) + (1 + answerType | trialNumber)`.

Neither the *risingDis* factor nor its interaction with the answer conditions were significant.

Even so, the ratings (as seen in Figure 5.7, next page) would not fully support what Commitment Space Discourse predicts; Commitment Space Discourse would predict high ratings for Y/N questions with raising accents in combination with answer conditions 1-3 (blue bars), and low (not acceptable) ratings for Y/N questions without raising accents on both disjuncts in combination with answer conditions 1-3 (red bars). Additionally, answer condition 4 ought to be considered unacceptable for any Y/N tune. While the cleft condition indeed has the lowest ratings, the ratings for the falling disjunct answer condition are *higher* when both disjuncts don't have raising accents; the ratings for the Yes + disjunct answer are exactly the same, regardless of accent type on the disjuncts. Only the rising disjunct answer condition adheres to the prediction by Commitment Space Discourse. However, a more complete study that gathers more tokens of

Y/N questions without raising accents on both disjuncts would need to be done in order to fully refute the predictions offered by Commitment Space Discourse.

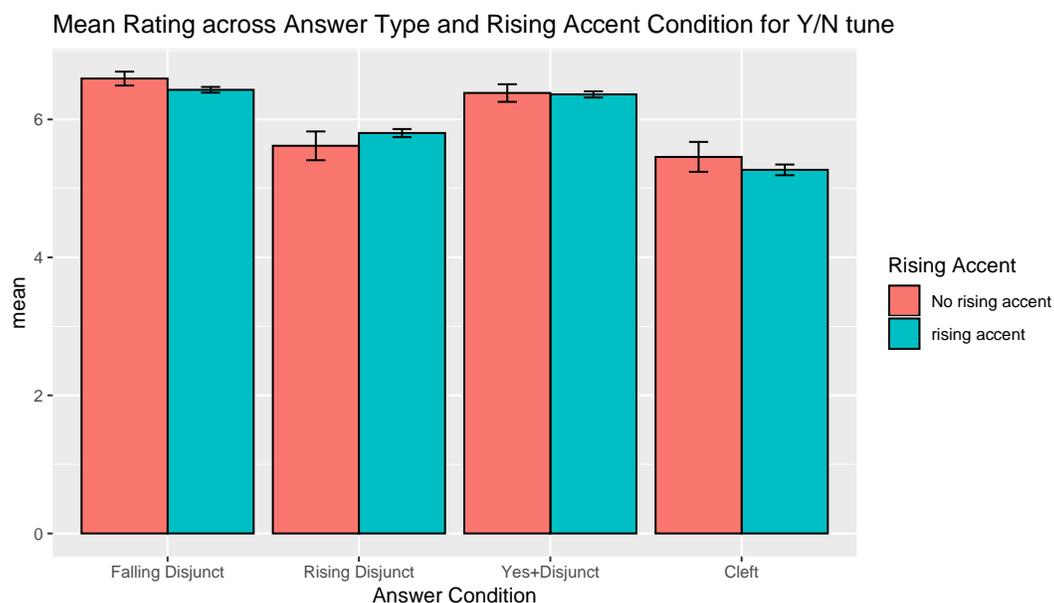


Figure 5.7: Mean rating for Y/N tune questions. Collapsed across answer condition and presence/absence of rising accent. The rising accent factor was not significant, and no significant interaction between this factor and any of the answer condition levels.

Taking away the tokens that do not have raising accents on both disjuncts does not drastically alter the overall average ratings across answer condition; Table 5.13 compares the ratings from all Y/N questions and those that only have raising accents on both disjuncts.

		Answer Condition			
		Disjunct↓	Disjunct↑	Yes + Disjunct	Cleft
All Y/N questions	Baseline	6.43	5.74	6.39	5.21
	<i>either</i>	6.46	5.83	6.34	5.36
Raising Accent on Disjuncts	Baseline	6.41	5.79	6.41	5.21
	<i>either</i>	6.45	5.82	6.32	5.33

Table 5.13: Comparison of average ratings when looking at all Y/N questions vs. those with raising accents on both disjuncts only, across either condition and answer conditions.

The hierarchy of acceptability remains the same when eliminating stimuli that don't have raising accents on both disjuncts; the simple disjunct answer had the highest ratings, followed closely by the Yes + disjunct answer. The rising disjunct answer was considerably lower, but the cleft answer was the lowest by nearly half a point. Thus, Commitment Space Discourse correctly predicts the top three acceptable answers, although it does not account for differences in ratings between these answers (why is the plain disjunct answer more acceptable than the rising disjunct answer?) and does not account for the variability in ratings when the disjunct location is moved around in the utterance.

5.6 Conclusion

Like the text experiment, the perception experiment found that the location of the disjunct affected the acceptability of the question-answer pair; however, the two experiments differed in terms of which disjunct location received the lowest acceptability ratings. The text experiment had a linear relationship, in which ratings were lowest overall when the disjunct occurred at the end of the utterance, and highest when the disjunct occurred at the start of the utterance (Figure 4.2). The perception experiment revealed that the middle location had significantly worse ratings than either the start or end locations (Figure 5.1); the start and end locations received comparable acceptability ratings. It remains unclear why participants would prefer the utterance-final disjunction when listening to conversation, but not prefer it when reading text (it was rated the worst disjunct location in the text experiment). Additionally, the answer condition, question tune, and even interaction with *either* were significant in determining acceptability of the question-answer pair.

The perception experiment tested the previously held assumption that disjunctive questions have a grammatical distinction available: inserting *either* into the question will

disallow the Alt. question interpretation. Experimental data refuted that assumption, revealing that inserting *either* into an alternative tune was acceptable. The tune contour alone determined question type. While inserting *either* did not change the interpretation of the question, it did allow more flexibility in acceptable answers. That is, while *either* can be freely inserted into both Y/N and alternative contexts and not alter the interpretation of said question, the presence of *either* does open up a wider range of potential answers, permitting answers that typically are only licit for the other context. Adding *either* into an alternative tune increased acceptability of the rising disjunct answer and the Yes + disjunct answer; adding *either* into a Y/N tune increased acceptability of the cleft answer. The very fact that participants not only accept *either* in an alternative tune, but also keep the alternative interpretation, goes against the currently held grammaticalization assumptions of disjunctive questions (Huddleston 1994, Haspelmath 2007). However, it simplifies the work of both semantic theories, as they don't need a mechanism to allow *either* to change the meaning of the question; instead, the presence of *either* merely needs to open up the possibility of additional acceptable answers.

The assumptions of licit answer conditions were also tested in this experiment. Specifically, a 'disjunct' response ought to be valid for an Alt. question, and neither intonational contour nor syntactic nature of that response is restricted. Evidence from it-cleft experiments (De Vaugh-Geiss et. al. 2018) suggest that clefts ought to be acceptable for only Alt questions; using a continuation rise on a disjunct response ('rising disjunct response') would signal a 'forward reference' (Pierrehumbert & Hirschberg 1990) that should only be compatible with Y/N questions. The data reveal that cleft answers are rated the least acceptable in response to Y/N questions, and Yes + disjunct responses and rising disjunct responses are rated the least acceptable for Alt. questions; however all question-answer pairs averaged a moderately

acceptable rating (despite the fillers indicating that participants were using the entire rating scale). It is possible that participants allow more flexibility in terms of intonation when trying to fit a question and answer into context; the filler analysis revealed that participants were willing to overlook incorrect utterances of contrastive prosody in order to fit a question and answer together. A previous study on disjunctive questions (Heidenreich 2014b) supports this analysis.

Both Inquisitive Semantics and Commitment Space Discourse offered predictions concerning licit answers, and the experiment confirmed that those answers were the most acceptable; however, the theories fail to account for the variability in acceptance between licit answers, as well as why certain question-answer combinations are more acceptable with *either*. They also did not correctly predict that all answers would be considered at least slightly acceptable, regardless of question type. Like the text experiment, these results again suggest that the location of the disjunct in the question affects the acceptability of the answer; neither theory offers insight into this phenomenon. Furthermore, it is possible that the intonational conditions required for Y/N type questions by Inquisitive Semantics and Commitment Space Discourse might not be accurate. Participants scored question-answer pairs in ways not predicted by Commitment Space Discourse when the question did not have rising accents on both disjuncts; however there were too few data points to refute the claim. Inquisitive Semantics demands that the question not have a phrasal break between disjuncts in order to allow responses with *Yes* (either bare or with a disjunct). There was but one stimuli in the experiment that did have a phrasal break; therefore the data cannot confirm or deny the claim.

Overall, the data support the notion that licit responses to disjunctive questions must be restricted by intonation, just like their question counterparts. The syntactic nature of the response also affects acceptability. It is not the case that any licit response for an Alt. question is

also like for its Y/N question counterpart, which brings the relationship between the two interpretations up for debate. Alt. questions are typically seen as a specific subset of Y/N questions (Haspelmath 2007), which is why their licit responses were also seen as a subset of Y/N question responses. The next chapter (Chapter 6), detailing the artificial language experiment, aims to explore this relationship and shed light on the assumption.

Chapter 6: Artificial Language Learning Experiment

6.1 Introduction

English is not unique in disambiguating Alt. questions and Y/N questions by intonation; however, there are other methods by which languages disambiguate disjunctive questions. Some languages have two different disjunctive connectives: one is used only in questions (interrogative disjunction) and carries the Alt. question interpretation, and the other is used in questions or declaratives (standard disjunction) and carries the Y/N question interpretation (when in questions, Haspelmath 2007). Furthermore, if a language has an interrogative connective, then it also has a standard connective; the reverse is not true (Mauri 2008). The restriction of the interrogative connective by utterance type and presence of a standard connective suggests a subset/superset relationship. However, some claim a complementary relationship between the two (Winans 2012). Understanding the nature of the relationship between the two semantic interpretations will facilitate development of an accurate framework.

6.2 Methodology

6.2.1 Design and Materials

The experiment was designed as a typical ‘artificial language paradigm’ experiment, which is composed of three phases: an exposure phase, a feedback phase, and a testing phase (Wilson 2006). The participant is expected to learn part of artificial language; each phase increased in difficulty in terms of what they needed to recall or apply about the learnt language. The experiment had an additional training phase before the exposure phase in order to introduce the participant to some key words in the experiment.

The words used in the artificial language experiment were taken from the ARC nonword database (Rastle et. al. 2002), limiting word choices to 4 letters for all nouns and 3-4 letters for

“yes”, “no”, “both”, “neither”, as well as every variation across conditions on the word “or”.

The only word(s) that changed between conditions were “or”; participants saw the same names for each animal across conditions, as well as the same words for “yes”, “no”, “both”, and “neither”. Additionally, participants were introduced to “yir” and “yirde” across all conditions, though they were not told the direct translation for either. Table 6.1 lists the words used in the experiment, their meaning, when the words were introduced, whether the participant was told the direct translation, and how many times the participant saw the word (in the exposure, feedback, and testing phases combined⁴⁸). The nouns were counterbalanced by ‘old’ nouns (introduced in training/given translation) and ‘new’ nouns (introduced in testing/not given translation).

Responses to questions were counterbalanced according to question type; for Alt questions, responses were the first disjunct listed half the time, and second disjunct the other half; for Y/N questions, responses were split evenly between four options: “Yes, both”, “No, neither”, “Yes, [first disjunct]”, and “Yes, [second disjunct]”. This accounts for the higher number of times “Yes” was seen in comparison to the other three particles. Each type of response to a Y/N questions was seen 10 times in the experiment (2 in the exposure phase, 4 each in the feedback and testing phases).

⁴⁸ This does not include instances where the participant had to supply the particular word on their own; i.e. it includes both questions and responses in the exposure phase but does not include responses that the participant had to provide in the feedback or testing phases. Additionally, it counts by stimuli tokens, and not within tokens (e.g. if the word occurred in both the question and the answer in the exposure phase, so the participant was given the word twice, the word is only counted once in the table).

Word	Meaning	Introduced	Given translation	# seen in question	# seen in response
sess	frog	training	yes	36	8
kurf	monkey	training	yes	36	8
rooc	jaguar	training	yes	36	8
lish	snake	training	yes	36	8
mun	bat	testing	no	4	1
eeph	turtle	testing	no	4	1
dwos	lizard	testing	no	4	1
rass	porcupine	testing	no	4	1
taz	yes	training	yes	0	18
twis	both	training	yes	0	6
omp	no	training	yes	0	6
fuf	neither	training	yes	0	6
yir	It is	training	no	4	0
yirde	Is it	training	no	84	0

Table 6.1: Words used in the experiment and counterbalancing measures

The training phase began the same across all conditions: participants learned the same four nouns, as well as the words for “Yes”, “No”, “Both”, and “Neither”. Next, they were introduced to phrases (Figure 6.1); they were not told what “yir” means directly. They were given four examples of this type of phrase, one for each animal.

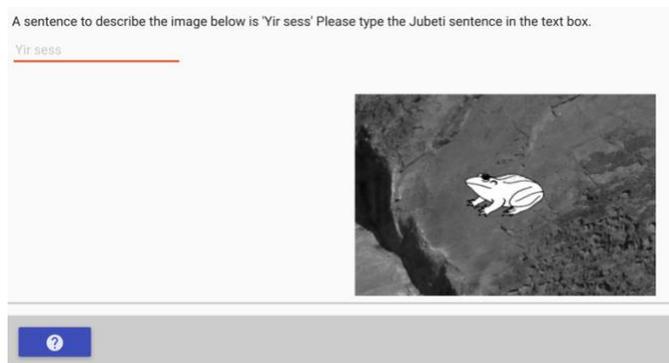


Figure 6.1: Introduction to declarative phrase in training

Then, they were introduced to a question-answer pair (Figure 6.2). They were given four examples of this type of question-answer combination; two with a “yes” response (the picture matched the animal name in the question) and two with a “no” response (the picture did not match the animal name in the question). This was the end of the training phase.



Figure 6.2: Introduction to interrogative phrase in training

In the exposure phase, they were introduced to the type of question-answer pairs they saw for the remainder of the experiment (Figure 6.3).



Figure 6.3: Exposure phase, alternative question example

Note that they were never told what the disjunctive connective means (either “ven” or “poz”, depending on the condition); the participants were meant to learn the meaning based on the context. There were sixteen stimuli like this; eight alternative question contexts and eight Y/N question contexts. The alternative question contexts always had only one animal image; the Y/N question contexts had 0-4 animal images, depending on the response. The alternative question response was one of the two disjuncts (counterbalanced between the first and the second disjunct); the Y/N question response was counterbalanced between the following four answers: “Yes, both”, “No, neither”, “Yes, [disjunct 1]”, and “Yes, [disjunct 2]”. The participant saw each animal as an answer an equal number of times (3 each; plus 2 “Yes, both” answers and 2 “No, neither” answers).

The feedback phase introduced the two types of question-answer pairs the participant needed to complete during the remainder of the experiment; the first type (Figure 6.4) had a complete question and required a response, based on the image provided. This was the same type of question seen in the exposure phase.

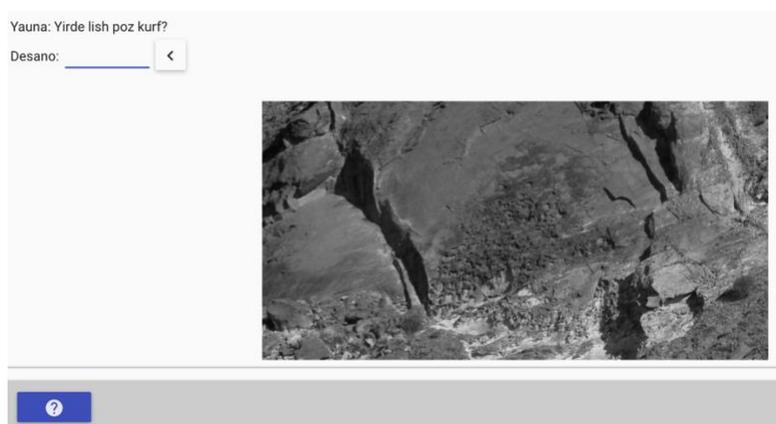


Figure 6.4: Feedback phase, response completion example

The second type (Figure 6.5) had an incomplete question (it was missing the “or”) and a complete answer, thus requiring the participant to complete the question correctly based on the answer and the image.

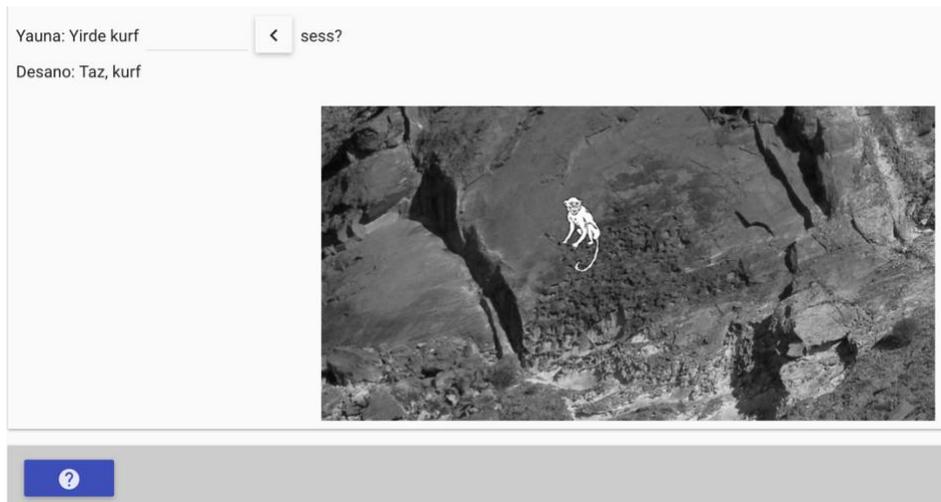


Figure 6.5: Feedback phase, question completion example

The experiment aimed to gauge how easy it is to learn each type of “or”, which means each participant not only had to learn the word(s), but also the correct type of answer, which was dependent on the image provided. This would indicate they were learning the correct meaning associated with it as well (especially in the case where the “or” word was the same for both types of disjunction). Each participant saw thirty-two stimuli; sixteen alternative question contexts and sixteen Y/N question contexts, which were further broken down into eight contexts with the response missing, and eight contexts with part of the question missing. Again, answers were counterbalanced across alternative and Y/N contexts, and each animal type was seen as an answer an equal number of times.

The testing phase was the same format as the feedback phase, only the participant was not told whether their answer was correct. Additionally, four new animals were added to the

available images, and their Jubeti names occurred in some of the questions. This was to test whether the participants were learning the ‘rules’ (i.e, alternative context vs. Y/N context) or just memorizing question-answer pairs. The testing phase was also counterbalanced according to type of question (alt. vs. Y/N), the part of the stimuli missing (response vs. “or”), the number of times a specific animal was seen as an answer (2 for each old animal type, 1 for each new animal type), and which disjunct was the answer.

Additionally, the whole experiment was counterbalanced so that the participants saw all animal names introduced in the exposure phase (‘old’ words) the same number of times in the question (36) and saw all animal names introduced in the testing phase (‘new’ words) the same number of times in the question (4). The number of times the participant saw each animal image was counterbalanced as much as possible, while keeping all other factors counterbalanced (between 28-30 times total for images introduced in the exposure phase, 8 times for each image introduced in the testing phase).

6.2.2 Language Conditions

Participants were sorted into one of the following conditions, which reflected either a possible language type observed naturally, or a language type that hasn’t been discovered yet and thus has been predicted to be impossible (Mauri 2008); It is predicted that this condition is more difficult to learn than the others. The six different conditions below each had two versions, in order to counterbalance the possible “or” words (poz/ven), for a total of twelve different lists. Each participant only saw one list.

1. *English-like language*: same disjunctive connective for both types of disjunctions

Standard disjunction: “or” (in Jubeti: poz/ven)⁴⁹

Interrogative disjunction: “or” (in Jubeti: poz/ven)

EXAMPLE: (standard and interrogative disjunction: “poz”)

Standard: “Yirde sess poz kurf?”

Gloss: *Is-there* *frog* **or** *monkey*

Interrogative: “Yirde sess poz kurf?”

Gloss: *Is-there* *frog* **or** *monkey*

2. *Finnish-like language*: A different disjunctive connective for standard disjunction vs. interrogative disjunction

Standard disjunction: “tai” (in Jubeti: poz/ven)

Interrogative disjunction: “vai” (in Jubeti: ven/poz)

EXAMPLE: (standard disjunction: “poz”; interrogative disjunction: “ven”)

Standard: “Yirde sess **poz** kurf?”

Gloss: *Is-there* *frog* **or-standard** *monkey*

Interrogative: “Yirde sess **ven** kurf?”

Gloss: *Is-there* *frog* **or-interrogative** *monkey*

⁴⁹ These were not in free variation: there was one condition in which the participant saw “poz” for all instances of “or”, and another condition in which the participant saw “ven” for all conditions of “or”. This was to counterbalance the ease of learnability across the two words.

3. *Somali-like language*: A different disjunctive connective for both disjunctions; but the standard disjunct occurs before each proposition (bisyndetically)

Standard disjunction: “amá” before each disjunct (Jubeti: poz..poz/ven...ven)

Interrogative disjunction: “misé” (Jubeti: ven/poz))

EXAMPLE: (standard disjunction: “poz...poz”; interrogative disjunction: “ven”)

Standard: “Yirde poz sess poz kurf?”

Gloss: *Is-there or-standard frog or-standard monkey*

Interrogative: “Yirde sess ven kurf?”

Gloss: *Is-there frog or-interrogative monkey*

4. *Korean-like language*: A disjunctive connective for standard disjunction; interrogative disjunction only marked via question particles

Standard disjunction: “animyen” (Jubeti: poz/ven))

Interrogative disjunction: overt question particle (“-kka”) or interrogative mood marker (“ni”) (expressed in both propositions) (Jubeti: “-de” on the verb “yir”)

Note that the interrogative mood marker (“-de”) appears for any interrogative; therefore, it occurs in both question types, but only the standard disjunction has a connective.

EXAMPLE: (standard disjunction: “poz”; interrogative disjunction: [BLANK₅₀])

⁵⁰ Participants had to choose to shrink the text box to indicate that no disjunctive connective ought to go there. This option was available across all conditions, but only conditions 4 and 5 utilized this condition (participants saw it multiple times in the exposure and feedback phase).

Standard: “Yirde sess poz yirde kurf?”⁵¹

Gloss: *Is-there frog or-standard is-there monkey*

Interrogative: “Yirde sess yirde kurf?”⁵²

Gloss: *Is-there frog is-there monkey*

5. **Unattested language:** A disjunctive connective for interrogative disjunction; standard disjunction only marked via particles. Mauri (2008) was unable to find any such language and predicted that such a language does not exist. We treated this as the opposite disjunction delineation of condition 4 but with the same restrictions, in that there is no NP disjunction but only clausal disjunction.

Standard disjunction: particle attachment on disjuncts (Jubeti: “-de”)

Interrogative disjunction: disjunction connective word (Jubeti: “poz/ven”)

EXAMPLE: (standard disjunction: [BLANK]; interrogative disjunction: “poz”)

Standard: “Yirde sess yirde kurf?”

Gloss: *Is-there frog is-there monkey*

Interrogative: “Yirde sess poz yirde kurf?”

Gloss: *Is-there frog or-interrogative is-there monkey*

6. *Mangarayi-like language:* No word for either type of disjunction connective; irrealis markers use to indicate “alternatives”.

⁵¹ It is impossible to have NP disjunction in these languages, so the disjunction is clausal (Mauri 2008, Yoon 2010).

⁵² “Korean allows sentential elements that are predictable from the discourse context or situation to be omitted (Sohn 1994, p. 266)”

Standard disjunction: the adverb “maṅaya”, meaning ‘perhaps’, before each disjunct⁵³
 (not considered a disjunctive connective because it also occurs when only one alternative is
 presented, e.g. “Is there a frog?”) (Jubeti: “poz...poz/ven...ven”)

Interrogative disjunction: the adverb “maṅaya” (Jubeti: “poz...poz/ven...ven”)⁵⁴

EXAMPLE: (standard and interrogative disjunction: “poz...poz”)

Standard: “Yirde poz sess poz kurf?”

Gloss: *Is-there perhaps frog perhaps monkey*

Interrogative: “Yirde poz sess poz kurf?”

Gloss: *Is-there perhaps frog perhaps monkey*

Note that this will alter the exposure phase examples of a single disjunct question, e.g
 “Yirde sess” in this condition will instead be “Yirde poz sess” *Is-there perhaps frog?*

Thus, the factors manipulated across these language conditions were the number of distinct
 disjunctive connectives (1 vs. 2) and the number of times each disjunctive connective (or
 equivalent) was seen across both standard and interrogative types (0, 1, or 2; cf. Table 6.2)

Language Condition	# distinct connectives	# times standard connective seen	# times interrogative connective seen
Condition 1	1	1	1
Condition 2	2	1	1
Condition 3	2	2	1
Condition 4	1	1	0
Condition 5	1	0	1
Condition 6	1	2	2

Table 6.2: Delineation of each language condition according to # of distinct connectives and type of construction of connective (juxtaposition = 0; monosyndetic = 1, bisyndetic = 2).

⁵³ This condition and condition 3 had a blank space before each disjunct (since at least one type of disjunction required a bisyndetic construction). Participants had to shrink the text box in order to indicate that no connective ought to go there.

⁵⁴c.f. Merlan (1982:33) who points out in Mangarayi there is no overt expression which directly corresponds to English *or*, neither at the phrase level nor at the clause level. Consideration of alternatives is often expressed by sequences with maṅaya *perhaps*.

6.2.3 Procedure

There was no sound associated with this experiment; all prompts were text only. Participants were told that an ancient language (named “Jubeti”) was discovered in some caves in the Amazon, and the pictures accompanying the text have been used to decode the language. A friend, Yauna, has asked for the participants’ help in learning and decoding the language. While punctuation and capitalization were not important in the experiment, spelling was important. Participants were instructed to pay careful attention to spelling. They also were allowed to take notes, if they wished to do so. In the training phase, Yauna taught them words she already knew (four nouns, “Yes”, “No”, “Both”, “Neither”), and introduced them to simple statements (“Yir sess”) and simple questions (“Yirde sess?”). They had to type in the word Yauna gave them; they were not allowed to continue until they correctly typed the word.

In the first phase (exposure phase) participants were introduced to the question/answer structure. They were told that another archaeologist, Desano, was providing the answer to Yauna’s question, based on the image provided. In the exposure phase, they merely typed Desano’s response in the grey text box (his response was given as greyed out text in the text box). They had to type the correct answer in order to proceed.

In the feedback phase, they were asked to type either disjunction connective(s) or a response, based on the pictures and text available. After the participant has input a response, a checkmark would appear if they got it correct; if they did not enter the correct response, an ‘X’ would appear, alongside the full, correct question-answer pair. They proceeded in the experiment whether or not they got the answer correct. In the testing phase, they were asked to type disjunction connective(s) or answer, based on the pictures and/or text available. They were given no feedback in this phase.

6.2.4 Participants

The artificial language experiment was conducted on Amazon Mechanical Turk (AMT) with requirements of greater than 90% HIT approval, US location and at least one previous HIT completion. One hundred twenty-three (123) participants completed the experiment. The experiment cycled through which list the next participant saw, so that at least eight participants completed each list. The experiment ran from 10-25-18 until 11-28-18. The entire experiment took 10-25 minutes. Participants also filled out a brief survey which asked their age, gender, ethnicity, and native language(s). They also filled out a survey at the end, which asked if they knew the meaning to the words “yirde”, “poz”, and “ven”, as well as if they took notes, and if they could guess what the experiment might be about. They were paid \$3 for their participation.

6.3 Predictions

The testing phase involved two fundamentally different tasks: in the first type of question-answer pair, the answer was given and the participant had to insert the correct disjunctive connective(s) into the incomplete question. The second task was involved typing in a full response based on the complete answer. These two tasks offer different predictions, depending on the relationship between Alt. questions and Y/N questions.

6.3.1 Superset/Subset Relationship

Haspelmath (2007) has argued that Alt. questions are a specific subset of the more general Y/N questions, based on the disjunctive connective that is used in each of the two question types. An interrogative disjunction can only be used in an Alt. question, and thus only ever appears in interrogatives; however, the standard disjunction is the disjunction seen in Y/N questions, as well as in declaratives. The intonational patterns in English support a

superset/subset relationship: the production experiment revealed a specific pattern than all Alt. questions must follow to be considered an Alt. question, while Y/N questions were allowed more variation. Previous studies have shown that the subset case is easier to learn than the superset case (Heidenreich 2014a), when holding the type frequency of each case constant. If this is the type of relationship between the two types of disjunction, then we would expect the following from the experiment:

1. Overall, participants will be better at learning the Alt. disjunctive connective (the subset) vs. the Y/N disjunctive connective (the superset) if they are not realized the same way, and they will be better at learning responses to Alt. questions vs. learning responses to Y/N questions.

2. A language condition in which an Alt. question is signaled using a subset of its Y/N question counterpart will be easier to learn than one where the Alt. question is not a subset; i.e. Language condition 4 (81) will be easier to learn than language conditions 2, 3, and 5, as condition 4 expresses its Alt. question as a lexical substring of its Y/N question counterpart.

(81) Language condition 4

Y/N: “Yirde sess poz yirde rooc?”

Alt: “Yirde sess [] yirde rooc”?

3. A language condition in which an Alt. question is signaled with a more compact construction (i.e. monosyndetic) than its Y/N question counterpart (i.e. bisyndetic) will be easier to learn than when the two are seen in the same construction; e.g. language condition 3 will be easier to learn than condition 2.

Note that conditions 1 and 6 are both ambiguous lexically between the two semantic interpretations; we would expect that this lack of distinction would make them the easiest to

learn (participants don't have to learn any distinction when completing the question). Thus, we might expect the language conditions to be ranked from easiest to learn to hardest to learn as follows:

Conditions 1 and 6; Condition 4, Condition 3, Condition 2, Condition 5

These predictions ought to be borne out in the question completion task; when participants have to respond to an Alt. or Y/N question, the only prediction is that Alt. questions will have higher response accuracy than Y/N questions.

6.3.2 Complementary Relationship

Winans (2012) has argued that Egyptian Arabic exhibits a complementary relationship between its standard disjunction, *aw*, and its interrogative disjunction, *wallaa*, noting that *wallaa* can actually appear in declarative sentences (83) when used in a negation of a disjunction from a previous assertion (82).

(82) Omar eind-ik aribiya aw bait.

Omar has-2sg.ma car or house.

Omar has a car or a house.

{Omar hears from across the room and shouts:}

(83) Ma.ein-ii.sh aribiya wallaa bait, eid-ii el etnain.

NEG.have-1sg car or house, have-1sg the two

*I don't have a car **or** a house, I have both!* (Winans 2012, 19)

A complementary relationship between the two disjunctive types would offer the following predictions:

1. Neither disjunct type will be significantly easier to learn than the other, in both the question completion task and the response task.

2. Language conditions 4 and 5 ought to be equally easy to learn, as one is the ‘mirror’ of the other.

3. A language condition that has both disjunctive connectives in the same type of construction (e.g. both monosyndetic) will be easier to learn than one in which the two connectives are in different constructions; e.g. language condition 2 will be easier to learn than conditions 3, 4 and 5.

Again, conditions 1 and 6 would be expected to have the best performance in the question completion task, but the ranking is as follows:

Conditions 1 and 6; condition 2; conditions 3, 4 and 5.

The only response task prediction is that both responses will be equally easy (or hard) to learn.

6.4 Results

6.4.1 Filtering

In order to filter out participants who completed the experiment but were not completing the task assigned, the testing phase of each participant was examined. For each error, it was determined if:

- 1) The response has a very obvious spelling mistake, which would have been the correct answer (e.g. kurff for kurf). These responses (24) were reverted to “correct”.
- 2) The response was either “omp” or “fuf” for “omp, fuf”. This translates to writing “No” or “neither” for a “No, neither” response. Both responses are sufficient alone to answer a Y/N question. These responses (36) were taken as a correct answer.
- 3) The response was simply “twis” for “taz, twis”. This translates to writing “both” for a “Yes, both” response. This was taken as a correct answer (4).

- 4) The response would have been the correct answer if the disjunct was the other type, i.e. if the response was the ALT response but the participant had the Y/N response, or vice versa. Or if the participant typed the ALT disjunction answer instead of the Y/N disjunction answer, or vice versa. This was marked as a “type error” (329). That is, the participant understood the task at hand, but input the wrong disjunctive type.

The correct answers and “type errors” were then added together to get a total number of “attempted” trials (Figure 6.6).

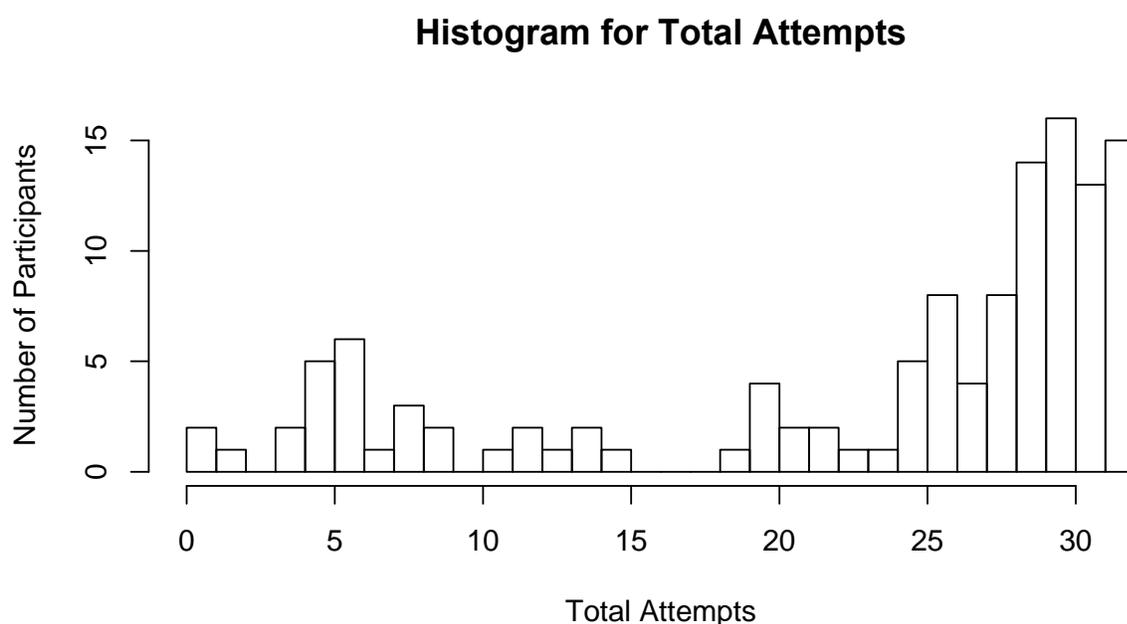


Figure 6.6: Histogram of total attempts by participant

The participants very clearly fell into one of two categories, after applying this filter: those who probably were not attempting the task at hand (15 or less attempted trials; average attempted 7.31; average correct 5.07) and those who were attempting the task at hand (19 or more attempted trials; average attempted 28.36; average correct 25.55). Data from 29 participants

were removed from the analysis in this manner⁵⁵. The remaining 94 participants' data was used in the analysis. There were at least five participants remaining from each list, at least eleven for each condition.

6.4.2 Analysis

6.4.2.1 Entire Testing Phase

A logistic-mixed effects regression was performed on the testing data of the remaining 94 participants. The dependent variable was whether the participant responded with the correct answer (1= correct, 0 = incorrect). Several independent variables were tested in the regression: the disjunct type (Alt vs. Y/N), the language condition (version), the part of the question-answer pair that was missing (“or” vs. response; this was called the “missing factor”), whether the words were old (already introduced pre-testing phase) or new, whether the pictures were old or new, and the number of pictures (0-7). Model pruning was performed, deleting factors one at a time that were not significant in the regression, or comparing two models with different factors when a model with both factors would not converge. Table 6.3 shows the estimates, standard error, z-value and p-values for the factors that contribute to the best fit model over the entire dataset. Both the type of disjunction and the missing factor were significant; there were also random slopes for participants by the missing factors⁵⁶ and random intercepts by participant and trial

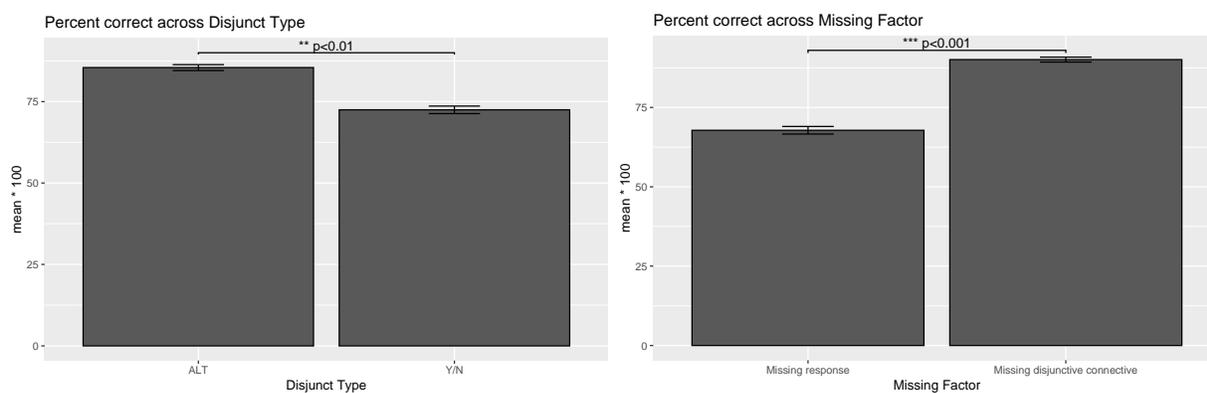
⁵⁵ It is possible that some of the 29 participants may have indeed been attempting the experiment but were unable to discern the pattern between the two disjunctive types; however, we would expect the ‘attempted’ trials to be higher, in this case, even if participants did not have a high ‘correct’ score. Very clearly some participants were not attempting the experiment; one participant often typed a variation of ‘adsfs’ into the answer, while another often clicked the arrow to leave a ‘blank’ space in a version that never had a blank space as a correct answer.

⁵⁶ The model that had random slopes for the missing Factor by participant accounted for more variance than a model that had random slopes for type of disjunction by participant ($p < 0.001$). The model that had random slopes for both factors by participants did not converge.

number. The negative estimate for disjunction type ($\beta = -1.03$, $SE = 0.32$, $z = -3.22$, $p < 0.01$) indicates that overall, participants performed better on the alternative question type than the Y/N question type (Figure 6.7). Additionally, the positive estimate for the missing factor ($\beta = 5.05$, $SE = 1.17$, $z = 4.32$, $p < 0.0001$) reveals that participants performed better when completing the question, rather than supplying an appropriate response (Figure 6.8). The significant positive intercept ($\beta = 1.74$, $SE = 0.32$, $z = 5.41$, $p < 0.0001$) reveals that participants were successful in learning the experiment (number of correct is above chance). See section 3.4.2 for the description of how the error bars in the figures were produced.

	Estimate	Std. Error	z-value	P-value
Intercept	1.7370	0.3210	5.411	6.27e-08
Disjunct Type	-1.0334	0.3208	-3.221	0.00128
Missing Factor	5.0458	1.1668	4.324	1.53e-05

Table 6.3: Best fit model for the entire dataset. Independent factors were disjunct type (Alternative question vs Y/N question) and missing factor ('or' missing or response missing). R code for model: `correct~disType+missingFactor+(1+missingFactor|participantID)+(1|trialNo)`.



Figures 6.7 and 6.8: Percent correct across the two significant factors: Disjunct Type (left) and the Missing Factor (right).

It is possible, however, that the ease in which participants learn each disjunct type differs across language conditions; however, a model that adds condition as a factor into the best fit model (above) did not converge. As the superset/subset relationship and the complementary relationship produce different predictions on the performance of the language conditions based on the task (whether the participant had to fill in the missing disjunctive connective or the missing response), the dataset was split by this “missing factor” in order to determine if the learnability differed among language conditions.

6.4.2.2 Disjunctive Connective

When the participant saw the complete response and had to simply fill in the disjunctive connective(s) or equivalent word, we see that participants indeed performed the best with condition 1 (the English-like language) and worst with condition 5 (the condition that is unattested in natural languages). Table 6.4 (next page) shows the total percent correct of the testing phase when the missing factor was the missing part of the question (the disjunctive connective), as well as how well participants performed for each disjunct type. Condition 6 was the only other condition that used a single disjunctive connective (but it occurred before each disjunct, including when the question only asked about a single noun), which may be why native English speakers scored second best in this condition. Interestingly, condition 2, which occurs in many attested languages, was the second worst percentage. Also interesting is again how differently participants performed in conditions 4 vs. 5, which were opposites of each other.

When participants had to memorize only one word for “or”, whether this was a word used in both disjunct types (condition 1, condition 6) or a word used in only one disjunct type and the other disjunct type did not have a disjunctive connective (condition 4, condition 5), participants were better at correctly identifying the interrogative (Alt. question) disjunctive connective.

However, when participants had two separate disjunct words to memorize (condition 2 and condition 3), they were better at identifying and recalling the standard (Y/N question) disjunctive connective. Table 6.4 also reveals that every condition had above 80% accuracy, except the condition that is unattested in natural languages (condition 5).

Disjunct Type	Condition 1	Condition 2	Condition 3	Condition 4	Condition 5	Condition 6
Alt	100%	83.6%	88.6%	95.8%	76.5%	99.2%
Y/N	100%	86.2%	96.6%	90.8%	73.5%	96.7%
Total	100%	84.9%	92.6%	93.3%	75%	97.5%

Table 6.4: Percent correct in testing phase when “or” was the missing factor, across condition and disjunct type.

To analyze whether these variations across condition and disjunct type were significant, a logistic-mixed effects analysis was performed with independent variables of condition (with the base level as condition 5, the condition unattested in natural language) and disjunct type (Table 6.5).

	Estimate	Std. Error	z-value	P-value
Intercept	2.94687	0.96677	3.048	0.0023
version 1	20.62102	93.77199	0.22	0.82594
version 2	0.54135	1.2471	0.434	0.66423
version 3	1.54084	1.48041	1.041	0.29796
version 4	2.80755	1.41711	1.981	0.04757
version 6	4.99272	1.73891	2.871	0.00409
disTypeY/N	-0.32014	0.40141	-0.798	0.42514
version 1: disTypeY/N	0.03401	296.53165	0	0.99991
version 2: disTypeY/N	0.65762	0.5757	1.142	0.25333
version 3: disTypeY/N	2.79717	1.11936	2.499	0.01246
version 4: disTypeY/N	-0.75622	0.74647	-1.013	0.31103
version 6: disTypeY/N	-0.7117	1.12664	-0.632	0.52758

Table 6.5: Model of best fit for testing phase data when the disjunctive connective was missing.

R code for model: $updatedCorrect \sim relevel(version, ref = "5") * disType + (1|participantID)$

The regression found that condition 4 (the Korean-like language: $\beta = 2.81$, $SE = 1.42$, $z = 1.981$, $p < 0.05$) and condition 6 (the Mangarayi-like language: $\beta = 4.99$, $SE = 1.74$, $z = 2.87$, $p < 0.01$) were significantly different from the unattested language (condition 5)⁵⁷. The participants in these language conditions performed better overall on the question completion task than the participants in the unattested language condition (Figure 6.9).

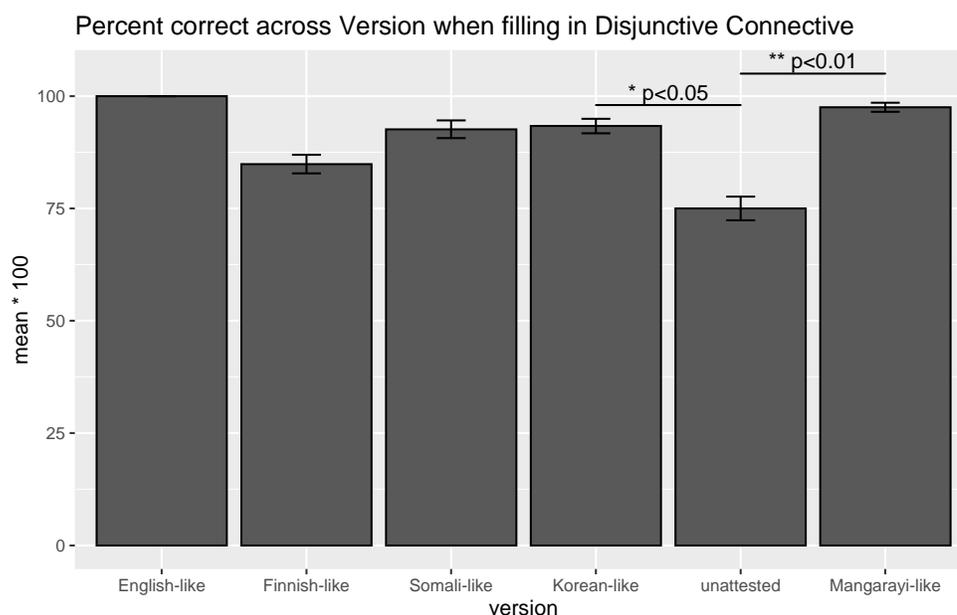


Figure 6.9: Percent correct across version when “or” was missing. See footnote 57 for the reason the English-like language is not considered significant by the model.

Participants may have performed significantly better in the conditions where they were required to memorize only one disjunctive connective, which did not change with the disjunct type (i.e.

⁵⁷ The logistic regression cannot truly say anything about condition 1, as every participant scored perfectly in this condition for this task. So the model does not list condition 1 as significant because none of these parameters distinguish a ‘correct’ answer from an ‘incorrect’ answer for that condition. However, it would obviously be considered ‘significantly’ different from condition 5, the baseline condition. A separate regression that removed condition 1 from the data revealed that conditions 6 and 4 were still significant ($p < 0.01$ and $p < 0.05$, respectively), and the interaction between condition 3 and disjunct Type was also still significant ($p = 0.01$). These were the only significant factors again.

the English-like condition and the Mangarayi-like condition). However, the fact that the Korean-like language condition is significantly different from the unattested language condition reveals that participants were better able to recall a single disjunctive connective that correlated with the Y/N question (standard disjunction) vs. the Alt. question (interrogative disjunction).

Furthermore, the analysis revealed a significant interaction between disjunct type within condition 3 (the Somali-like condition: $\beta = 2.80$, $SE = 1.12$, $z = 2.50$, $p < 0.05$). That is, participants were much better at remembering the Y/N connective in condition 3, where participants in condition 5 were better at remembering the Alt. connective (Figure 6.10). Therefore, it was not the case that across all conditions, participants were better at recalling one type of disjunction over another; rather, the nature of the two disjunctive connectives influenced how well participants were able to recall the correct connective. In the two language conditions that had two distinct disjunctive connectives (Finnish-like language and Somali-like language), participants were better at correctly recalling the Y/N connective. In all other conditions, participants were better at recalling the Alt connective.

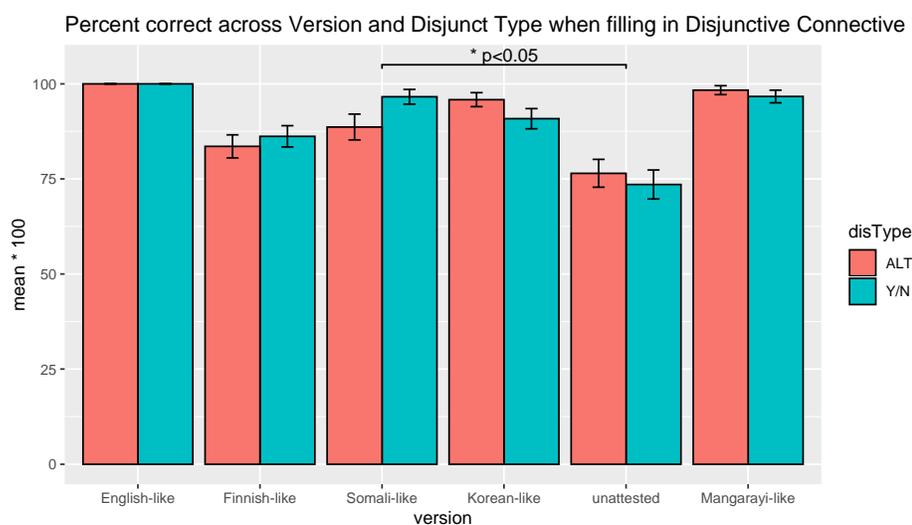


Figure 6.10: Percent correct of connective across Version and Disjunct Type

6.4.2.3 Response to Question

Participants were required to give the correct response, given a complete question, in the other half of the testing phase. We would predict here that perhaps conditions 1 and 6 would perform the worst, as both the Y/N and Alt disjunctive questions are realized the same lexically, and the participants would have to rely on image clues. Therefore, anytime there was a single animal image in the context, *either* type of disjunctive question would fit the context, and thus the context was ambiguous (this would mean every single intended alternative question context was ambiguous (8), and one intended Y/N question context was ambiguous). Conditions 2, 3, 4, and 5, however, do have different ways to realize the two disjunctive questions, and therefore do not contain any ambiguous contexts. Table 6.6 shows the average percent correct of the testing phase missing response stimuli within both disjunct Types (Alt. and Y/N), as well as the average correct across condition (Total).

Disjunct Type	Condition 1	Condition 2	Condition 3	Condition 4	Condition 5	Condition 6
Alt	80.9%	73.0%	85.2%	92.5%	81.6%	74.2%
Y/N	48.5%	59.2%	51.1%	66.7%	61.0%	40.8%
Total	64.7%	66.1%	68.2%	79.6%	71.3%	57.5%

Table 6.6: Percent correct in testing phase when the response was the missing factor. Collapsed across condition and disjunct type.

Overall, conditions 1 and 6 were indeed the worse performing; surprisingly, conditions 4 and 5 were the best performing, both achieving over 70% correct. However, there is a clear distinction here between Alt. responses and Y/N responses, with participants correctly identifying Alt. responses better than Y/N responses across all conditions. In fact, a logistic-mixed effects regression analysis reveals that here, disjunct type is significant ($\beta = -1.65$, $SE =$

0.59, $z = -2.80$, $p < 0.01$), but no condition is significant, nor any interaction between condition and disjunct types⁵⁸ (Table 6.7). Participants did not perform significantly better or worse in condition 5 than in the other conditions, though condition 4 was near significance ($z = 1.70$, $p = 0.089$)⁵⁹. These results suggest that while conditions 4 and 5 both seem learnable in terms of choosing appropriate responses, there is something inherently more difficult in learning a disjunctive connective for alternative questions when Y/N questions do not have one (condition 5), than the reverse situation (condition 4). This might explain why condition 5 is unattested in natural languages.

	Estimate	Std. Error	z-value	P-value
Intercept	2.3471	0.584	4.019	5.85e-05
disTypeY/N	-1.6452	0.5887	-2.795	0.0052
version 1	-0.3003	0.6666	-0.451	0.6523
version 2	-0.7486	0.6427	-1.165	0.2441
version 3	0.4795	0.7838	0.612	0.5407
version 4	1.2778	0.7509	1.702	0.0888
version 6	-0.8421	0.6784	-1.241	0.2145
disTypeY/N:version 1	-0.5463	0.4917	-1.111	0.2665
disTypeY/N:version 2	0.685	0.4678	1.464	0.1431
disTypeY/N:version 3	-1.1505	0.6098	-1.887	0.0592
disTypeY/N:version 4	-0.9134	0.5855	-1.56	0.1188
disTypeY/N:version 6	-0.4525	0.4936	-0.917	0.3593

Table 6.7: Model of best fit for testing phase data when the response was missing. R code for model: $updatedCorrect \sim disType * relevel(version, ref = "5") + (1|participantID) + (1|trialNo)$. This model was significantly better than one without version as a factor ($p < 0.01$).

⁵⁸ The base level was again condition 5, due to the fact that it is the unattested condition that we are comparing natural language against. Condition 4 was near significance ($p = 0.0888$), as well as the interaction between disjunct type and condition 3 ($p = 0.0592$). The best fit model had random intercepts by participant and trial number.

⁵⁹ The interaction between disjunct type and language condition 3, which was significant in the missing “or” analysis, was near significance as well ($z = -1.89$, $p = 0.059$)

Participants performed significantly better on the alternative question responses than on the Y/N question responses (Figure 6.11), which is to be expected since the alternative question responses are limited to the two disjuncts in the question; participants have a 50% chance of getting the answer right on a guess. The Y/N responses, however, are more varied; participants had four different answers to choose from (“Yes, both”; “Yes, [disjunct 1]”; “Yes, [disjunct 2]”; “No, neither”), and would have to remember the word(s) for “yes”, “both”, “no”, and “neither”. Conditions 1 and 6 had the worst Y/N percentages, although this may be expected as these conditions did not differ lexically between the alternative question and Y/N question; participants would have to rely on the images in order to correctly identify the disjunctive question type.

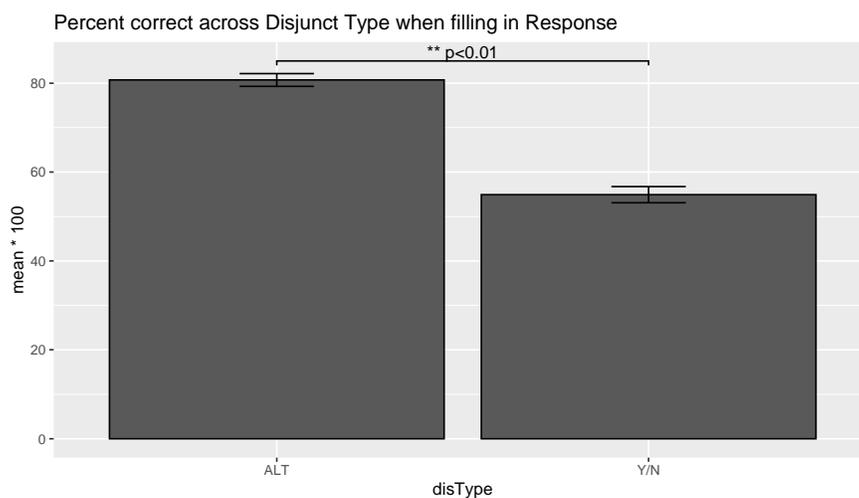


Figure 6.11: Percent correct of response across Disjunct Type

Based on the higher accuracy of alternative question responses, however, it appears that participants did associate a single image with the alternative question, thereby reducing the number of ambiguous stimuli down to a single Y/N type trial that also had a single image. Most of the participants in conditions 1 and 6 did miss this question, suggesting that they were

delineating the disjunctive questions by the number of images (they split the two types by number of images: alternative questions- one image; Y/N questions- any other number of images). Table 6.8 shows the difference in percent correct when the single ambiguous Y/N stimuli is removed from the dataset.

Y/N Disjunct Type	Condition 1	Condition 6
All data	48.5%	40.8%
W/out ambiguous trial	53.8%	44.8%

Table 6.8: Percent correct of Y/N questions when the response was missing

In fact, removing this trial from all conditions, even conditions in which the trial was not ambiguous (even though there was only one image, which could indicate either disjunctive type, the question itself in the other language conditions disambiguated between the two types) slightly increased the percent correct for all conditions (except condition 4, which had the same percent correct for this trial as for the entire Y/N dataset). This suggests that not only were participants paying attention to the number of images, but additionally were using that factor to disambiguate the types of responses, perhaps even relying on images more than lexical clues.

Y/N	Condition 1	Condition 2	Condition 3	Condition 4	Condition 5	Condition 6
Disjunct	1	2	3	4	5	6
All data	48.5%	59.2%	51.1%	66.7%	61.0%	40.8%
W/out #66	53.8%	59.4%	51.9%	66.7%	61.3%	44.8%

Table 6.9: Comparison of percent correct of Y/N questions when response was missing; full dataset vs. eliminating trial 66, which was ambiguous in terms of number of images.

Although the percentages do increase when eliminating the ambiguous trial, the Y/N disjunct type data still remains significantly lower than the Alt disjunct type data. Additionally, the condition predictions that were seen when “or” was the missing factor are not upheld by this response data; condition 5 is not the worst performing- rather, conditions 4 and 5 were the easiest

to learn in terms of response patterns. Thus, the two conditions that did differentiate between the two disjunctive question types and yet only had a disjunctive connective for one of the two types (the other type signaled by clausal juxtaposition) had the best performance in this task (Table 6.10).

	Condition 1	Condition 2	Condition 3	Condition 4	Condition 5	Condition 6
Percent correct	64.7%	66.1%	68.2%	79.6%	71.3%	57.5%
Percent correct, no trial 66	68.2%	66.7%	69.7%	80.4%	72.2%	60.4%

Table 6.10: Percent correct of testing phase where the response was the missing factor.

Updated to not include the Y/N trial that is ambiguous in conditions 1 and 6

6.5 Discussion

6.5.1 Type Errors

The artificial learning experiment revealed that participants were able to indirectly learn the two different types of disjunctive questions and their appropriate responses across different language conditions. Participants performed differently in the stimuli when they had to complete the answer with the disjunctive connective vs. when they had to enter the correct response.

When the disjunctive connective was the missing factor, the data upheld the predictions outlined for each condition: condition 5 had the worst percent correct, followed by the conditions that required the memorization of two disjunctive connectives. Conditions that had only one disjunctive connective had the highest percentage correct. It is not the case that condition 5 merely had individuals that may have put in a random answer when “or” was the missing factor;

Table 6.11 shows the percent correct across condition, as well as the percent of answers that were a type error (recall a type error is defined as an error where the participant input the correct answer for the opposite disjunctive type).

	Condition 1	Condition 2	Condition 3	Condition 4	Condition 5	Condition 6
Percent Correct	100%	84.9%	92.6%	93.3%	75%	97.9%
Percent Type Error	0%	13.5%	5.7%	5.8%	24.3%	0%
Total	100%	98.4%	98.3%	99.2%	99.3%	97.9%

Table 6.11: Percent correct, percent type error, and total percent stimuli accounted for by the two types when “or” was the missing factor, across conditions.

Condition 5 had an overwhelming number of type errors, considering it is simply the opposite of condition 4 (one disjunct type had no disjunctive connective while the other did); nearly 25% of the responses were type error responses. This would support the hypothesis that condition 5 is more difficult to learn than, say, condition 4, and explains why this type of language condition does not occur in natural languages. The fact that the percent correct and percent type error summed up to at least 97.9% for each condition also suggests that participants did understand the task, but probably misremembered the correct disjunctive connective if they were in error.

Looking closer at condition 5, there is further support that participants, for some reason, have a harder time learning this dichotomy of the two disjunction questions. There was one participant in condition 5 that somehow ended up responding as if they had learned condition 4. That is, they were supposed to learn (and were exposed to and had feedback for) a language that

had an interrogative (Alt. question) disjunctive connective, but simply had two interrogative clauses for the Y/N question; however, participant 52 actually learned the opposite- they used the disjunctive connective they learned as the *standard* disjunctive connective and they had two interrogative clauses juxtaposed for the Alt. question. Table 6.12 illustrates why participant 52, while only achieving two correct trials in the testing phase, had a total of 29 total attempts (that is, they had 27 answers that would have been correct, had the disjunctive question type been the opposite).

Missing Factor	Disjunct Type	Correct	Type error
response	Alt	0	8
	Y/N	2	3
“or”	Alt	0	8
	Y/N	0	8
Total		2	27

Table 6.12: Participant 52 data from the testing phase. Broken down into percent correct and percent type error by missing factor and disjunct type. There were 32 tokens in total.

Had participant 52 been in condition 4, they would have had 84.4% correct in the testing phase, which is near the average of 86.5% for that condition. That is, even though participant 52 was exposed to consistent stimuli in condition 5, they still learned condition 4 nearly as well as those exposed to that condition. This suggests that language evolution change may have influenced the lack of natural languages with a condition 5 dichotomy. Even when exposed to a language with only an *interrogative* disjunctive connective, participant 52 learned a language with only a *standard* disjunctive connective.

Even without participant 52’s data, condition 5 only had 77.3% correct in the testing phase; while no longer the lowest percent correct (condition 2 had 75.5% correct), it was still

considerably lower than its opposite, condition 4. There is no real reason to discount participant 52's data, either; they were clearly performing the task at hand, and their unusual testing phase data only indicates that it may be somehow inherently easier to learn the disjunctive question differentiation in condition 4 over condition 5.

It is possible that the alternative question and Y/N question relationship is one of a default vs. exception; the Y/N question is a default question, while a certain set of parameters must be met for the alternative question interpretation (in English, this translates to intonational parameters), which is restricted not only to the question itself, but also to the appropriate responses. Haspelmath (2007) has defined the relationship between the standard disjunctive connective and interrogative disjunctive connective as a subset relationship; he noted that standard disjunctive connectives can occur both in declarative and interrogative clauses, while the interrogative disjunctive connective is restricted to interrogative clauses. This relationship would explain why participants did not seem to find it difficult to learn a language with a standard disjunctive connective (while the exception was simply two interrogative clauses juxtaposed) but did find it difficult to learn a language with only an interrogative disjunctive connective (and the default was two interrogative clauses juxtaposed). They seemed inclined to associate the default with the lexical item, if only one lexical item were introduced.

This idea that perhaps alternative questions (and the connective used) are a specific subset of disjunctive questions (which are typically, or by default, polar questions, using the standard connective) is further supported by the licit answerhood conditions in Inquisitive Semantics; while a polar question allows various answers, the alternative question allows only a subset of those answers (namely, the disjuncts; Roelofsen and Van Gool, 2010). Perhaps participants associated the smaller subset of licit answers to a smaller *lexical phrase*. That is,

when the participant was given the incomplete question and full response and was required to fill in the missing “or” information, they may have associated the more restrictive responses (a simple disjunct answer) to a more restricted “or” connective.

For example, consider condition 3, which had two separate disjunctive connectives for standard disjunction and interrogative disjunction, but the standard disjunctive connective was repeated before each disjunct. This condition had a total percent correct closer to those conditions with only one lexical disjunctive connective (c.f. Table 6.9, condition 4), rather than condition 2, which had a different disjunctive connective for both types of disjunctive, but both occurred before the second disjunct alone. One would expect condition 2 and 3 to perhaps be comparable in learning difficulty, or even condition 2 easier to learn because the Y/N question had one less lexical item, but the opposite held true; participants seemed to associate more lexical items with the Y/N question, and fewer lexical items with the Alt. question. One would expect that a language that uses two separate disjunctive connectives, but the interrogative disjunctive connective is repeated before each disjunct, would be more difficult to learn as well; this language condition has not been found in any natural language.

6.5.2 Ranking Performance by disjunctive connectives

Natural languages follow the rule of economy in terms of diachronic change; ease of learnability will affect how the language evolves. Many languages only have one disjunctive connective; this connective can be used in both Alt. and Y/N questions (English-like), or as only the standard disjunction (used in Y/N questions, Korean-like), with interrogative disjunction constructed with two juxtaposed clauses. No language uses a single disjunctive connective in only Alt. questions (as the interrogative disjunction); in fact, this seems more difficult to learn. Additionally, when a language has two separate disjunctive connectives, they can each be used

monosyndetically between disjuncts (Finnish-like language), or the interrogative disjunction may be monosyndetic while the standard disjunction might be repeated before each disjunct in a bisyndetic construction (Somali-like language); however, there is no attested language that repeats the interrogative disjunctive, but not the standard disjunction.⁶⁰ This seems to suggest that there may be some inherent relationship between the number of words used for the interrogative disjunctive connective and the number of words used for the standard disjunctive connective; participants were more easily able to learn languages in which the standard disjunctive (for Y/N questions) had more lexical items than the interrogative disjunctive (for alternative questions); this is evident by the high percent correct for conditions 3 and 4 (c.f. Table 6.11)

Conditions 1 and 6 had a similar percent correct when looking only at the data where “or” was the missing factor (c.f. Figure 6.9), but the same disjunctive connective was used across both types of disjunctive questions. That is, participants weren’t really learning a distinction between the two types of questions for this part of the task; regardless of the response given, the missing “or” word(s) would always be the same. Therefore, these participants only had one distinct type to memorize for this task, which, naturally, was easier to learn than the other conditions which did delineate the two types of disjunctive question types for the task. Additionally, the fact that Condition 1 is an English-like language condition, modeled after participants’ native language, probably also contributed to its high percent correct (c.f. Figure 6.9).

⁶⁰ Mangarayi doesn’t really have a true disjunctive connective, as it occurs even before a single Y/N question, but regardless is it repeated before both disjuncts for both types of questions, not just alternative questions.

Table 6.13 supports the notion that the learnability of the condition seems to be affected by the number of distinct connective items in the system (one vs. two) as well as the number of times the disjunctive connective was repeated for each type of disjunct type (monosyndetically, between disjuncts vs. bisyndetically, before each disjunct).

Condition	% Correct, “or” factor	# words standard disjunction	# words interrogative disjunction	# distinct disjunctive connectives
1	100%	1	1	1
6	97.9%	2	2	1
4	93.3%	1	0	1
3	92.6%	2	1	2
2	84.9%	1	1	2
5	75%	0	1	1

Table 6.13: Language conditions ranked by percent % when participants had to fill in the disjunctive connective

When the participant had to fill in the missing “or”, the two best conditions (condition 1 and condition 6) always had the same answer, regardless of disjunct type. Then, it appears that the conditions performed better when they had only one distinct disjunct vs. two, and when the number of times the standard disjunctive connective occurs is more than the number of times the interrogative disjunctive connective occurs. Condition 5, unattested in natural language, was the only condition in which the number of times the standard disjunctive connective occurs was less than the number of times the interrogative disjunctive connective occurs. It was also the worst performing across the language conditions. Table 6.13 ranks the conditions based on accuracy when filling in the disjunctive connective, and highlights similarities between condition 5 and the other conditions with one distinct disjunctive connective; whereas participants were successful in learning those language conditions, condition 5 had the worst performance when it ought to have

similar performance, if only considering the number of distinct disjunctive connectives. Recall that conditions 1, 6, and 4 were considered statistically significant in comparison to condition 5 in the logistic regression analysis. Therefore, it must not be just the number of distinct disjunctive connectives that influenced performance, but also the relationship between the two connectives. Despite the disjunct type not being significant in the “or” missing factor task, the ranking in Table 6.13 is the exact ranking predicted based on a superset/subset relationship between Y/N questions and Alt. questions (c.f. section 6.3.1). Importantly, it is very different from what a complementary relationship would predict (c.f. section 6.3.2). However, the complementary relationship is based on an interrogative disjunction in a declarative in “a nonconventional use of negation” (Winans 2012, 41). Indeed, reviewing (82) and (83), reprinted below, indicates that Omar may use the interrogative disjunction in order to dismiss the presumption that the disjunction carries- the ‘exactly one’ stipulation. In languages in which the Alt. interpretation and Y/N interpretation are realized lexically, the interrogative disjunction carries the ‘exactly one’ marker that the intonational contour carries in English.

Examples taken from Egyptian Arabic:

(82) Omar eind-ik aribiya **aw** bait.

Omar has-2sg.ma car or house.

Omar has a car or a house.

{Omar hears from across the room and shouts:}

(83) Ma.ein-ii.sh aribiya **wallaa** bait, eid-ii el etnain.

NEG.have-1sg car or house, have-1sg the two

*I don't have a car **or** a house, I have both!*

(Winans 2012, 19)

Further research into the relationship between negation and interrogative disjunction is needed, however; the use of negation in disjunctive questions and the use of either standard or interrogative disjunction in declarative disjunction is beyond the scope of this dissertation.

6.5.3 Ranking Performance by Entire Testing Phase

The language conditions ranked differently when looking at the total percent correct; here it becomes more evident that the language conditions in which the standard disjunctive connective is seen more times than the interrogative disjunctive connective affect the learnability of the system; conditions 4 and 3 were two of the three languages with over 80% correct (c.f. Table 6.14); the only other language condition was condition 1, which may have performed so well because this condition matched the native language of all the participants in that condition (English).

Condition	% Correct	# words standard disjunction	# words interrogative disjunction	# distinct disjunctive connectives
4	86.5%	1	0	1
1	82.4%	1	1	1
3	80.4%	2	1	2
6	77.5%	2	2	1
2	75.5%	1	1	2
5	73.2%	0	1	1

Table 6.14: Language conditions ranked by % correct in testing phase, overall.

That is, when number of distinct disjunctive connectives is held constant, a language condition in which the standard disjunction is repeated more than the interrogative disjunction was easier for participants to learn. It is possible that people associated a question restricted to fewer lexical words with the semantic interpretation that had the corresponding more restrictive answer group: Alt. questions. This would explain Mauri's observation that "there is no language

which uses a disjunctive connective only to convey interrogative alternative relations, while standard alternatives are expressed by juxtaposition (p. 46).” From this experiment, we may hypothesize that this observation might be extended to the effect that for any language that has interrogative disjunctive connective *x*, the language must have a standard disjunctive connective *y* where *y* is the same number of lexical units or more than *x*. That is, there is no attested language in which the interrogative disjunction occurs bisyndetically while the standard disjunction occurs monosyndetically; nor would we predict one to occur.

6.6 Conclusion

The artificial language experiment was conducted to explore the relationship between Alt. question disjunction and Y/N question disjunction. Haspelmath (2007) has argued for a superset/subset relationship, while Winans (2012) has given evidence for a complementary relationship. The results from the experiment indicated a relationship along the lines of a default (Y/N) and exception (Alt.) rule, which is consistent with the results from the text experiment as well (Chapter 4). This supports Haspelmath’s hypothesis of a superset/subset relationship.

Six different language conditions were used in the experiment, which varied both the distinct number of disjunctive connectives (1 vs. 2) and their position (monosyndetic vs. bisyndetic). The language experiment encompassed two specific tasks: insert the appropriate disjunctive connective when given an incomplete question and full response; or insert the correct response when given the complete question. Participants performed better overall when having to complete the question, with the unattested language condition having the worst performance, and the percent correct was statistically significantly different from three other conditions. It was not easier for participants to learn one type of disjunction over the other in this task, however. When there were two distinct “or” words, participants performed better on the Y/N

disjunction; when there was only one distinct “or” word (but it might only occur in one disjunct type), participants performed better on the Alt. disjunction. This was not predicted by the superset/subset relationship that Y/N questions and Alt. questions were presumed to have. Possibly, participants focused on learning the connective associated with the ‘default’ when there were two different words; when there was only one distinct lexical word, participants were better at identifying the ‘exception’. Further work on learning disjunctions can shed light on these hypotheses.

Comparison of how the language conditions would rank in terms of learnability supports the superset/subset relationship over the complementary relationship. Furthermore, the response task revealed the disjunct type was significant, with participants scoring higher on Alt. responses vs. Y/N responses across all language conditions. Such a result indicates that Alt. responses are easier to learn than Y/N responses, regardless of how the language delineates the two interpretations, and is indicative of a default and exception relationship (Heidenreich 2014a). The previous evidence in Egyptian Arabic for a complementary relationship was shown to have specific interaction with polarity particles (only occur in negation). That is, the interrogative disjunction could be used in a declarative only when the disjunction occurred with negation (c.f. (83) in 6.5.2). This specific use of an interrogative disjunctive connective in the declarative may have the specific function of negating the ‘exactly one’ presumption that the interrogative disjunction carries. This is an area for future work.

Chapter 7 General Discussion

7.1 Introduction

Each experiment outlined in the previous chapters offered new insights into the theories on disjunctive questions. It is evident from the results of these experiments that certain modifications need to be made on the current theories in order to account for the experimental data. First, the predictions and observations for each experiment are reprinted below. Next, the necessary modifications that need to be addressed for each theory are discussed. Each theory already contains certain mechanisms that will allow for some of the modifications; some have already been used in conjunction with disjunctive questions, while others will be introduced into the analysis. Finally, the predictions from these modified theories will be compared to the results from the experiments. Future work will be addressed as the final conclusions from the experiments are laid out.

7.2 Theory predictions and experimental observations

The production experiment (Chapter 3) required participants to produce alternative disjunctive questions and Y/N disjunctive questions based solely on image context. The participants saw a general Y/N response for intended Y/N questions (“Yes” or “No”) and a general alternative question response for intended alternative questions (first or second disjunct). The predictions and observations are outlined on the next page in Table 7.1 (Table 3.25 from production chapter).

Factor	Question Type	Predictions			Observations	Conclusion
		Inquisitive Semantics	Commitment Space Discourse	Previous Empirical Results	Current Study	Theories supported
Final Contour	Alt	Final Fall	Final Fall	Final Fall	Final Fall	All
	Y/N	Final Rise	Final Rise	Final Rise	Final Rise	All
Prosodic Focus	Alt	Both disjuncts	Both disjuncts	Both disjuncts	Both disjuncts or whole phrase	None
	Y/N	Both disjuncts or whole phrase	Both disjuncts ⁶¹ or whole phrase	Both disjuncts ⁶¹ or whole phrase	Both disjuncts or whole phrase	All ⁶¹
Phrase Break	Alt	Phrasal Break	Phrasal Break	Yes or no, D1 ends high	Yes or no, D1 ends high	Previous results only
	Y/N	Yes ⁶² or no	Yes ⁶² or no	Yes or no	Yes or no	All ⁶²

Table 7.1: Theory predictions, observations, and conclusions from the production data.

In particular, the observed focus structure and the phrase break for both types of disjunction did not completely match the predictions from the two theories and previous empirical work. For the theories to accurately reflect the production results, the following criteria must be met:

1. Y/N disjunctive questions with focus on both disjuncts (having a phrasal break) must allow *yes* and *no* answers.

⁶¹ The theory supports focus on both disjuncts, but then does not license the responses seen in the experiment. Participants would have had to be disregarding responses.

⁶² The theory supports phrase breaks in Y/N questions, but then does not license the responses seen in the experiment. Participants would have had to be disregarding responses.

2. Alternative disjunctive questions can occur without a phrasal break, or without focus on both disjuncts
3. Alternative disjunctive questions must have the first disjunct ‘end high’

The first criterion is dependent on the participants taking Jordi’s responses into account during the experiment. It is quite plausible, however, that they were not. This will be reexamined after the results from the other experiments are discussed.

The results from the text experiment indicated that when participants are given the ambiguous disjunctive question in lexical form alone, the question is biased toward a Y/N question interpretation. The results from the artificial language experiment support this; people seem to hold alternative questions as an ‘exception’ to the ‘rule’ of disjunctive questions, where the ‘rule’ is the Y/N or polar interpretation. Thus, the results from the artificial language experiment further support the conclusion from the text experiment that participants interpret ambiguous strings in their ‘default’ interpretation, which is the polar interpretation.

Theory	Disjunct Type	Response					
		Disjunct	Yes + disjunct	Cleft	Yes	No	Yes, both
IS	Y/N: Open Complex	✓				✓	
	Y/N: Block	✓	✓		✓	✓	✓
	ALT	✓		✓			
CSD	Y/N: Two monopolar	✓	✓				✓
	Y/N: Bipolar				✓	✓	
	ALT	✓		✓			

Table 7.2: Permissible answers to alternative questions and various interpretations of Y/N questions across Inquisitive Semantics (IS) and Commitment Space Discourse (CSD).

Both IS and CSD have multiple Y/N question subtypes, which have different licit answerhood conditions (Table 7.2, previous page).

The results from the text experiment (Table 7.3) indicated that all answers had an acceptable score (a score above 4 on a 7-point scale, where 4 was neutral, 1 was very unacceptable, and 7 was very acceptable), which means that participants also must have considered ‘non-default’ interpretations (e.g. alternative questions, other interpretations of Y/N questions). Based on the predictions in Table 7.2 and the observations in Table 7.3, it seems that participants were, by default, interpreting the questions as two monopolar questions (under CSD) or as a Block Y/N interpretation (under IS).

Tier	Answer Conditions					
	Disjunct	Yes, disjunct	Cleft	Yes	No	Yes, both
Borderline Acceptable (4-4.99)				4.26		
Quite Acceptable (5-5.99)			5.53		5.71	
Very Acceptable (6-7)	6.67	6.48				6.57

Table 7.3: Average rating across answer condition, categorized into tiers of acceptability

The predictions offered by CSD’s two monopolar questions perfectly predict the ‘very acceptable’ tier in Table 3; IS’s Block interpretation also accounts for *yes* and *no* as acceptable answers; however, if the framework takes into account the fact that bare particle responses are considered less felicitous (Farkas and Roelofsen 2015), then the lower (yet still acceptable) ratings for these licit responses are also accurately predicted. Neither theory encompasses any sort of rule or mechanism by which people might interpret ambiguous lexical strings, so the various degrees of acceptability can be explained by a combination of preference for the default

interpretation, with lower ratings given to exception interpretations (all other interpretations) and to the bare particles.

In addition, neither theory offered predictions concerning answerhood conditions somehow altered by the location of the disjunctive in the question. However, the location of the disjunct swayed answer acceptability for both the text and perception experiments. Table 7.4 combines the acceptability ratings for the text and perception experiments.

Answer Condition	Disjunct Location	Text Experiment	Perception Experiment	
			Y/N tune	Alt tune
Disjunct (Falling intonation)	Initial	6.60	6.38	6.49
	Middle	6.59	6.36	6.59
	Final	6.81	6.60	6.75
Disjunct (Rising intonation)	Initial		5.86	5.61
	Middle		5.67	5.68
	Final		5.83	5.78
Yes, Disjunct	Initial	6.50	6.41	5.54
	Middle	6.43	6.32	5.40
	Final	6.50	6.36	5.49
Cleft	Initial	5.82	5.51	6.09
	Middle	5.27	5.14	5.89
	Final	5.49	5.21	5.68
Yes	Initial	4.58		
	Middle	4.36		
	Final	3.85		
No	Initial	5.73		
	Middle	5.93		
	Final	5.45		
Yes, both	Initial	6.50		
	Middle	6.64		
	Final	6.56		

Table 7.4: Acceptability ratings for text and perception experiments. Collapsed across answer condition and disjunct location in question (significant for both experiments).

Although the two experiments had some different answer conditions, a comparison of the answer conditions across experiments reveals that participants, indeed, seemed to be treating the ambiguous questions in the text experiment similar to the questions with Y/N tune contours in the perception experiment. This is especially apparent when comparing the “Yes + disjunct”

answer condition: the Y/N tune ratings from the perception experiment are extremely similar to the text ratings, while the alternative tune ratings are significantly lower. The cleft answer comparison and (Falling) disjunct answer comparison also support this conclusion; the latter answer is licit for both disjunctive question types, and so the text ratings are even higher than the perception ratings (this can be explained by a participant's preference to say/read the sentence as they would, instead of listening to a separate speaker who could have idiosyncrasies that lead to slightly lower ratings). However, the cleft answer ratings for the text experiment are *lower* than the alternative tune ratings from the perception experiment, indicating that participants did not read the questions in the text experiment as an alternative question, by default (but may have considered it as an attempt to fit the question-answer pair together).

The disjunct location does not universally lower or higher ratings across answer types (or even across the two tunes in the perception experiment). The following Figures 7.1-3 compare the answer conditions of plain disjunct (falling disjunct for the perception experiment), the Yes + disjunct answer, and the cleft answer by the three distinct disjunct locations in the question, across the text and perception experiments. The perception experiment is further broken down by the tune of the question.

Figure 7.1 further supports the notion that the text experiment questions were biased to a Y/N interpretation; the ratings across disjunct location for the text experiment mimic the ratings across disjunct location for the perception experiment, Y/N tune. The perception experiment ratings are probably lower because participants might be biased to their own prosodic tunes (which they can use in the text experiment) vs. the tune provided in the perception experiment.

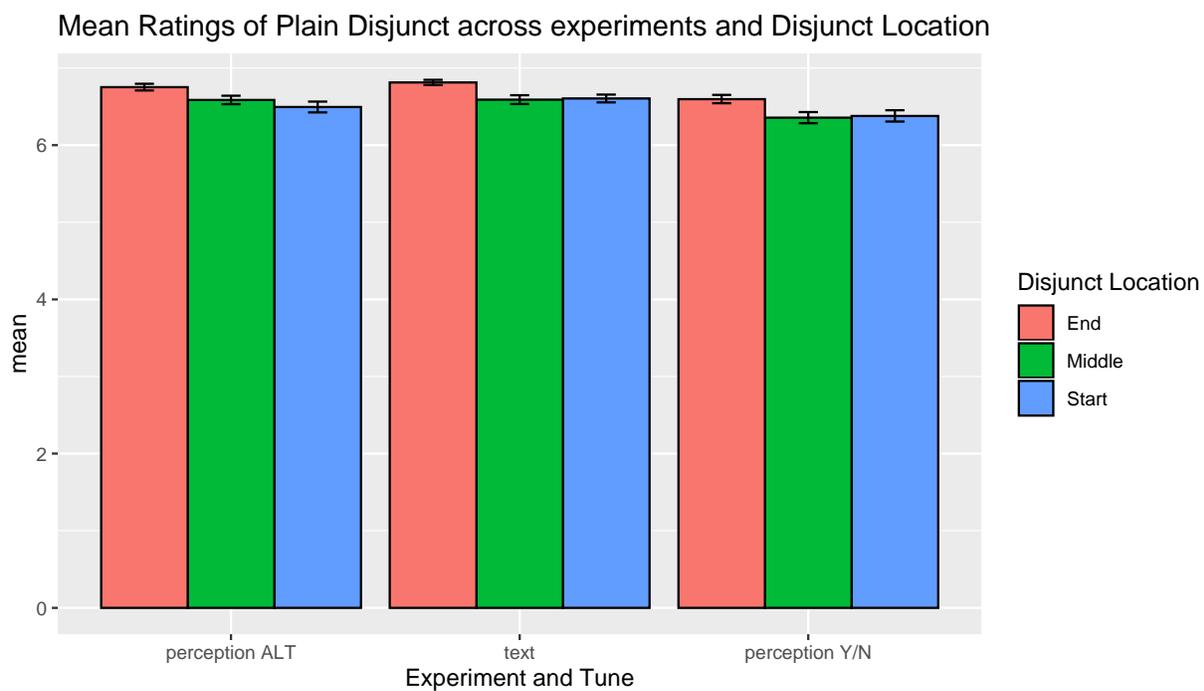


Figure 7.1: Plain disjunct answer condition across text and perception experiments. Perception experiment is further broken down by tune.

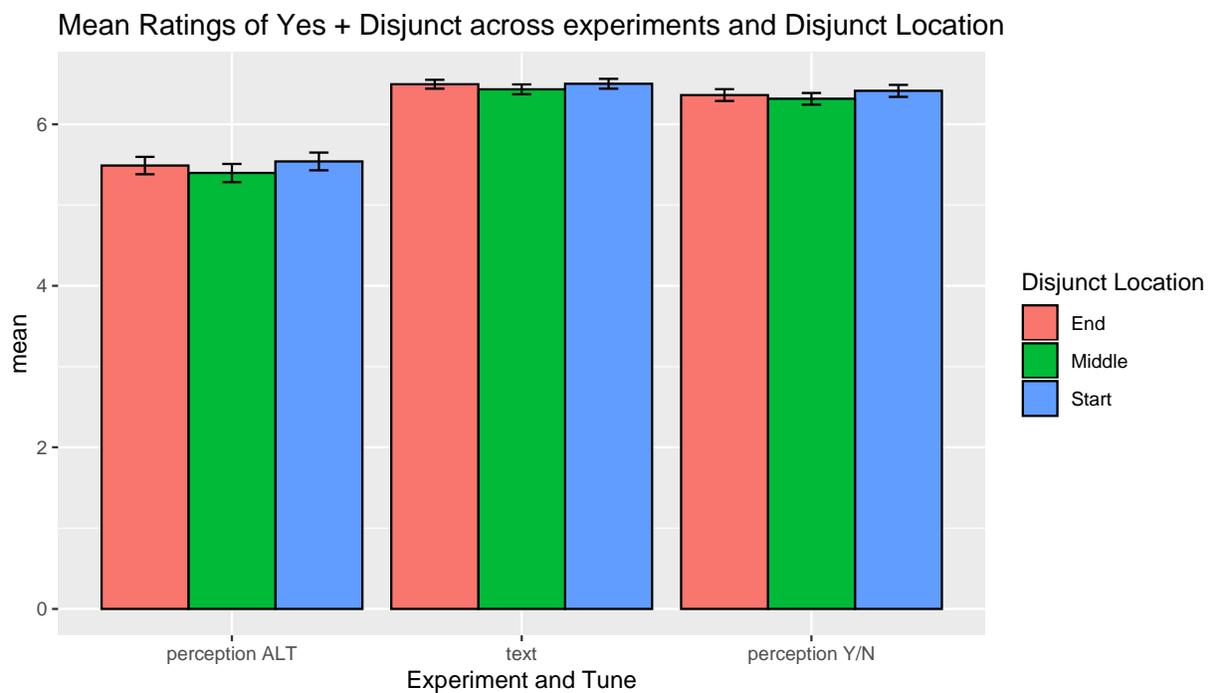


Figure 7.2: Yes + disjunct answer condition across text and perception experiments. Perception experiment is further broken down by tune.

Comparison of the Yes + disjunct answer condition (Figure 7.2) and cleft answer condition (Figure 7.3) offers similar results. The ratings for the Yes + disjunct answer condition (Figure 7.2) show that the text experiment and Y/N tunes in the perception experiment have the same contour and very similar ratings, again with the text experiment ratings slightly higher. The alternative tune for the perception experiment was rated significantly lower for this answer condition. When comparing the cleft disjunct answer (Figure 7.3), the alternative tune for the perception experiment was not only higher than both the Y/N tune and the text experiment results, but also had a different contour shape; the answer was rated higher as the disjunct moved from the end to the beginning of the question. Meanwhile, the text experiment and Y/N tune for the perception experiment had lowest ratings for when the disjunct was in the middle of the question.

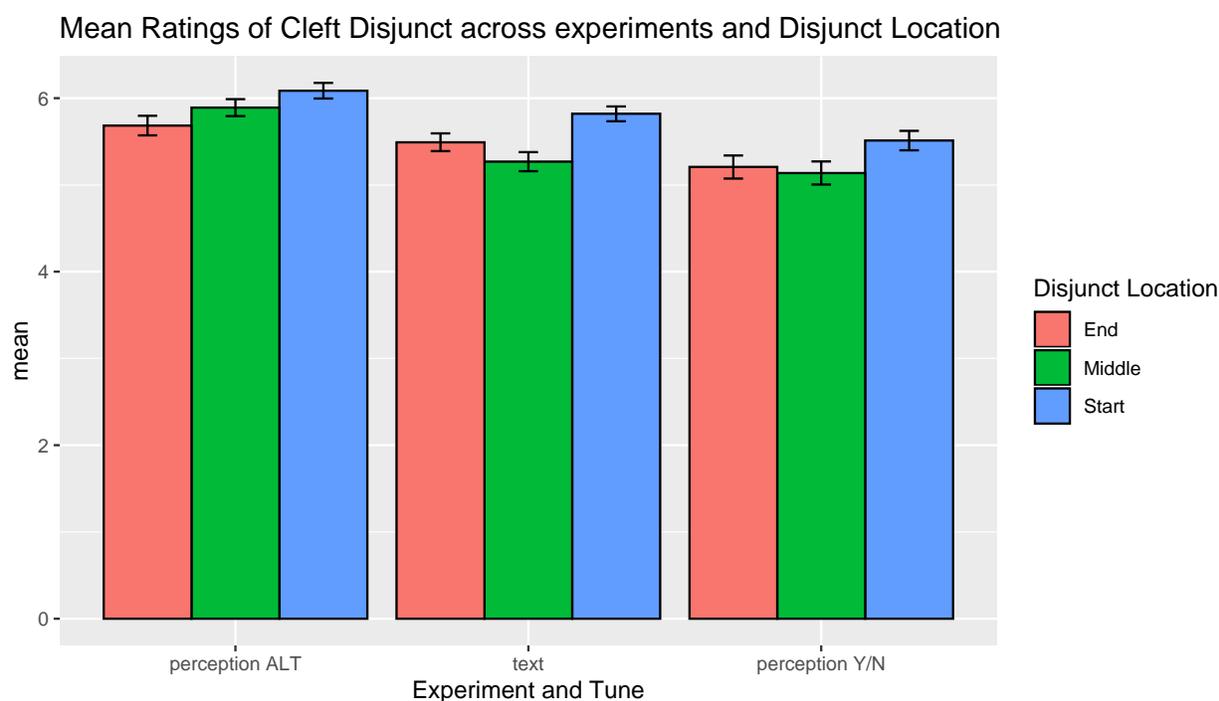


Figure 7.3: Cleft disjunct answer condition across text and perception experiments. Perception experiment is further broken down by tune.

While it might be argued that both tunes in the perception experiment seem to mimic the text experiment ratings with the plain disjunct answer condition (Figure 7.1), it becomes clear that there is a bias to the Y/N interpretation when looking at the cleft answer (Figure 7.3). If participants had both interpretations available and simply choose the one which fit the context better, then we would expect the results in Figure 7.2 (text experiment mimics Y/N tune ratings) but also expect the text experiment to mimic the Alt tune ratings for the cleft disjunct. As Figure 7.3 reveals, this is not the case. Despite the fact that participants have been shown to attempt to accommodate a question and answer pair (Heidenreich 2014b), they consistently rated the question-cleft answer pair in the text experiment lower than the Alt. tune counterpart; the results are more consistent with the Y/N tune ratings, indicating a Y/N question interpretation.

The perception experiment also revealed that participants accept *either* in an alternative question tune; there was no significant difference between alternative tune question ratings with *either* vs. without *either* (combined across all other factors). Even the production experiment garnered some recordings of alternative question tunes with *either*, when the participant was not told specifically what (not) to say, suggesting that both tunes ought to allow *either*. While *either* does not seem to change the interpretation of the question, it does seem to increase the subset of acceptable answers. Recall from the perception experiment that adding *either* into an alternative tune significantly increased acceptability of the “Yes + disjunct” answer. The presence of *either* in an alternative tune also slightly increased the acceptability of the rising disjunct answer; adding *either* into a Y/N tune slightly increased the acceptability of the cleft answer. These latter two trends were not significant in the statistical analysis, however. While this simplifies the models of the questions, it does require the presence/absence of *either* to interact with the set of

possible answers. Therefore, the set of criteria that must be met for the semantic theories to account for the data from the experiments has been updated to include the following:

1. Y/N disjunctive questions with focus on both disjuncts (having a phrasal break) must *yes* and *no* answers.
2. Alternative disjunctive questions can occur without a phrasal break, or without focus on both disjuncts
3. Alternative disjunctive questions must have the first disjunct ‘end high’
4. *Either* can occur in both question tunes without altering the meaning
5. The tune of the answer (falling vs. rising disjunct) needs to be taken into account
6. The presence of *either* in the alternative question tune must increase acceptability of answers typically only licit for Y/N question tune (the reverse, if predicted as well, would not go against empirical data)
7. Answers to disjunctive questions are on more of a continuous scale vs. categorical one
8. Answers to disjunctive questions are affected by where in the question the disjunct is located

Finally, the artificial language experiment corroborated the text experiment results in suggesting that Y/N questions are the ‘default’ disjunctive question, with alternative questions being a specific subtype. The ranking of language condition performance when inserting the disjunctive connective matched the predictions outlined by a superset/subset relationship. When participants had to give the correct answer, given a question, they were more accurate with the Alt. questions than Y/N questions. This is consistent with morphological results from Heidenreich (2014a) which showed that, holding type frequency constant, participants can more

easily learn the ‘core’ case than the ‘periphery’ case of a default vs. exception relationship. Furthermore, the condition in which the alternative question had a connective while the Y/N question did not was the worst performing of all language conditions. The results support the typological data that such a language does not exist, possibly because it is harder to learn. Indeed, there was one participant who was trained on this language condition and actually learned the opposite distinction! The data lend further credence to the idea that alternative questions are a specific type of Y/N question (Haspelmath 2007). Both theories ought to acknowledge this relationship via their interpretational methods.

Now that the results across experiments have been amassed, we can return to the first criterion. Participants put focus on both disjuncts (and a phrasal break in between disjuncts) in the production experiment for Y/N questions about a fourth of the time, and Jordi always responded with either “Yes” or “No”. Both theories would predict a lack of focus on both disjuncts and no phrasal break to allow these answers. However, it is possible the participants were disregarding Jordi’s (previous) responses to these questions. The text experiment revealed that participants rate the bare particle answers *yes* and *no* as less acceptable, and, based on the perception experiment results, it seems that participants were interpreting the text questions as Y/N questions. While the text experiment cannot say whether participants were interpreting focus on both disjuncts, the text and perception ratings were similar across shared answers, and the perception experiment revealed that these ratings were based on Y/N questions that did not have focus on both disjuncts or phrase breaks between them (only one Y/N question tune in the experiment had a phrasal break). Therefore, since participants typically ranked *Yes* and *No* as not acceptable for Y/N questions without a phrasal break, it can logically be deduced that participants were not taking Jordi’s responses of “Yes” and “no” into account when producing

the Y/N questions in the production experiment, and therefore criteria #1 may be eliminated.

Further research may offer insight into whether or not theories ought to reexamine the first criteria.

Therefore, the updated list of criteria is listed below:

1. Alternative disjunctive questions can occur without a phrasal break and focus on both disjuncts
2. Alternative disjunctive questions must have the first disjunct ‘end high’
3. *Either* can occur in both question tunes without altering the meaning
4. The tune of the answer (falling vs. rising disjunct) needs to be taken into account
5. The presence of *either* in the alternative question tune must increase acceptability of answers typically only licit for Y/N question tune (the reverse, if predicted as well, would not go against experimental data)
6. Answers to disjunctive questions are on more of a continuous scale vs. categorical one
7. Answers to disjunctive questions are affected by where in the question the disjunct is located

These criteria can further be broken down into those which affect the mechanisms of modeling the question only, and those which affect the licit answerhood conditions (Table 7.5, next page). As both theories use the licit answerhood conditions to formulate the models for each type of disjunction, both columns must be accounted for in order to accurately depict the productions and perceptions of disjunctive question-answer pairs. Each criterion will be examined in turn; both frameworks will be given the opportunity to account for the results.

	Criteria affecting question only	Criteria affecting question-answer relationship
Criteria from experiments that theories must account for	<ol style="list-style-type: none"> 1. Alternative questions can occur without a phrasal break 2. Alternative questions must have first disjunct end ‘high’ 3. <i>Either</i> can occur in both disjunctive questions without altering meaning 	<ol style="list-style-type: none"> 1. The tune of the answer must be taken into account 2. Presence of <i>either</i> in question increases acceptability of non-licit answers 3. Answer acceptability on a continuous scale 4. Answer acceptability affected by where in the question disjunction occurs

Table 7.5: Results from experiments that are currently not accounted for by formal theory

7.3 Alternative Questions: phrasal breaks and first disjunct ending ‘high’

The first two criteria involve the intonation patterns of alternative questions.

Canonically, it seems that alternative questions have a phrasal break after the first disjunct (H-), which aids in the disambiguation process (the final fall is also significant). However, the production experiment revealed that this phrasal break is not obligatory (about 23.5% of correct alternative question productions did not contain a phrasal break); rather, the first disjunct needs to end with a high pitch, but that could be expressed with a phrasal break (H-) or a pitch accent (L*+H)⁶³. The perception experiment stimuli also revealed that an alternative question need not have this phrasal break but did need the first disjunct to end ‘high’. Thus, both theories must be able to account for alternative questions potentially not having a phrasal break, but all alternative questions having the first disjunct end in a high pitch range.

⁶³ 99% of correct alternative tunes had either the high phrasal accent after the first disjunct, or a L*+H pitch accent on that disjunct.

7.3.1 Inquisitive Semantics and Alternative Questions

Inquisitive Semantics uses phrasing, final pitch contour, and word order to delineate the various subtypes of disjunctive questions. Specifically, IS stipulates that alternative questions are closed intonation patterns with a necessary phrase break in between the disjuncts. While this certainly seems to be the preferred or default intonational contour for alternative questions, it is not the only one. Thus, IS must be modified to account for alternative questions without phrasal breaks.

A closed disjunctive interrogative without a phrase break was touched on as a ‘special case’ (Roelofsen 2013b) in Inquisitive Semantics; it was determined that one ought to treat the disjuncts as separate list items, even though no prosodic break was perceived (Pruitt & Roelofsen 2013). This strategy, however, does not account for the second criteria; namely, that the first disjunct does need to end high. There seems to be a solution for this dilemma. Already, IS models a closed interrogative with an upward arrow after the first disjunct and a downward arrow after the second disjunct. If this upward arrow after the first disjunct can be instantiated by a phrase break (H-) or an accent (L*+H), then the model would accurately portray alternative questions. It has been shown that when two accents are close together (as they would be on two discourse-new nouns in a disjunctive question), it can be difficult to differentiate an intermediate phrase break (H-) from an accent (L*+H). Therefore, this additional mechanism would simply allow for this intonational ambiguity. This would alter the interpretation procedure for identifying disjunctive questions, as indicated by the bold below:

1. Determine basic list items
 - a. Detect the prosodic phrase boundaries
2. Determine whether the list is open (or block) or closed

- a. If the list is open or block, *no* is licensed, and so perform list completion
- b. If the list is closed, apply exclusive strengthening, **unless there is only one list item, then do the following:**

- i. **Look to see if the first disjunct ends ‘high’. If so, return to step 1 and mark the question as having two lists.**

3. Determine whether the list is declarative or interrogative

Thus, IS would account for the (84) as seen by the progress below (Figure 7.4):

(84) Is he going to Spain or Italy?

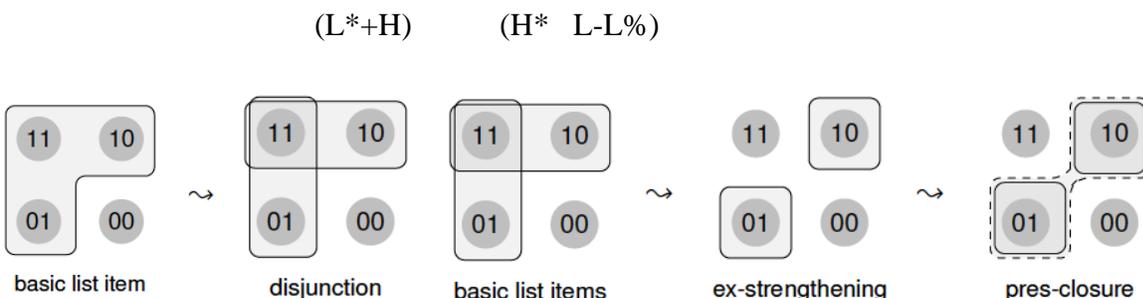


Figure 7.4: IS account of an alternative question without a phrase break; after determining only one list item, IS would see the final fall as a closed list, causing the disjunction to be reevaluated as two basic list items, and then restart with said list items.

The highlighting (exclusive strengthening) and suggesting (closure) mechanisms in IS already allow for an alternative question interpretation once the two disjuncts are viewed as separate ‘list’ items. These two mechanisms already use ‘focus features and prosody’ (Roelofsen and van Gool, 2010) to capture what a question may highlight or suggest; therefore it is reasonable to conclude that IS would be able to use the accent on the first disjunct in determining the correct delineation of a question like (84). The updated criteria list for IS now looks as follows:

Criteria	Accounted For	Mechanism Used
Alternative Questions can occur without a phrasal break; first disjunct must end high	YES	Allow high accent to delineate end of basic list item in closed lists
'Either' can occur in both question tunes without altering meaning	c.f. section 7.4	
The tune of the answer needs to be taken into account	c.f. section 7.5	
The presence of 'either' in question increases acceptability of illicit answers	c.f. section 7.6	
Answers to disjunctive questions on continuous scale	c.f. section 7.7	
Answers are affected by where in the question the disjunct is located	c.f. section 7.8	

Table 7.6: Criteria list for Inquisitive Semantics. Includes whether the theory can account for the experimental result and what mechanisms it uses.

7.3.2. Commitment Space Discourse and Alternative Questions

Commitment Space Discourse analyzes questions by restricting the possible continuations in the commitment space. Specifically, CSD must analyze alternative questions differently than polar questions, as the licit continuations to these two types of disjunctive questions differ. CSD claims that an alternative question, such as (85), is a disjunct of two monopolar questions (Figure 7.5, next page).

(85) Did I win the race, or not?

However, this alternative question example contained two propositions that were inherently in negation to one another (“I won the race” vs “I did not win the race”). When the disjuncts are not in negation, such as in (86), the framework analyzes the question as in Figure 7.6 (next page).

(86) Did Ed meet Ann[↑], Beth[↑], or Carla[↓]?

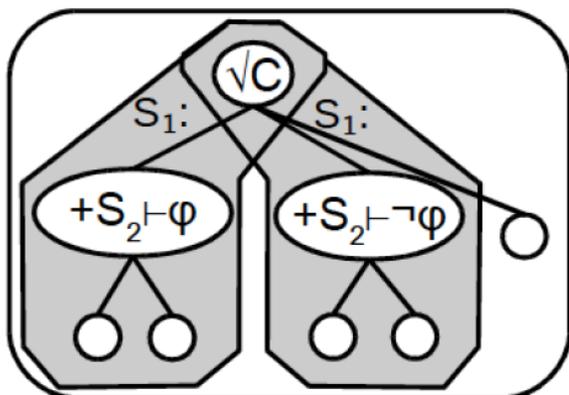


Figure 7.5: Alternative question: Disjunction of two monopolar questions (a proposition and its negation)

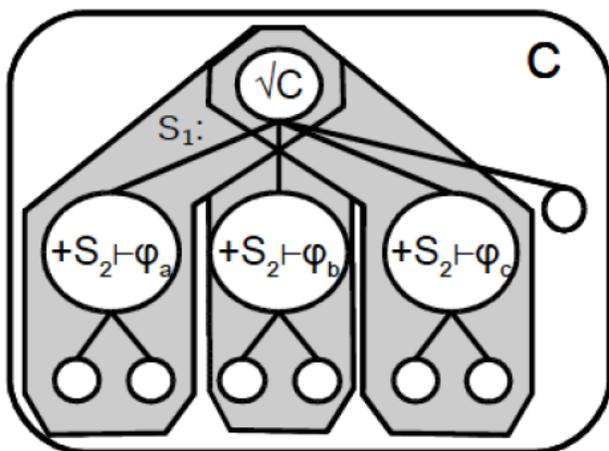


Figure 7.6: Alternative question: Disjunction of monopolar questions

In this particular example, it is not the case that “Ed met Ann”, “Ed met Beth” and “Ed met Carla” are in negation to each other; however, an alternative question would convey the expectation that if Ed met Ann, then he did not meet Beth and did not meet Carla. If we could assume (in Figure 7.6) that by S2 asserting proposition φ_a , they are removing φ_b and φ_c as

possible adjacent *or future* continuations, then the analysis would still hold. However, CSD analyzes a polar disjunctive question, such as (87), as in Figure 7.7.

(87) Did Ed meet Ann[↑] or did Ed meet Beth[↑]?

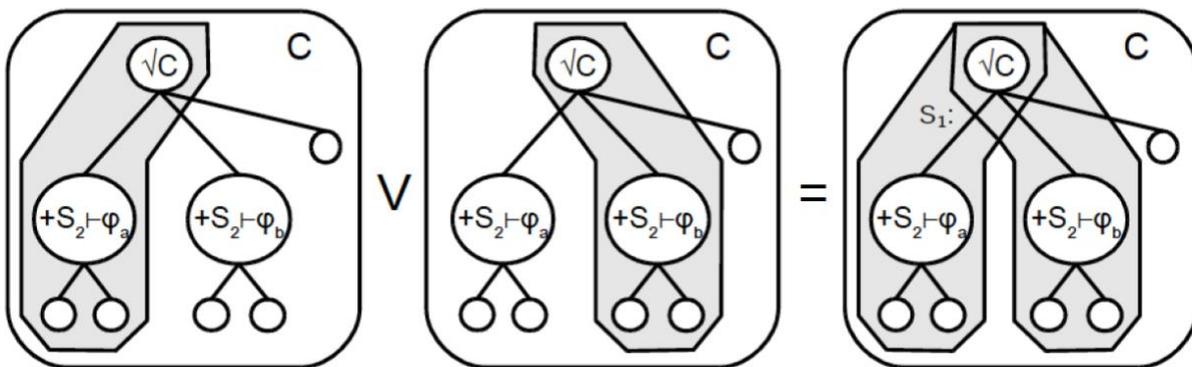


Figure 7.7: Polar question: Disjunctive question of two monopolar questions

A licit continuation to (87) can be “Ed met both of them”; thus, we may assume that in Figure 7.7, S2 ought to be able to assert both continuations. Thus, the alternative question interpretation needs a way to convey that by the second speaker asserting φ_a , they are simultaneously confirming the negation of φ_b . That is, when S1 utters an alternative question like (86), they are already presuming the ‘exactly one’ expectation; S2 needs to confirm this presumption via a single disjunct response, which then adds the negation of the other disjuncts into the common ground. Therefore, the licit continuations for S2 require *two* things: the assertion of one of the disjuncts and the confirmation of the negation of all others. Thus (88), which has an alternative question tune, ought to have the licit continuations in Figure 7.8.

(88) Did Ed meet Ann[↑] or did Ed meet Beth[↓]?

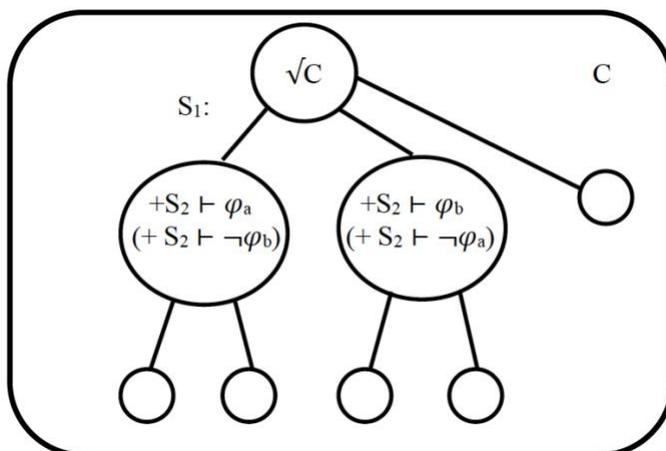


Figure 7.8: *Alternative Question with licit continuations*

The negated disjunct is included in parenthesis, because it is not a true assertion. Krifka (2017) states that the only admissible continuations are those in which the addressee makes an assertion that answers the question (p. 20). However, (89) reveals that simply asserting the negation of a disjunct does not constitute a licit response:

(89) S1: Did Ed meet Ann[↑] or did Ed meet Beth[↓]?

S2: #He didn't meet Ann.

[S1 would most likely ask for clarification, e.g. "So he met Beth?"]

Similarly, including the negated disjunct *as* an assertion in the answer also seems slightly less acceptable.

(90) S1: Did Ed meet Ann[↑] or did Ed meet Beth[↓]?

S2: ?He met Ann and not Beth.

Thus, it appears that alternative questions, like negated polarity questions, are complex speech acts, a combination of assertions and questions (Krifka 2017). He uses the ASSERT operator to combine the speaker's commitment to the proposition, as well the speaker calling on the addressee to be committed to that proposition. This is the operator that appears to be

functioning in the asking of the question, rather than the accepting of the question. Rather than the ASSERT function, it seems that the negation of the remaining disjuncts is done with the ACCEPT operation (Krifka 2017):

ACCEPT: S2 takes on the obligation imposed by S1

That is, when S2 asserts ‘He met Ann’ in (8), they are accepting the presumption by S1 that exactly one of the disjuncts may be chosen; by nature the ACCEPT operation can be done implicitly by lack of any reaction that would involve a rejection (Krifka 2017). Importantly, this presumption by S1 that ‘exactly one’ disjunct holds does not occur for Y/N questions, which explains the difference between (91), an alternative question, and (92), a polar question.

(91) S1: Did Ed meet Ann[↑] or did Ed meet Beth[↓]?

S2: Ed met Ann.

[C updated to include (Ed met Ann, ¬Ed met Beth)]

S1: #What about Beth?

(92) S1: Did Ed meet Ann[↑] or did Ed meet Beth[↑]?

S2: Ed met Ann.

[C updated to include (Ed met Ann)]

S1: What about Beth?

An alternative question response updates the common ground such that one disjunct is asserted and all other disjuncts have their negation added to the common ground; an answer to a Y/N question simply removes the other licit continuations. These continuations can re-enter with a separate question (i.e. ‘What about Beth’) after a Y/N question, but not after an alternative question, as the proposition’s negation has already been added to the commitment space.

Thus, CSD must account for the presumption by S1 that exactly one disjunct must be chosen for an alternative question. Although Krifka (2017) assumes that prosody plays a major role in the assertion of propositions (even going so far as to describe how the incredulity contour can affect a declarative question), the specific tones used in alternative questions to convey such a presumption have not been described. However, the assert operator is already linked to prosody via the nuclear stress H* (Pierrehumbert & Hirschberg 1990), which Krifka (2017) describes in (93):

(93) *There's a vegeTarian REStaurant around here.*

H* H* L- L%

Note that the end of this phrase is the type of contour that an alternative question seems to require; a high pitch followed by a fall, with the entire utterance ending L-L%. However, alternative questions can occur even when no H* is present; consider (94-95), taken from the production experiment, and (96-97), from the perception experiment.⁶⁴

(94) Do you want to see the gazelle or the zebra?

L* H- L* L-L%

(95) Would you like to see the porcupine or the dingo next?

L*+H L* L-L%

(96) Is Pamela going to knit a blanket or an outfit for the baby shower?

L* H- L* L-L%

(97) Were hamburgers or hotdogs served at the orientation picnic?

L*+H L* L-L%

⁶⁴ The perception experiment recordings did not demand specific contour tunes; rather, each stimulus was reviewed aurally and was deemed an alternative question if Shari Speer, Carl Pollard, and myself all agreed on its interpretation.

In (94) and (96), there is no H*, but a H- is present at the end of the first disjunct; similarly, in (95) and (97), there is only the L*+H at the end of the first disjunct. Thus, the ASSERT operator must take any sort of high pitch marker at the end of the first disjunct in conjunction with the final low pitch of the utterance. The relationship between these two points, or their pitch slope, is the same pitch slope given as evidence for the ASSERT operator in (93); and in fact, Krifka (2017) notes the difference between (93) and a response like (98), which does not have the same contour and does not assert that the proposition should become part of the common ground.⁶⁵

(98) *Of course we can find a decent place to eat. Just remember!*

There's a vegeTarian REStaurant around here.

L* L* L- L%

This isn't the first time that Commitment Space Discourse operators have been linked to intonation contours; Prieto and Borràs-Comes (2018) tested intonation patterns in Catalan, and found that question intonation, in particular, can encode different levels of ASSERT and REJECT. It is possible that the ASSERT function can take multiple different specific intonational patterns; regardless, the ASSERT operator (at least, in English) seems to take the pitch slope as a necessary component for its use. Thus, we may conclude that CSD can now account for alternative questions requiring the first disjunct to end 'high', by incorporating the ASSERT and ACCEPT operations, already a part of Commitment Space Discourse, into the analysis of disjunctive questions.

Krifka's (2015, 2016) account of alternative questions seems to correlate the breakdown of a disjunctive question into two monopolar propositions with a phrasal break; similarly, a

⁶⁵ Krifka argues that one of the possible meanings of the prosody in 15 is that S1 thinks that the proposition is already part of the common ground, and is merely reminding S2

phrasal break in a Y/N question determines the types of licit continuations one can expect.

However, constituent questions, such as “Which woman did Ed meet?” are also analyzed as a disjunction of monopolar questions (Figure 7.9).

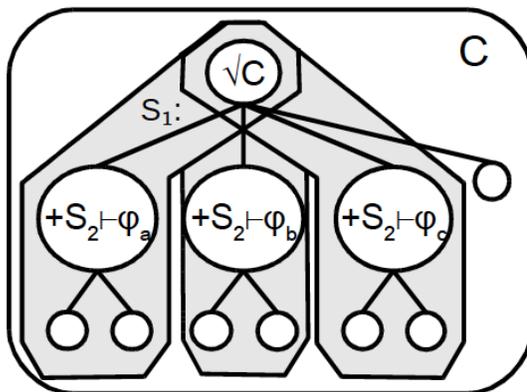


Figure 7.9: Constituent as disjunction of monopolar questions (taken from Krifka 2016).
Question was “Which woman did Ed meet?”

Therefore, phrasal breaks must not be necessary to impose a monopolar interpretation. Thus, how CSD handles alternative questions does not depend on the presence or absence of a phrasal break. With this in mind, as long as the ASSERT operator uses the intonation pattern of alternative questions, the first criteria in Table 7.7 (next page) is accounted for.

Criteria	Accounted For	Mechanism Used
Alternative Questions can occur without a phrasal break; first disjunct must end high	YES	ASSERT operator on alternative question contour
'Either' can occur in both question tunes without altering meaning	c.f. section 7.4	
The tune of the answer needs to be taken into account	c.f. section 7.5	
The presence of 'either' in question increases acceptability of illicit answers	c.f. section 7.6	
Answers to disjunctive questions on continuous scale	c.f. section 7.7	
Answers are affected by where in the question the disjunct is located	c.f. section 7.8	

Table 7.7: Criteria list for Commitment Space Discourse Includes whether the theory can account for the experimental result and what mechanisms it uses.

7.4 The presence of *either* in disjunctive questions

The next criterion alters the stance in English grammar in general, rather than altering the frameworks specifically. Haspelmath (2007) stated “The emphatic disjunctive markers *either...or* only express standard disjunction (*Do you want either tea or coffee?* cannot be an alternative question) (p. 26). Huddleston (1994) similarly stated “the polar question is differentiated from an alternative question by intonation [..]; the only respect in which the distinction is grammaticalized is that the *or* of an incidental (standard) disjunction can be paired correlatively with *either*, whereas that of an alternative disjunction cannot: *Is it either a boy or a girl?* Is unambiguously a polar question” (p. 420). While this may have been the case in the past, results from the production and perception experiments show that participants accept *either* in an alternative tune context, and the presence of *either* does not alter the meaning of the sentence to a polar interpretation. Rather, the intonational contour alone determines which of the

two interpretations the disjunction will have. The production experiment revealed that participants were comfortable producing *either* unprompted in a polar context *and* an alternative context, although no single participant used it in both contexts. The perception experiment found no significant difference between ratings for alternative tune questions without *either* vs. those with *either*. Gibson et. al. (2011) notes that many judgments, especially concerning the acceptability of a sentence / meaning pair, don't survive experimental evaluation. In fact, "there are many examples in published syntax and semantics articles where the reported judgments do not match those of typical native speakers of the language in question" (Gibson et. al. 2011). Although these experimental results are certainly surprising, the judgments of *either* in an alternative question were never tested empirically, and this is not the first instance of empirical results suggesting behavior different from prescriptive grammar or language experts.

Possibly, the grammatical change occurred because declarative sentences like (99) exist: an alternative tune with the presence of *either*.

(99) I will either walk to school[↑] or take the bus[↓].

It may be the case that participants may accept *either* in an alternative question because this intonational pattern can exist in the declarative counterpart. Regardless, neither theory offers specific accommodations for restricting *either* in an alternative question, and therefore both theories already allow this surprising result by not explicitly restricting it. The tables are updated on the next page accordingly.

Criteria	Accounted For	Mechanism Used
Alternative Questions can occur without a phrasal break; first disjunct must end high	YES	Allow high accent to delineate end of basic list item in closed list
'Either' can occur in both question tunes without altering meaning	YES	Already accounted for in framework
The tune of the answer needs to be taken into account	c.f. section 7.5	
The presence of 'either' in question increases acceptability of illicit answers	c.f. section 7.6	
Answers to disjunctive questions on continuous scale	c.f. section 7.7	
Answers are affected by where in the question the disjunct is located	c.f. section 7.8	

Table 7.6 (updated): Criteria list for Inquisitive Semantics. Includes whether the theory can account for the experimental result and what mechanisms it uses.

Criteria	Accounted For	Mechanism Used
Alternative Questions can occur without a phrasal break; first disjunct must end high	YES	ASSERT operator on alternative question contour
'Either' can occur in both question tunes without altering meaning	YES	Already accounted for in framework
The tune of the answer needs to be taken into account	c.f. section 7.5	
The presence of 'either' in question increases acceptability of illicit answers	c.f. section 7.6	
Answers to disjunctive questions on continuous scale	c.f. section 7.7	
Answers are affected by where in the question the disjunct is located	c.f. section 7.8	

Table 7.7 (updated): Criteria list for Commitment Space Discourse. Includes whether the theory can account for the experimental result and what mechanisms it uses.

7.5 The tune of the response

The experiments not only found a relationship between prosody and the question, but also between prosody and the response. In particular, the perception experiment tested two different versions of the ‘Disjunct’ response, one with falling intonation (default assertion) and the other with rising intonation (continuation/uncertainty contour).

Question Tune	Answer Condition			
	Disjunct↓	Disjunct↑	Yes + Disjunct	Cleft
Alternative	6.61	5.69	5.48	5.89
Y/N	6.44	5.79	6.36	5.29

Table 7.8: Ratings from Perception Experiment, across question tune and answer condition.

Table 7.8 reveals the need for the theories to take into account the tune of the answer. The theories need to explain why an answer like “Dairy” or “Soy” is very acceptable when accompanied by a falling tone, but not very acceptable when accompanied by a rising tone (answer condition 2).

7.5.1 Inquisitive Semantics and response tunes

Recall that the rising disjunct tone was represented as L-H%, or a continuation rise (Beckman and Hirschberg 1994). This continuation rise communicates a potential unfinished answer with its ‘forward reference’; that is, according to the established framework of IS, the two disjuncts have different response ‘worlds’ based on their intonation:

(100) A closed declarative with a single item (next page)

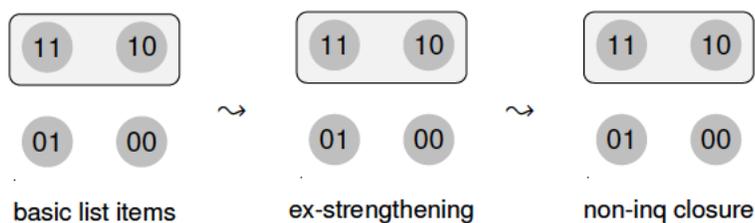


Figure 7.10: *He is going to Spain* ↓, analyzed in the IS framework (Roelofsen 2013b)

(101) An open declarative with a single item



Figure 7.11: *He is going to Spain* ↑, analyzed in the IS framework (Roelofsen 2013b)

While Figure 7.10 reveals that the falling tone disjunct is non-inquisitive, having only one list item, Figure 7.11 shows that a rising tone on a declarative (such as ‘She’s allergic to dairy’ ↑) remains inquisitive with a final rise, and thus such a response can be taken as answering a question with an incomplete answer. Specifically, the rising disjunct response would resolve differently, depending on the interpretation of the question:

Alternative Question Interpretation: When in response to an alternative question, a continuation rise would signal non-finality; however, the ‘exactly one’ mandate by an alternative question would contradict the idea that more needs to be said. Therefore, the answer, although a ‘Disjunct’, becomes illicit.

Polar Question Interpretation: When in response to a polar question, a continuation rise would, again, signal non-finality (forward reference). This could be seen as an incomplete answer to the question, as seen in the example below:

S1: Is Mary allergic to dairy↑ or soy↑?

S2: She's allergic to dairy... (continuation rise)

S1: What about soy?

That is, participants might feel as though the speaker was withholding known information with this type of intonation, thereby flaunting the Maxim of Quantity (Grice 1975). However, it might also imply that the speaker isn't sure about soy, and therefore leaves that proposition open, rather than the falling tone "She's allergic to dairy". The ambiguity left in interpreting the continuation rise could cause a lower rating, though it might not necessarily be deemed 'illicit'.

Both possibilities cause the rising tone disjunct answer to be rated lower, regardless of the disjunctive question interpretation chosen. It explains why the rising tone disjunct is viewed as illicit in an alternative question tune (although a disjunct answer is needed for such a question), and also explains why such an answer is still rated lower than the falling tone disjunct answer when paired with a Y/N question tune. The updated criteria for Inquisitive Semantics are listed in Table 7.6 (updated on the next page).

Criteria	Accounted For	Mechanism Used
Alternative Questions can occur without a phrasal break; first disjunct must end high	YES	Allow high accent to delineate end of basic list item in closed list
'Either' can occur in both question tunes without altering meaning	YES	Already accounted for in framework
The tune of the answer needs to be taken into account	YES	Rising intonation already interpreted as inquisitive
The presence of 'either' in question increases acceptability of illicit answers	c.f. section 7.6	
Answers to disjunctive questions on continuous scale	c.f. section 7.7	
Answers are affected by where in the question the disjunct is located	c.f. section 7.8	

Table 7.6 (updated): Criteria list for Inquisitive Semantics. Includes whether the theory can account for the experimental result and what mechanisms it uses.

7.5.2. Commitment Space Discourse and response tunes

It has already been shown how the ASSERT operator in commitment space discourse can be used in questions in order to convey the 'exactly one' expectation the speaker has when posing an alternative question. This operator addresses both disjuncts for an alternative question in each continuation. In a Y/N question like (102), answer (a) would also address both continuations, while answer (b) or (c) would address merely one, removing the other continuation from the commitment space but not adding anything about that proposition into the common ground.

(102) Did Ed meet Ann[↑] or did Ed meet Beth[↑]?

- a. Yes, Ed met both of them
- b. Ed met Ann
- c. Ed met Beth

Replying with (b) or (c) with a continuation rise, however, indicates non-finality. In this case, it is possible that the other continuation in the commitment space *remains* as a licit continuation, rather than being removed. It also indicates that S2 did not fully address the propositions laid out by S1 and thus explains the lower rating for a rising disjunct answer than for a falling disjunct answer. This type of analysis also offers the following predictions:

1) For alternative question tunes, the rising disjunct ought to be unacceptable. S2 is giving a licit continuation, but they are also *not* accepting S1's presumption that exactly one of the propositions holds when using continuation rise.

2) For Y/N question tunes, the rising disjunct ought to be acceptable, though not as helpful. That is, S2 is giving a licit continuation, and they are not rejecting any sort of presumption by S1 that only one disjunct can be chosen. However, S2 is leaving the question and the remaining continuations open, and therefore they might be violating the Maxim of Quantity (Grice 1975).

Both these predictions seem to suggest that answers would be accepted on a continuous scale rather than categorical one, and that is fully supported by the experimental results. The continuous scale of acceptable answers is addressed in section 7.7. For now, the updated criteria for CSD are listed in updated Table 7.7 (updated, next page)

Criteria	Accounted For	Mechanism Used
Alternative Questions can occur without a phrasal break; first disjunct must end high	YES	ASSERT operator on alternative question
'Either' can occur in both question tunes without altering meaning	YES	Already accounted for in framework
The tune of the answer needs to be taken into account	YES	Rising tune in response leaves continuations not added to common ground in commitment space
The presence of 'either' in question increases acceptability of illicit answers	c.f. section 7.6	
Answers to disjunctive questions on continuous scale	c.f. section 7.7	
Answers are affected by where in the question the disjunct is located	c.f. section 7.8	

Table 7.7 (updated): Criteria list for Commitment Space Discourse. Includes whether the theory can account for the experimental result and what mechanisms it uses.

7.6 Presence of *either* in questions altering answerhood acceptability

The experimental results showed that *either* could occur in both types of disjunctive questions without affecting question interpretation; however, that isn't to say that *either* was without effect: the acceptability of *answers* changed when inserting *either* into a disjunctive question. Specifically, inserting *either* into the question opened up licit answers to both types of interpretations, although the interpretation that fit the question tune was still preferred. The perception experiment results showed the most dramatic increase in acceptability for an alternative tune with *either* when paired with a "Yes + Disjunct" answer (c.f. Figure 5.3). It seems that *either* is somehow interacting with the potential continuations or answerhood conditions of the question.

It has been argued that *either*, like *any*, is systematically ambiguous between a negative polarity item (an existential) and a “free-choice” item (a universal) (Higginbotham 1991).⁶⁶ Thus *either* introduces a level of ambiguity that isn’t there when *either* is omitted. It is possible that this additional level of ambiguity might allow responses for the other interpretation to be *more* available, than when *either* is absent. Table 7.9 reprints the results from the perception experiment.

Question Tune	Either	Answer Conditions			
		Disjunct↓	Disjunct↑	Yes + Disjunct	Cleft
Alternative	Absent	6.61	5.66	5.29	5.96
	Present	6.61	5.72	5.66	5.82
Y/N	Absent	6.43	5.74	6.39	5.21
	Present	6.46	5.83	6.34	5.36

Table 7.9: Ratings across answer condition, tune, and *either* in perception experiment.

The falling disjunct answer (licit for both conditions) does not change with the addition of *either*; as the falling disjunct answer is rated slightly higher for Alt. questions than for Y/N questions, this provides evidence that *either* is not altering the interpretation of the question. However, the other answer conditions do have different ratings when adding *either* into the question. In particular, adding *either* in an alternative tune increases acceptability of the Yes + disjunct answer considerably, while slightly increasing the rising disjunct answer and slightly

⁶⁶ Higginbotham makes the comparison between *any* and *either* by showing that *either* without *or* has the same distribution of *any*, e.g.:

- (1) Any/Either of them will work
- (2) John plays any/either game.
- (3) *John played any/either game
- (4) I don’t know any/either of them
- (5) *I know any/either of them

Higginbotham states “As a negative polarity item, [either] is existential, as in (4); as a free-choice item, it is universal, as in (1) and (2).”

decreasing the cleft answer. Adding *either* to a Y/N question also slightly increased the acceptability of the rising disjunct, while slightly decreasing the acceptability of the Yes + Disjunct and cleft disjunct answers. Adding *either* did not change the relative ordering of acceptable answer (i.e. the answer conditions were ordered most to least acceptable in the same manner regardless of whether *either* was present). The statistical analysis from the perception experiment found that only the interaction between *either* and the Yes + disjunct answer condition was significant; therefore, it seems that both theories must provide a rationale for why adding *either* to an alternative question tune allows the canonical Y/N answer, Yes + disjunct, to be more acceptable.

7.6.1 Inquisitive Semantics and the interaction with *either*

Inquisitive Semantics may be able to capture this phenomenon if *either* can be considered a lexical item that, like focus, can affect what a sentence *highlights*. In general, when *either* is included in a lexically ambiguous utterance (like disjunctive questions), it allows responses for both interpretations. The intonational contour of the question, meanwhile, disambiguates the question and also prefers those answers that align with that interpretation. For example, (103a) and (103b) are an alternative question without and with *either*, respectively. Each one will be examined to determine how IS can account for the different answerhood conditions.

(103a) Is Mary allergic to dairy[↑] or soy[↓]?

(103b) Is Mary allergic to either dairy[↑] or soy[↓]?

Example (103a) has an alternative question tune. Therefore, according to IS, the following steps are taken to determine licit answers:

(103a) Is Mary allergic to dairy[↑] or soy[↓]?

- c. $[[\alpha]]_P$ proposes three possibilities: Mary is allergic to dairy, Mary is allergic to soy, or neither
- d. $[[\alpha]]_H$ highlights two possibilities: Mary is allergic to dairy, Mary is allergic to soy
- e. $[[\alpha]]_S$ removes the overlap between the highlighting suggestions, resulting in the two distinct possibilities.
- f. Licit answers: Dairy; Soy

The results from the perception experiment support this conclusion that the two disjuncts are significantly better than any other answers (6.61 acceptability rating for disjunct answer, as compared to next highest answer, cleft disjunct, at 5.96). One might predict that answers such as “I think it’s soy”, which provide additional information (uncertainty) in answering the question, would be equally acceptable; this is an area for further research. Importantly, the Yes + Disjunct answer was rated significantly lower (though still in the range of acceptable). We might theorize that the answer remains slightly open due to the ambiguous nature of the question itself; although intonation highlights and suggests answers, it may not fully negate the full set of possible answers across both interpretations.

However, when adding *either* to the question, the Yes + disjunct answer becomes more acceptable (significantly). It is possible that *either* forces the possible answer sets through both types of interpretation, making both answer sets available; then the intonation of the question still allows the answers for that interpretation to be preferred answer(s). So while the proposition remains the same $[[\alpha]]_P$ for the utterance, there are two separate H-sets and S-sets, one for the alt

interpretation and the other for the polar interpretation. Thus, (103b) may be analyzed as follows:

(103b) Is Mary allergic to either dairy[↑] or soy[↓]?

- a. $[[\alpha]]_P$ proposes three possibilities: Mary is allergic to dairy, Mary is allergic to soy, or neither
- b. $[[\alpha, ALT]]_H$ highlights two possibilities: Mary is allergic to dairy, Mary is allergic to soy. $[[\alpha, ALT]]_s$ removes the overlap between the highlighting suggestions, resulting in the two distinct possibilities: Dairy; Soy
- c. $[[\alpha, Y/N]]_H$ highlights three possibilities: Mary is allergic to dairy, Mary is allergic to soy, Mary is allergic to neither (list completion) $[[\alpha, Y/N]]_s$ closes the set, states that one of the three possibilities must hold (Mary is allergic to dairy, Mary is allergic to soy, Mary is allergic to neither).

Importantly, because the Y/N interpretation allows *No*, *neither* as an answer, the response *Yes*, *dairy* or *Yes*, *soy* may now be felicitous (one would argue that since *No* is available, the alternative *Yes* before disjuncts is also now available). Therefore, (103b) offers the following answer predictions:

(103b) Is Mary allergic to either dairy[↑] or soy[↓]?

- a. Alternative interpretation (Preferred due to intonation): Dairy, Soy
- b. Y/N interpretation (allowed due to *either*): Dairy, Soy, No, Yes + disjunct

This analysis successfully explains why the “Yes + Disjunct” answer was rated higher when paired with the alternative question with *either*. It would also explain why the cleft disjunct

response, typically illicit with a Y/N question tune, would be rated slightly higher when the question contained *either*. Table 7.6 is updated below.

Criteria	Accounted For	Mechanism Used
Alternative Questions can occur without a phrasal break; first disjunct must end high	YES	Allow high accent to delineate end of basic list item in closed list
'Either' can occur in both question tunes without altering meaning	YES	Already accounted for in framework
The tune of the answer needs to be taken into account	YES	Rising intonation already interpreted as inquisitive
The presence of <i>either</i> in question increases acceptability of illicit answers	YES	<i>either</i> opens both interpretations; intonation still elevates preferred interpretation
Answers to disjunctive questions on continuous scale	c.f. section 7.7	
Answers are affected by where in the question the disjunct is located	c.f. section 7.8	

Table 7.6 (updated): Criteria list for Inquisitive Semantics. Includes whether the theory can account for the experimental result and what mechanisms it uses.

7.6.2 Commitment Space Discourse and the *either* interaction

It seems that the presence of *either* in a disjunctive question allows responses from both interpretations to varying degrees, rather than the responses indicated by the question tune. Commitment Space Discourse already analyzes alternative questions and Y/N questions in mostly the same manner (monopolar questions); only the alternative question (now) uses the ASSERT operator to indicate the 'exactly one' stipulation herein. The juxtaposition of this ASSERT operator with an ambiguous lexical item (*either*) creates a clash between a presumption (knowing/asserting) and ambiguity; it is this dichotomy that lends itself to the higher ratings for the Yes + Disjunct response (which are still rated the lowest of all four answer conditions in the perception experiment for the alternative question). That is, typically the ASSERT operator is

the tool by which the alternative question subset is derived from the Y/N question superset.

When this tool's effectiveness is compromised (such as by adding uncertainty), the line marking the delineation between alternative question responses and Y/N question responses appears to be blurred, which increases acceptability of responses reserved solely for the other interpretation.

Although the single significant interaction between *either* and an answer condition was with “Yes + disjunct”, this analysis also predicts the higher ratings for the cleft answer with a Y/N tune. Additionally, the increase of ambiguity in the question in terms of interpretation would make an uncertain response more acceptable; and indeed, the rising disjunct response is more acceptable with *either* than without *either* for both question tunes.

Thus, the ambiguity introduced by the presence of *either* hinders the effectiveness of the ASSERT operator, which is the only difference in analysis in CSD when interpreting alternative questions vs. Y/N questions.

Criteria	Accounted For	Mechanism Used
Alternative Questions can occur without a phrasal break; first disjunct must end high	YES	ASSERT operator on alternative question
‘Either’ can occur in both question tunes without altering meaning	YES	Already accounted for in framework
The tune of the answer needs to be taken into account	YES	Rising tune in response leaves continuations not added to common ground in commitment space
The presence of ‘either’ in question increases acceptability of illicit answers	YES	‘Either’ increases ambiguity; alters ASSERT effectiveness
Answers to disjunctive questions on continuous scale	c.f. section 7.7	
Answers are affected by where in the question the disjunct is located	c.f. section 7.8	

Table 7.7 (updated): Criteria list for Commitment Space Discourse. Includes whether the theory can account for the experimental result and what mechanisms it uses.

7.7 Continuous Scale of Acceptability for Answers

Table 7.9 in section 7.6 opens discussion for another criteria: it seems clear that answers to disjunctive questions do not separate into distinct categories, but rather are rated on a continuous scale of acceptability. The perception experiment indicated, via fillers, that participants were able and willing to use the entire Likert scale in order to rate the question-answer pairs; however, many participants rated all stimuli question-answer pairs in the ‘acceptable’ range. There was significant difference between the answer conditions as well; participants rated canonically licit answers as ‘very acceptable’ and then rated canonically ‘illicit’ answers as ‘moderately acceptable’ or ‘slightly acceptable’. Both theories currently operate on the assumption that answers are separated into two distinct categories, ‘licit’ and ‘illicit’.

Heidenreich (2014b) revealed participants’ desire to fit questions into a ‘consistent’ or understandable context when dealing with disjunctive questions; even when a preceding context (fitting an alternative interpretation) had a question (with a Y/N tune), participants claimed the two were consistent over 40% of the time. Reversing the context and tune still had participants claiming consistency over 30% of the time, when by all accounts the context and following question tune ought to be dismissed categorically. In these experiments, therefore, even when the question and answer ought not to fit together, participants may have *wanted* the two to fit together, and therefore pursued interpretations in order to find one that would allow such an answer for such a question. While these results support the notion that the lower rated answers may actually be considered unacceptable, both theories must still account for the range across answer conditions; why were some licit answer more acceptable than others, and why were some illicit answers less acceptable than others?

7.7.1 Inquisitive Semantics and continuous scale

Roelofsen and van Gool (2010) hinted at the idea of a continuous value feature over binary features; they mention that the “rising-and-falling pitch contour that was taken to signal closure may be pronounced more or less dramatically, and this seems to correlate with the strength of the corresponding ‘exactly one’ suggestion” (p. 393). If phonological content informs semantic meaning of the question on a continuous scale, then it follows that answers themselves would lie on a continuous rather than categorical scale.

But what sort of phonological content could contribute to the scale of acceptability? Roelofsen and van Gool (2010) had suggested the slope of the alternative pitch contour would affect alternative interpretation. Specifically, they suggest that the closure function, taken from the rising-and-falling pitch contour, correlates with the strength of the ‘exactly one’ suggestion. The sharper the contour, the stronger the suggestion. The perception experiment offers data into this suggestion. The difference between the pitch at the end of the first disjunct and the end of the second disjunct was taken for alternative questions, in order to measure the ‘sharpness’ of the slope. On the other end, the ‘rise’ in Y/N questions was measured, taking the difference from the pitch on the second disjunct to the end of the phrase. For alternative questions, this resulted in a positive number, indicating a ‘fall’; for Y/N questions, this difference was a negative number, indicating a ‘rise’.

Both question types allow the falling disjunct as a licit answer; therefore the strength of this ‘exactly one’ suggestion ought to be irrelevant across that answer condition. The rising disjunct was shown to be slightly acceptable for Y/N questions, but not acceptable for alternative questions. It is unlikely that the Alt. question slope would have an effect on the acceptability of Alt. questions with this response; it is possible that the Y/N question slope could affect

acceptability of the rising disjunct, though it is unclear as to which direction the effect would take. Y/N questions scored significantly higher with the “Yes + disjunct” answer; therefore we would predict, if Roelofsen and van Gool’s theory were to be upheld, that the more positive the slope (the sharper the fall), the worse the rating for this answer type; the more negative slopes (the sharper the rise), however, ought to have better ratings. The predictions are the opposite for answer condition 4 (the cleft condition); the more positive the slope (fall), the better the ratings ought to be for this answer condition, as the cleft answer ought to be licit for only Alt. question interpretations. Table 7.10 reveals the average slope of questions, broken down by question type, answer condition, and rating given.

Rating	Alternative Question				Y/N Question			
	Falling Disjunct	Rising Disjunct	Yes + Disjunct	Cleft Disjunct	Falling Disjunct	Rising Disjunct	Yes + Disjunct	Cleft Disjunct
1	103.61	107.97	90.72	102.01	---	-84.14	---	-86.69
2	---	104.85	102.70	97.66	-83.45	-81.47	-69.64	-95.61
3	73.42	96.27	102.16	96.95	-100.14	-73.01	-96.97	-82.22
4	114.30	105.27	105.20	100.39	-87.47	-89.86	-86.63	-84.99
5	108.18	103.95	101.40	100.51	-82.87	-85.86	-82.10	-89.96
6	101.98	102.59	103.57	102.07	-87.02	-87.72	-87.05	-85.07
7	100.75	98.88	98.91	102.43	-86.42	-87.15	-86.79	-86.81

Table 7.10: Average ‘closure’ slope by question type, answer condition, and given rating.

The results show that a less dramatic ‘final fall’ (or closure slope) did not translate automatically to a weaker alternative interpretation for alternative questions. Table 7.10 reveals that when rating the alternative question and Yes + disjunct combination as very unacceptable (a

1 rating), the final fall was actually comparatively small (lower positive number), when the closure slope predictions indicated that it should be the largest. Similarly, when rating the alternative question and cleft disjunct combination as very unacceptable, the closure slope was comparatively larger (larger final fall), which again was the opposite of what was predicted. The Y/N question slopes are all quite comparable; though to be fair, Roelofsen and van Gool (2010) did not suggest using the final rise as a mechanism to determine the strength of the Y/N interpretation. However, they did imply that the alternative interpretation may be influenced by the intonational contour of Alt. questions. Figure 7.12 shows the average ‘closure slope’ across each rating (least acceptable on the left, most acceptable on the right) for each answer condition.

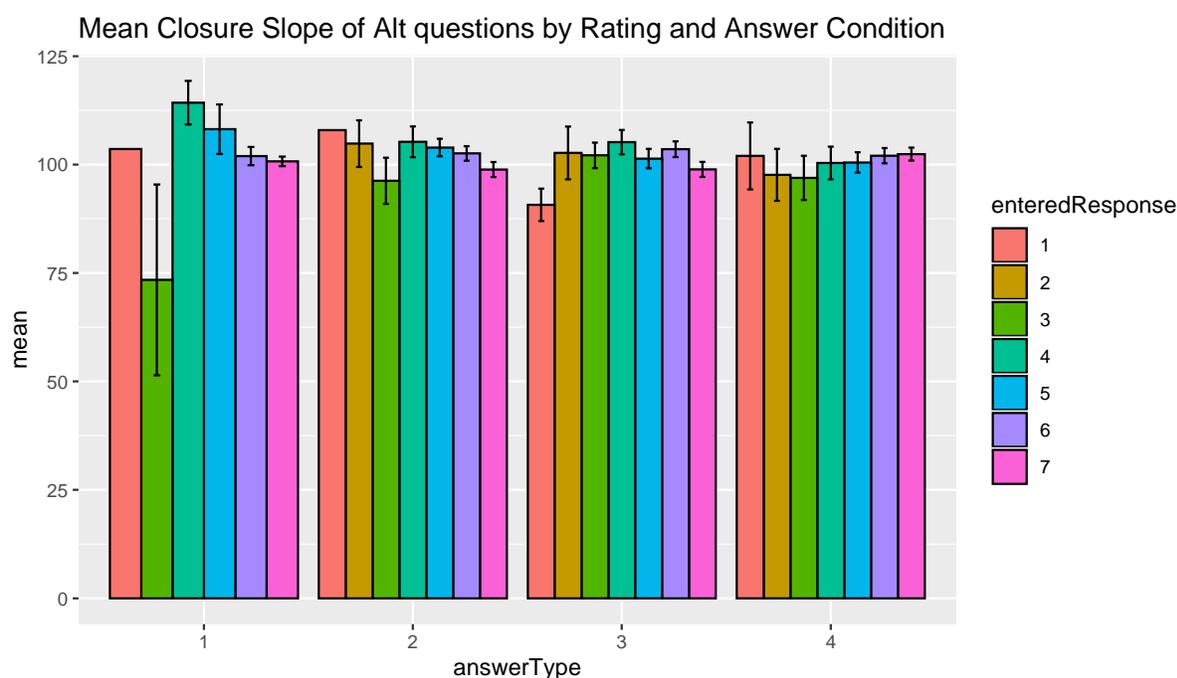


Figure 7.12: Average closure slope on acceptability scale for Alternative Questions. Shown with all answers.. 1= Falling Disjunct; 2 = Rising Disjunct; 3 = Yes + Disjunct; 4 = Cleft Disjunct

The Figure 7.12 indicates that closure slopes do not appear to determine the acceptability of certain answer types. If the closure slope were affecting acceptability, we would expect to see the closure slope increase in answer condition 4 (the cleft disjunct) as the rating increased. We

would expect the opposite effect in answer condition 3 (the Yes + disjunct) as the rating increased. Figure 7.12 demonstrates that this was not observed in the experiment. Even substituting in this measurement for the categorical ‘alternative question’ or ‘Y/N question’ factor in the ordinal regression analysis revealed that the categorical factor, although adding a degree of freedom to the model, was still a better fit ($p < 0.001$)⁶⁷. Thus the closure slope does not correctly predict the variability in response acceptability. Additionally, this would still fail to account for the variability seen in the text only experiment.

Biezma & Rawlins (2012) have claimed that certain typically ‘illicit’ answers (e.g. *both* and *neither* for Alt. questions) are actually licensed in certain contexts. They offer examples such as (104) to indicate that the acceptability of certain responses is dependent on the power dynamic between the participants.

(104) Scenario: A is a waiter, B is a restaurant customer.

A: Would you like coffee or tea?

B: Neither, thanks.

B’: Both, please.

Even here, they suggest that B’ is, if not inappropriate, somewhat unexpected or marked. Thus, the continuous scale of acceptability may lay in the pragmatic application of all answer conditions, or “how easy the power dynamic makes it to reject certain possible

⁶⁷ The model code for the closure slope regression was `enteredResponse ~ answerType * closureSlope + either + disLoc + (1 + answerType | participantID) + (1 + answerType | trialNumber)` and comparing that model in an ANOVA analysis to a model with the categorical disjunct Type instead of closureSlope yielded a p-value of (4.42e-06, ***) in regards to the categorical model.

assumptions(/presuppositions) on the part of the speaker about the context” (Biezma & Rawlins, 2012). Regardless, Inquisitive Semantics has no direct way to account for the continuous scale of acceptability. Table 7.6 (reprinted below) is updated accordingly.

Criteria	Accounted For	Mechanism Used
Alternative Questions can occur without a phrasal break; first disjunct must end high	YES	Allow high accent to delineate end of basic list item in closed list
‘Either’ can occur in both question tunes without altering meaning	YES	Already accounted for in framework
The tune of the answer needs to be taken into account	YES	Rising intonation already interpreted as inquisitive
The presence of ‘either’ in question increases acceptability of illicit answers	YES	‘either’ opens both interpretations; intonation still elevates preferred interpretation
Answers to disjunctive questions on continuous scale	NO	Slope of final fall/rise to indicate strength of closure
Answers are affected by where in the question the disjunct is located	c.f. section 7.8	

Table 7.6 (updated): Criteria list for Inquisitive Semantics.. Includes whether the theory can account for the experimental result and what mechanisms it uses.

7.7.2 Commitment Space Discourse and answers on a continuum

Commitment Space Discourse correctly predicts the best answers for the text experiment⁶⁸, but does not explain why the other answers were still rated as borderline or moderately acceptable. According to CSD, continuations either exist or they don’t; a proper continuation will be considered fully acceptable, while any continuation that does not exist in the commitment space ought to be outright dismissed. Thus, the bare *yes* and *no*, as well as the cleft

⁶⁸ If we accept as true that participants interpreted the text experiment as Y/N questions by default

answer, ought all to be equally unacceptable. However, the results showed that bare *yes* was equivocally worse than both *no* and the cleft answer, which seemed to be rated in a ‘middle’ tier of acceptability. Thus, CSD must account for this difference.

Similarly, the perception experiment revealed a continuum of acceptability for both types of question tunes. An alternative tune was predicted to have two licit answers, but the falling disjunct was significantly better than the cleft answer; the rising disjunct answer was also rated significantly better than the Yes + disjunct answer, although both aren’t licit continuations. A Y/N tune should have three licit answers, but the rising disjunct was significantly less acceptable than the falling disjunct and Yes + disjunct answer. Table 7.11 summarizes the continuum of acceptability that the text and perception experiments reveal, and that CSD must therefore account for.

Categorical Scale	Continuous Scale	Alternative Question	Y/N Question
Licit Answers	Very Acceptable	Falling Disjunct	Falling Disjunct Yes + Disjunct Yes + Both
	Moderately Acceptable	Cleft Disjunct	Rising Disjunct
Illicit Answers	Slightly Acceptable	Rising Disjunct	No Cleft Disjunct
	Barely Acceptable	Yes + Disjunct	Yes

Table 7.11: Continuum of Acceptability, based on text and perception results

The alternative question continuum is based solely on the perception experiment (recall that the text experiment was shown to be biased to a Y/N interpretation). CSD already offers the correct predictions based on the categorical scale; it must therefore offer motivation for why the

falling disjunct is better than the cleft disjunct, and similarly why the rising disjunct is better than the Yes + disjunct answer (though both are ‘illicit’).

The Y/N question continuum was formed by combining the text and perception experiments. The Rising Disjunct was rated as licit (see section 7.5), though it was significantly worse than the other three licit answers; additionally, the cleft answer and bare *no* response were significantly better than the bare *yes* response. Thus, all responses have been ordered on a makeshift continuum of acceptability, based on response ratings.

First, to address the alternative question responses and ordering. Recall the ASSERT operator, that takes in the specific alternative question tune pattern in order to convey the ‘exactly one’ expectation on the part of the speaker. Therefore, a licit response must not only confirm this assertion, but also choose a legal continuation (namely, one of the two disjuncts). It was shown via (89), reprinted below, that simply negating one of the valid continuations, although possibly still confirming ‘exactly one’ disjunct, does not constitute a valid answer.

(89) S1: Did Ed meet Ann[↑] or did Ed meet Beth[↓]?

S2: #He didn’t meet Ann.

[S1 would most likely ask for clarification, e.g. “So he met Beth?”]

Therefore, both must be necessary, and it seems that the valid continuation is more important than the acceptance of the assertion. That is, adhering to both is necessary for a licit answer, but the valid continuation (which needs to be voiced) is also more important than the acceptance of the presumption (which can be done implicitly, c.f. 7.3.2). Therefore, an answer (such as the cleft answer), which is phrased in order to emphasize the ‘exactly one’ assertion (It’s dairy) would not be rated as highly as the falling disjunct answer (She’s allergic to dairy). There seems

to be a preference for the confirmation of the assertion to be done implicitly, which is why examples like (90), reprinted below, aren't as good.

(90) S1: Did Ed meet Ann[↑] or did Ed meet Beth[↓]?

S2: ?He met Ann and not Beth.

This analysis perfectly predicts the acceptability rankings for alternative questions (Table 7.12).

Ranking	Tune	Order of criteria met
Licit- more acceptable	Falling Disjunct	Valid continuation Confirms assertion
Licit- less acceptable	Cleft	Confirms assertion Valid continuation
Illicit- more acceptable	Rising Disjunct	Valid continuation
Illicit- less acceptable	Yes + Disjunct	Confirms assertion

Table 7.12: Rankings of alternative question responses in CSD. Includes the emphasis placed on each of the two necessary components for a valid answer.

The analysis offers a specific prediction that any answer which is not a legal continuation and also rejects the 'exactly one' assertion would be rated even lower than the Yes + Disjunct answer. This is a possible avenue for further research.

Returning to the Y/N rankings, deriving the continuum takes a little more effort. A Y/N question has no additional operator that provides an added level of analysis; answers ought to simply be categorized based on the presence of the continuation or not. The rising disjunct answer, however, has already been analyzed in such a way as to showcase its less acceptable, though still valid, rating (c.f. section 7.5.2). Therefore, CSD merely needs a way in order to demonstrate that the *yes* answer is significantly worse than the *no* and cleft answers.

The cleft answer is a valid answer for the alternative interpretation; therefore, it might be considered ‘not as bad’ due to its acceptability for another type of disjunctive question. What about bare particle *yes* and *no*, however? Already they have been shown to be less acceptable in general (Farkas and Roelofsen 2015), but this does not account for the disparity between them. Of course, Inquisitive Semantics correctly predicts *yes* to be less felicitous than *no* by way of two propositions being present instead of one (Roelofsen and van Gool 2010), so why can this analysis not be extended to CSD, which also analyzes Y/N questions as a disjunction of two monopolar propositions?

Suppose, then, S1 utters sentence (105) with a Y/N tune. Responses (a)-(c) are valid continuations, and thus licit. Response (d), though valid lexically, has the uncertainty contour, which leaves the other proposition in the continuation space, making it less appealing. Responses (e)-(g) are illicit, though (e) and (f) are better than (g).

(105) Do you want to go to the movies or the bars tonight?

- (a) The movies (L-L%)
- (b) Yes, the movies
- (c) Yes, both
- (d) The movies (L*+H L-H%)
- (e) It’s the movies
- (f) No
- (g) Yes

The cleft answer, as described above, seems to emphasize that ‘exactly one’ ought to be chosen, an assertion that S1 did not make. Therefore, it is illicit, even though ‘the movies’ is a valid continuation. Speakers are also allowed to reject an assertion or question; this is ultimately

what *no* amounts to. That is, when (105) is analyzed as a disjunction of monopolar questions, the question contains a *bias* that the speaker believes one or both of the things to be true or chosen. The response *No* in (f) would reject both licit continuations, thus going against the bias.

The response *Yes*, however, is neither a licit continuation for an alternative question nor a licit rejection of the question itself; it is rather a half answer. The response indicates that one (or both) of the propositions is true or is chosen, but the speaker is unwilling to specify which, flagrantly flaunting the Maxim of Quantity. Ultimately, it is the notion that the response is not licit as a rejection or as a response for any type of disjunctive question that relegates answer (g) to the bottom of the ratings.

Thus, using the general principles of disjunction of monopolar questions, as well as the ASSERT operator for alternative questions, Commitment Space Discourse is able to account for the scale of acceptability that the experiments reveal. Table 7.7 is updated accordingly on the next page.

Criteria	Accounted For	Mechanism Used
Alternative Questions can occur without a phrasal break; first disjunct must end high	YES	ASSERT operator on alternative question
'Either' can occur in both question tunes without altering meaning	YES	Already accounted for in framework
The tune of the answer needs to be taken into account	YES	Rising tune in response leaves continuations not added to common ground in commitment space
The presence of 'either' in question increases acceptability of illicit answers	YES	'Either' increases ambiguity; alters ASSERT effectiveness
Answers to disjunctive questions on continuous scale	YES	Ranking of criteria that must be met for each question type
Answers are affected by where in the question the disjunct is located	c.f. section 7.8	

Table 7.7 (updated): Criteria list for Commitment Space Discourse. Includes whether the theory can account for the experimental result and what mechanisms it uses.

7.8 Location of Disjunction in Question

A surprising result of the experiments revealed participant bias based on where in the question the disjunct occurred. There wasn't a clear delineation on how the location of the disjunct would affect the acceptability of the answer, even within the same question type. For alternative questions, acceptability increased for both the falling and rising disjunct as the disjunct location moved from the beginning to the end of the sentence (c.f. Figure 7.1); however, the cleft disjunct answer *decreased* in acceptability as the location of the disjunct moved from beginning to end (c.f. Figure 7.3). Therefore, the location of the disjunct affected the two licit answer conditions differently. For Y/N questions, the 'very acceptable' responses (Falling disjunct, Yes + Disjunct, Yes + both) also did not share the same pattern (Figure 7.13).

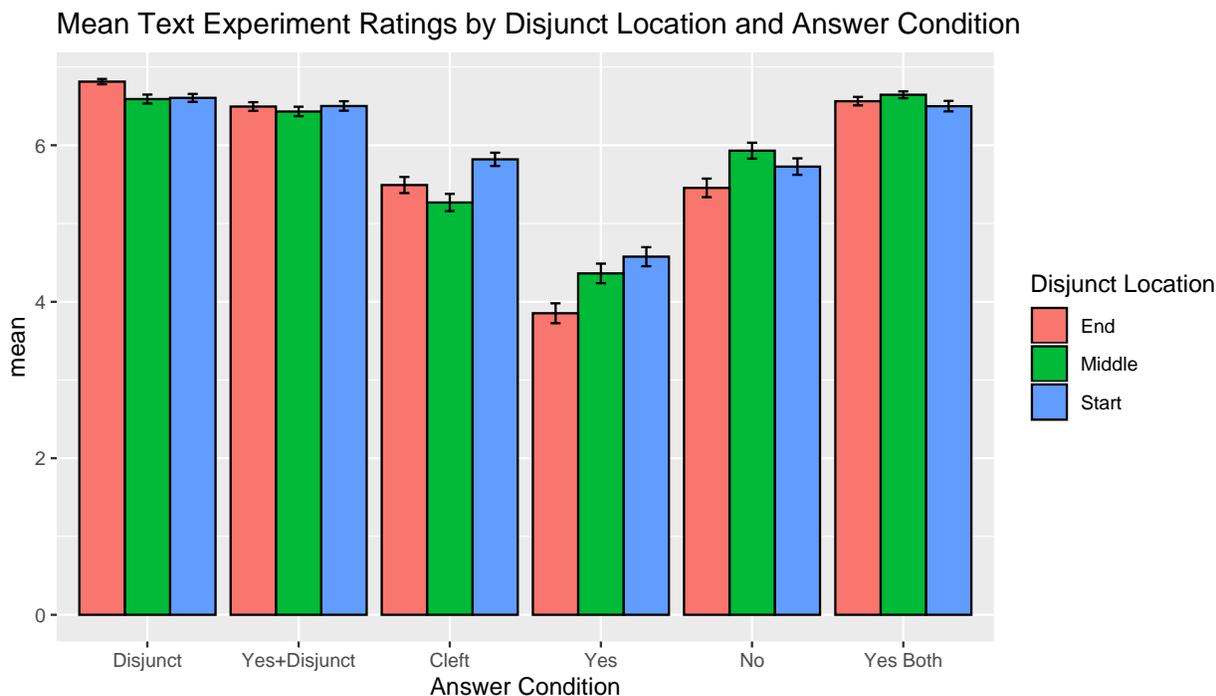


Figure 7.13: Ratings averaged over disjunct location in question, by answer condition.

Participants rated the falling disjunct response the best when the question disjunction was utterance initial or utterance final but rated the “Yes + both” response best when the question disjunction was utterance middle. It is clear that the responses remain in their separate tiers of acceptability, despite the location of the disjunct; however, the ranking, or preference, within those tiers changes with the disjunct location. For example, the *No* response was rated significantly better than the cleft response when the disjunct occurred in the middle of the sentence; in the other locations, the ratings were comparable. Finally, all the questions contained NP disjunction, so the syntactic category did not contribute to acceptability. Importantly, the disjunct location was found to be significant in both the text and perception experiments. So how exactly does placement of the disjunct affect response acceptability?

7.8.1 Inquisitive Semantics and disjunct location

Although the closure slope discussed in the last section could not account for the differences in acceptability between similarly licit answers and similarly illicit answers, there does seem to be a relationship between the closure slope and the location of the disjunct. Table 7.13 reveals that alternative questions had a much smaller ‘closure slope’ when the disjunct occurred at the end of the utterance, versus the beginning or middle. This smaller slope would allow IS to correctly predict that the cleft answer would be rated lower at the end of the disjunct vs. the beginning or middle.

Disjunct Location	Answer Condition			
	Falling Disjunct	Rising Disjunct	Yes + Disjunct	Cleft Disjunct
Initial	108.53	108.81	108.35	108.54
Middle	111.01	110.88	111.13	111.01
Final	85.12	85.58	85.35	85.31

Table 7.13: Average ‘closure slope’ of alternative question. Collpased by location of disjunct and answer condition.

This closure slope also correctly predicts that the rising disjunct would have the highest rating when the disjunct occurred at the end of the utterance (i.e. a weakened ‘exactly one’ assertion would allow the uncertainty contour). However, the closure slope is comparable when the disjunct occurred in the middle vs the beginning of the sentence, so the variable cannot account for the difference in cleft answer acceptability in those two locations, as well as the “Yes + Disjunct” answer; both answer conditions were significantly better when the disjunct occurred utterance initial vs. the middle of the utterance, even though the cleft answer is ‘licit’ while the “Yes + disjunct” answer is not.

It is possible that the difference in disjunct location acceptability is a combination of interactions of different prosodic cues, the contour slope merely being one of them.

Unfortunately, there is no mechanism in IS that would account for such variability based on prosody and location of the disjunct; therefore, we must conclude that this is an area for further research. Table 7.6 is updated below to reflect the final criterion for Inquisitive Semantics to account for.

Criteria	Accounted For	Mechanism Used
Alternative Questions can occur without a phrasal break; first disjunct must end high	YES	Allow high accent to delineate end of basic list item in closed list
<i>either</i> can occur in both question tunes without altering meaning	YES	Already accounted for in framework
The tune of the answer needs to be taken into account	YES	Rising intonation already interpreted as inquisitive
The presence of <i>either</i> in question increases acceptability of illicit answers	YES	<i>either</i> opens both interpretations; intonation still elevates preferred interpretation
Answers to disjunctive questions on continuous scale	NO	Slope of final fall/rise to indicate strength of closure
Answers are affected by where in the question the disjunct is located	NO	Closure Slope explains some variation, but not all

Table 7.6 (updated): Criteria list for Inquisitive Semantics. Includes whether the theory can account for the experimental result and what mechanisms it uses.

7.8.2 Commitment Space Discourse and question disjunct location

Commitment Space Discourse was able to account for the variability of response acceptance via a ranking system of the ASSERT mechanism and licit continuation criteria for alternative questions, and through an analysis of monopolar disjunction interpretation. However, it is unclear how these factors may account for the differing acceptability when the location of

the disjunct in the question is altered. Syntactically, Krifka (2015, 2016) interprets monopolar questions via an Act phrase head [_{Act} ?], which requests that the commitment denoted by the complement of the ActP to be performed by the addressee (this is the mechanism by which S1 requests S2 to commit to a proposition). However, as previously noted, the complement of the ActP in the experiments was always a NP (e.g. Melanie or Emily, dairy or soy, etc.), thereby eliminating the complement space as a factor in accounting for the variability. Possibly the location of the ActP in the TP can affect acceptability, though the disjuncts scope out to the SpecActP on the logical form level (Krifka 2015), which is how the disjunctive question is analyzed as a disjunction over speech acts. Additionally, a corpus study has shown that NP-coordinations are produced far more often than S-coordinations, and that participants prefer readings with NP-coordination when faced with NP versus sentence coordination (Hoeks et. al. 2006). As coordination and disjunction have been shown to exhibit similar behaviors within a language (Haspelmath 2007), one would think that the frequency of NP coordination would nullify any sort of parsing issue based on unusual sentence structure.

If, indeed, the variability associated with the location of the disjunct lies in the complexity or form of the syntactic analysis, then one would need to account for the similar ratings across disjunct location in the text experiment (text only) and the perception experiment (audio only). The text experiment and Y/N questions for the perception experiment had the same pattern of acceptability across similar answer types for the three distinct disjunct locations, indicating that the variability, at least for the ‘default’ disjunctive question, cannot be due to specific prosodic cues (from the perception experiment) or some type of reading ambiguity (from the text experiment), but rather something that both experiments have in common. Commitment Space Discourse does integrate the semantics of disjunction questions with the prosody and the

syntax but does not directly provide reasons or predictions for variable acceptability of monopolar questions based on where in the utterance the disjunction occurs. This is an area for further research. Table 7.7 is reprinted below to reflect the final criterion.

Criteria	Accounted For	Mechanism Used
Alternative Questions can occur without a phrasal break; first disjunct must end high	YES	ASSERT operator on alternative question
'Either' can occur in both question tunes without altering meaning	YES	Already accounted for in framework
The tune of the answer needs to be taken into account	YES	Rising tune in response leaves continuations not added to common ground in commitment space
The presence of 'either' in question increases acceptability of illicit answers	YES	'Either' increases ambiguity; alters ASSERT effectiveness
Answers to disjunctive questions on continuous scale	YES	Ranking of criteria that must be met for each question type
Answers are affected by where in the question the disjunct is located	NO	ActP does not account for different continuations based on disjunct location

Table 7.7 (updated): Criteria list for Commitment Space Discourse. Includes whether the theory can account for the experimental result and what mechanisms it uses.

7.9 Conclusion and Further Work

Disjunctive questions can be interpreted as their default, a Y/N question with disjunction, or as a special subset- alternative questions. Rather than being unbiased in text form, the default interpretation occurs; only through prosodic contour does the alternative interpretation occur. Specifically, the first disjunct must end in the high pitch range, whether thru a pitch accent or phrasal accent, and the phrase must end in the low pitch range. The assumption of a necessary phrase break between disjuncts for Alt. questions was disproved. Historically, inserting *either* was thought to disambiguate the two interpretations, disallowing an Alt. question interpretation; this assumption was shown to be false. While *either* can occur with both interpretations, its

presence does allow typically illicit responses for that interpretation to become more acceptable; this is especially true when *either* is paired with an alternative question contour. The tune of the response also affects the acceptability of the response in regard to the question; a licit continuation with the uncertainty contour will make the response less acceptable (with a Y/N question) or no longer acceptable (with an alternative question). A cleft answer, pragmatically exhaustive, will only be licit for an Alt. question. Therefore, previous assumptions concerning the acceptability of any disjunct answer (regardless of syntactic structure or tune) was dismissed. In fact, rather than a categorical scale of ‘licit’ and ‘illicit’, the responses seemed to adhere to a continuous scale of acceptability. Finally, the location of the disjunct in the question was also shown to affect answerhood conditions.

Two theories, both of which were developed in order to handle not only questions but their licit responses, were examined in light of the empirical results. The initial predictions by both theories were not entirely accurate, as they were formed with certain assumptions in mind which were shown to be false. For example, neither theory portrayed the alternative contour appropriately, and both had definitive categorical vs. continuous distinctions in answerhood acceptability. Inquisitive Semantics was able to use prosodic accents and the suggesting and highlighting mechanisms in order to account for the empirically derived alternative tune pattern, as well as how *either* affects response acceptability. The slope of the ‘closure’ predicted some of the variation occurring when the disjunct occurred utterance final in alternative questions; overall, however, Inquisitive Semantics lacks the mechanisms to grade responses on a continuous, rather than categorical, scale.

Commitment Space Discourse was also successful, using the ASSERT operator, in modeling the alternative question contour. It is this operator that interacts with ambiguous *either*

that allows Y/N-only responses to receive higher acceptability ratings. Additionally, the analysis of disjunctive questions as a disjunction of monopolar questions, as well as the ASSERT operator in the case of alternative questions, allows Commitment Space Discourse to account for a continuous scale of acceptability for the responses studied in this paper. Like Inquisitive Semantics, however, Commitment Space Discourse fails to predict the significance of disjunct location.

It is possible that the difference in disjunct location acceptability is a combination of interactions of different prosodic cues; Inquisitive Semantics revealed that the contour slope is potentially one of them. Perhaps there is a relationship between interpretation and the amount of distance between the fall on the second disjunct and the end of the utterance. Heidenreich (2014b) conducted the first empirical study on disjunction questions when the disjunct was not utterance-final; therefore, much of relationship between location of the disjunction and response acceptability remains new and uncharted territory. The factors that drive such variability have yet to be determined; additionally, current question-driven Semantic theories must find a way to account for this difference.

Finally, the superset/subset relationship between Y/N questions and Alt. questions was supported by the text and artificial language experiments. However, there was outstanding evidence that suggested the disjunctive connective(s) used in disambiguating the two types of disjunction may have a different relationship when inserting negation and declaratives. This is also an area for further work.

References

- Adank, P., Smits, R., & Van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America*, 116(5), 3099-3107.
- Aloni, Maria, and Robert van Rooy. (2002). The dynamics of questions and focus. In *Proceedings of SALT 12*, ed. by Brendan Jackson. Available at <http://elanguage.net/journals/salt/issue/view/295>.
- Alonso-Ovalle, L. (2006). Disjunction in Alternative Semantics. Ph. D. thesis, University of Massachusetts, Amherst, MA.
- Ariel, M., & Mauri, C. (2019). An ‘alternative’ core for or. *Journal of Pragmatics*, 149, 40-59.
- Bartels, Christine. (1999). *The intonation of English statements and questions: a compositional interpretation*. Routledge.
- Beck, Sigrid, and Shin-Sook Kim. (2006). Intervention effects in alternative questions. *The Journal of Comparative Germanic Linguistics* 9:165–208.
- Beckman, Mary E. (1996). The Parsing of Prosody. *Language and Cognitive Processes*, 11:1-2, 17-68. DOI: 10.0180/016909696387213.
- Beckman, Mary, and Julia Hirschberg. (1994). *The ToBI Annotation Conventions*. Columbus, OH: Ohio State University.
- Biezma, M. (2009). Alternative vs polar questions: The cornering effect. In *Proceedings of SALT 19*, ed. by Ed Cormany, Satoshi Ito, and David Lutz, 37–54. Available at <http://elanguage.net/journals/salt / issue / view / 278>.
- Biezma, M., & Rawlins, K. (2012). Responding to alternative and polar questions. *Linguistics and Philosophy*, 35(5), 361-406. Doi: 10.1007/s10988-012-9123-z.

- Brennan, S. E. (1995). Centering attention in discourse. *Language and Cognitive processes*, 10(2), 137-167.
- Christensen, R.H.B. (2019). Cumulative Link Models for Ordinal Regression with the R Package *ordinal*. To appear in *Journal of Statistical Software*.
- Ciaredelli, I., Groenendijk, J., & Roelofsen, F. (2013a). *Inquisitive semantics: A new notion of meaning*. *Language and Linguistics Compass*, 7(9), 459-476. Doi:10.1111/lnc3.12037.
- Ciaredelli, I., Groenendijk, J., & Roelofsen, F. (2013b). Towards a logic of information exchange: An inquisitive witness semantics. In G. Bezhanishvili, V. Marra, S. Löbner, & F. Richter (Eds.), *Logic, Language, and Computation: Revised Selected Papers from the Ninth International Tbilisi Symposium on Logic, Language and Computation*. Berlin: Springer.
- Ciaredelli, I., Groenendijk, J., & Roelofsen, F. (2015). On the semantics and logic of declaratives and interrogatives. *Synthese* 192:1689-1728.
- Cohen, Ariel & Manfred Krifka. (2014). Superlative quantifiers and meta-speech acts. *Linguistics and Philosophy* 37, 41-90.
- Cutler, A. (1976). Beyond parsing and lexical look-up.
- Cutler, A., Dahan, D., & Van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and speech*, 40(2), 141-201.
- De Veugh-Geiss, Joseph P., Swantje Tönnis, Edgar Onea & Malte Zimmermann. (2018). That's not quite it: An experimental investigation of (non-)exhaustivity in clefts. *Semantics and Pragmatics* 11(3). <https://doi.org/10.3765/sp.11.3>.
- Duan, Manjuan. (2015). A categorical analysis of Chinese alternative questions. In *Proceedings for ESSLLI 2015 Workshop 'Empirical Advances in Categorical Grammar' (CG 2015)*.

- Farkas, Donka F. & Floris Roelofsen. (2015). Polar initiatives and polar particle responses as a window onto the interpretation of questions and assertions. *Language* 91, 359-414. Doi: <http://dx.doi.org/10.1353/ln.2015.0017>.
- Featherston, S. (2005). Magnitude estimation and what it can do for your syntax: Some wh-constraints in German. *Lingua*, 115, 11, 1525–1550.
- Frazier, Lyn, Katy Carlson, and Charles Clifton Jr. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences*, Vol. 10, No. 6.
- German, J., Pierrehumbert, J. B., & Kaufmann, S. (2006). Evidence for phonological constraints on nuclear accent placement. *Language*, 82(1), 151-168.
- Gibson, Edward, Steve Piantadosi, and Kristina Fedorenko. (2011). “Using Mechanical Turk to Obtain and Analyze English Acceptability Judgments.” *Language and Linguistics Compass* 5.8: 509–524. Web.
- Grice, H. P. (1975). Logic and conversation. In P. Cole and J. J. Morgan, eds. *Syntax and semantics, volume 3: speech acts*. New York: Academic Press.
- Haida, A. (2010). On the Semantics and Pragmatics of Yes / No-Questions , Yes / No-Question Disjunctions , and Alternative Questions Evidence from Chadic. Draft version of October 8, 2010.
- Hamblin, C. L. (1971). Mathematical models of dialogue 1. *Theoria*, 37(2), 130-155.
- Hamblin, C. L. (1973). Questions in Montague English. *Foundations of Language*, 10, 41-53.
- Han, C. & Romero, M. (2004a). Disjunction, focus, and scope. *Linguistic Inquiry* 35: 179-217.
- Han, C. & Romero, M. (2004b). The syntax of *whether/Q . . . or* questions: Ellipsis combined with movement. *Natural Language and Linguistic Theory* 22:527–564.
- Haspelmath, M. (2007). Coordination. *Language typology and syntactic description* vol. II:

Complex constructions. 2nd ed.:1-51.

- Hedberg, Nancy, Juan M. Sosa, and Lorna Fadden. (2004). Meanings and configurations of questions in English. In *Proceedings of Speech Prosody 2004* (Nara, Japan), 309–312.
- Heidenreich, S. (2014a). *Models in Morphological Learning*. Master's Thesis, Northwestern University, Evanston, IL.
- Heidenreich, S. (2014b). *Disjunctive Questions: Disjunctive Meanings*. Qualifying Paper, Northwestern University, Evanston, IL.
- Hoeks, John C.J., Petra Hendriks, Wietske Vonk, Colin M. Brown & Peter Hagoort. (2006). Processing the noun phrase versus sentence coordination ambiguity: Thematic information does not completely eliminate processing difficulty, *The Quarterly Journal of Experimental Psychology*, 59:9, 1581-1599, DOI: [10.1080/17470210500268982](https://doi.org/10.1080/17470210500268982)
- Horn, Larry. (1981). Exhaustiveness and the semantics of clefts. In Victoria E. Burke & James Pustejovsky (eds.), *North Eastern Linguistic Society (NELS)*, vol. 11, 125-142. Amherst: University of Massachusetts.
- Horn, Larry. (2004). Information structure and the landscape of (non-)at-issue meaning. In Caroline Féry & Shinichiro Ishihara (eds.), *The Oxford handbook of information structure*, Oxford University Press.
<https://doi.org/10.1093/oxfordhb/9780199642670.013.009>. Online Publication Date: Dec 2014.
- Huddleston, R. (1994). The contrast between interrogatives and questions. *Journal of Linguistics* 30:411-439.
- Hurford, J. (1974). Exclusive or inclusive disjunction. *Foundations of Language* 11(3):409-411.
- Jaeger, T. Florian. (2008). *Categorical data analysis: Away from ANOVAs (transformation or*

- not) and towards logit mixed models. *Journal of Memory and Language* 59:434–446.
- Karttunen, L. and Peters, S. (1976). What indirect questions conventionally implicate. In *CLS 12: Papers from the Twelfth Regional Meeting*, pages 351-368. Chicago Linguistic Society.
- Kiss, Katalin E. 1998. Identificational Focus versus Information Focus. *Language* 74: 245-273.
- Kratzer, A. and Shimoyama, J. (2002). Indeterminate pronouns: The view from Japanese. In Y. Otsu, editor, *The Proceedings of the Third Tokyo Conference on Psycholinguistics*, pages 1-25.
- Krifka, M. (2015). Bias in Commitment Space Semantics: Declarative questions, negated questions, and questions tags. *Proceedings of SALT 25*: 328-345.
- Krifka, M. (2016). Constituent, alternative, and yes/no questions as multipolar, bipolar, and monopolar requests for assertions. Workshop on Questions and Enquiry, King's College, London.
- Krifka, M. (2017). Negated polarity questions as denegations of assertions. In *Contrastiveness in information structure, alternatives and scalar implicatures* (pp. 359-398). Springer, Cham.
- Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. *Glossa*, 7, 107-122.
- Levy, R. (2007a). Linguistics 251 lecture notes, Fall 2007. Lecture 14: Introduction to generalized mixed-effects models.
- Levy, R. (2007b). Linguistics 251 lecture notes, Fall 2007. Lecture 15: Mixed-effects logistic regression.
- Li, Charles and Sandra Thompson. (1981). *Mandarin Chinese: A Functional Reference Grammar*. Berkeley: University of California Press.

- Martin, Scott & Carl Pollard. (2014). A dynamic categorial grammar. Paper presented at Formal Grammar 2014.
- Mauri, C. (2008). The irreality of alternatives: Towards a typology of disjunction. *Studies in Language* 32:1; 22-55.
- Merlan, F. (1982). *Magarayi*. Amsterdam: North-Holland.
- Nolan, F. (2003). Intonational equivalence: an experimental evaluation of pitch scales. In *Proceedings of the 15th international congress of phonetic sciences* (Vol. 771, p. 774). Barcelona, Spain: International Phonetic Association.
- Ostendorf, M., P.J. Price, S. Shattuck-Hufnagel. (1995). The Boston University Radio News Corpus. Technical Report ECS-95-001, Electrical, Computer and Systems Engineering Department, Boston University, Boston, MA.
- Pierrehumbert, Janet. (1980). The phonology and phonetics of English intonation. Doctoral dissertation, MIT, Cambridge, MA.
- Pierrehumbert, Janet, and Julia Hirschberg. (1990). The meaning of intonational contours in the interpretation of discourse. In *Intentions in communication*, ed. by Philip R. Cohen, Jerry Morgan, and Martha E. Pollack, 271–311. Cambridge, MA: MIT Press.
- Percus, Orin. (1997). Prying open the cleft. In Kiyomi Kusumoto (ed.), *North Eastern Linguistic Society (NELS)*, vol. 27, 337-351.
- Pinker, S. (1994). *The Language Instinct*. New York: Morrow.
- Pollard, Carl & Murat Yasavul. (2015). Anaphoric *it*-clefts: The myth of exhaustivity, <https://web.archive.org/web/20170514171536/http://www.ling.ohiostate.edu/~yasavul.1/c1s50paperfinal.pdf>. To appear in *Proceedings of the Chicago Linguistic Society (CLS)* 50.
- Price, P.J., Ostendorf, S., Shattuck-Hufnagel, S. & Fong, C. (1991). The use of prosody in

- syntactic disambiguation. *Journal of the Acoustical Society of America*, 9, 2956-2970.
- Prieto, P. & Borràs-Comes, J. (2018). *Natural Language & Linguistic Theory*, vol 36: 563-586.
<https://doi.org/10.1007/s11049-017-9382-z>
- Pruitt, K., & Roelofsen, F. (2013). The interpretation of prosody in disjunctive questions.
Linguistic inquiry, 44(4), 632-650.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A Comprehensive Grammar of the English language*. London: Longman.
- Rando, E. (1980). Intonation in discourse. *The melody of language*, 243-278.
- Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC Nonword Database. *Quarterly Journal of Experimental Psychology*, 55A, 1339-1362.
- Rawlins, K. (2008). (Un)conditionals: An investigation in the syntax and semantics of conditional structures. Doctoral dissertation, University of California, Santa Cruz.
- Roelofsen, F. (2013a). Algebraic foundations for the semantic treatment of inquisitive content. *Synthese* 190.1.79-102.
- Roelofsen, F. (2013b). An inquisitive perspective on meaning: The case of disjunction. Colloquium talk given at Stanford, Rochester, UMass Amherst, University College London, and the University of Stuttgart.
- Roelofsen, Floris, and Kathryn Pruitt. (2011). Disjunctive questions: Prosody, syntax, and semantics. Presented at a seminar at the Georg August Universität Göttingen, April 2011. Available at <http://www.illc.uva.nl/inquisitive-semantics>.
- Roelofsen, F., & van Gool, S. (2010). Disjunctive questions, intonation, and highlighting. In M. Aloni, H. Bastiaanse, T. de Jager, & K. Schulz (Eds.), *Logic, language, and meaning*:

- Selected papers from the seventeenth Amsterdam colloquium (pp. 384–394). Amsterdam: Springer.
- Romero, Maribel, and Chung-hye Han. (2003). Focus, ellipsis and the semantics of alternative questions. In *Empirical issues in formal syntax and semantics 4*, ed. by Claire Beyssade, Olivier Bonami, Patricia Cabredo Hofherr, and Francis Corblin, 291–307. Paris: Presses Universitaires de Paris-Sorbonne.
- Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*. 1:75-116
- Saeed, J. (1993). *Somali reference grammar* (2nd revised edition). Kensington MD: Dunwoody Press.
- Schafer, A.J. (1997). Prosodic Parsing: The role of prosody in sentence comprehension. Doctoral Dissertation, University of Massachusetts Amherst.
- Schafer, A.J., Speer, S., Warren, P., and White, S.D. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research*, Vol. 29, No. 2, 2000
- Schubiger, M. (1958). *English intonation, its form and function*. M. Niemeyer Verlag.
- Schütze, C. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press.
- Simpson, A. P. (2009). Phonetic differences between male and female speech. *Language and linguistics compass*, 3(2), 621-640.
- Sohn, H. (2001). *Korean*. London: Routledge.
- Sulkala, H and M. Karjalainen. (1992). *Finnish*, Routledge, London and New York.
- Szabolcsi, Anna. (1982). Model theoretic semantics of performatives. In: Kiefer, Ferenc (ed.),

Hungarian Linguistics, 515-535. Amsterdam: John Benjamins. Doi:

<http://dx.doi.org/10.1075/llsee.4.16sza>

- Terken, J., & Nootboom, S. G. (1987). Opposite effects of accentuation and deaccentuation on verification latencies for given and new information. *Language and cognitive processes*, 2(3-4), 145-163.
- Truckenbrodt, Hubert. (2015). Intonation phrases and speech acts. In: Kluck, Marlies, Dennis Ott & Mark de Vries (eds.), *Parenthesis and ellipsis: Crosslinguistic and theoretical perspectives*, 301-349. Berlin: Walter de Gruyter. doi:
<http://dx.doi.org/10.1515/9781614514831.301>.
- Vainikka, (1987). ‘Why can *Or* mean *And* or *Or*?’, in J. Blevins and A. Vainikka (eds.), *UMOP 12. Issues in Semantics*, vol. 12, GLSA, Amherst, MA., pp. 148-186.
- Velleman, Dan, David Beaver, Emilie Destruel, Dylan Bumford, Edgar Onea, & Liz Coppock. (2012). It-clefts are IT (inquiry terminating) constructions. In *Proceedings of SALT 22*, 441–460.
- Venables, W. N. & Ripley, B. D. (2002) *Modern Applied Statistics with S*. Fourth Edition. Springer, New York. ISBN 0-387-95457-0
- Ward, G., & Hirschberg, J. (1985). Implicating uncertainty: The pragmatics of fall-rise intonation. *Language*, 747-776.
- Warren, P. (1985). *The Temporal Organization and Perception of Speech*. Doctoral Dissertation, Cambridge University.
- Wasow, T. & Arnold, J. (2005). Intuitions in linguistic argumentation. *Lingua*, 115, 1481-1496.
- White, M., Rajkumar, R., Ito, K., & Speer, S. R. (2009). Eye tracking for the online evaluation of prosody in speech synthesis: Not so fast!. In *INTERSPEECH* (pp. 2523-2526).

Wightman, C., Price, P., Pierrehumbert, J., & Hirschberg, J. (1992), October. ToBI: A Standard for labeling English prosody. In Proceedings of the 1992 International Conference on Spoken Language Processing, ICSLP (pp.12-16).

Wilson, Colin (2006). Learning Phonology With Substantive Bias: An Experimental and Computational Study of Velar Palatalization. *Cognitive Science* 30, 945-982.

Winans, L. (2012). Alternative Disjunctions in Egyptian Arabic. Master's Thesis, University of California, Los Angeles.

Appendix A: Text Experiment Stimuli

1. Did (either) Melanie or Emily go to the party tonight?
 1. Melanie did.
 2. Yes, Melanie did.
 3. It was Melanie.
 4. Yes.
 5. No
 6. Yes, they both did.
2. Will (either) Isabel or Andrea go to the baseball game today?
 - 1: Andrea will go.
 - 2: Yes, Andrea will go
 3. It'll be Andrea
 4. Yes
 5. No
 6. Yes, they will both go.
3. Did you invite (either) Annabelle or Adrian out to the movies?
 1. I invited Adrian.
 2. Yes, I invited Adrian.
 3. It was Adrian.
 4. Yes
 5. No
 6. Yes, I invited them both.
4. Are (either) podiatrists or ophthalmologists prejudiced against nurses?
 1. Podiatrists are.
 2. Yes, podiatrists are.
 3. It's podiatrists.
 4. Yes
 5. No
 6. Yes, they both are.
5. Did Lauren see (either) Angelo or Julia perform at the tap dancing recital?
 1. She saw Julia.
 2. Yes, she saw Julia.
 3. It was Julia.
 4. Yes
 5. No
 6. Yes, she saw them both.
6. Do (either) botanists or pharmacists attend the college of Arts and Sciences?
 1. Botanists do.
 2. Yes, botanists do.
 3. It's botanists.
 4. Yes
 5. No
 6. Yes, they both do.
7. Did (either) Dario or Gabriel get a puppy this summer?
 1. Gabriel did.

2. Yes, Gabriel did.
 3. It was Gabriel.
 4. Yes
 5. No
 6. Yes, they both did.
8. Did (either) journalists or scientists write the article?
1. Scientists wrote it.
 2. Yes, scientists wrote it.
 3. It was scientists.
 4. Yes
 5. No
 6. Yes, they both wrote it.
9. Did (either) Chelsey meet with urologists or neurologists during her hospital visit?
1. She met with neurologists.
 2. Yes, she met with neurologists.
 3. It was neurologists
 4. Yes
 5. No
 6. Yes, she met with both.
10. Did (either) the new school principal meet with Mario or Darius before class?
1. He met with Mario.
 2. Yes, he met with Mario.
 3. It was Mario
 4. Yes
 5. No
 6. Yes, he met with both of them.
11. Did (either) Solomon or Sullivan vandalize the old warehouse last night?
1. Sullivan did.
 2. Yes, Sullivan did
 3. It was Sullivan.
 4. Yes
 5. No
 6. Yes, they both did.
12. Does the senator like (either) lobbyists or activists knocking on his door?
1. He likes lobbyists
 2. Yes, he likes lobbyists
 3. It's lobbyists.
 4. Yes
 5. No
 6. Yes, he likes both.
13. Is Leah listening to (either) philosophers or psychologists at the conference this weekend?
1. She's listening to psychologists.
 2. Yes, she's listening to psychologists.
 3. It's psychologists.
 4. Yes

5. No
 6. Yes, she's listening to both.
14. Is Martha videotaping (either) Lillian or Dorian in the school play?
1. She's videotaping Lillian
 2. Yes, she's videotaping Lillian.
 3. It's Lillian
 4. Yes
 5. No
 6. Yes, she's videotaping them both.
15. Did (either) Cameron or Jillian feed the dog yesterday?
1. Cameron did
 2. Yes, Cameron did
 3. It was Cameron
 4. Yes
 5. No
 6. Yes, they both did.
16. Is Alex inviting (either) Everett or Ellery to spend the night?
1. He's inviting Everett
 2. Yes, he's inviting Everett
 3. It's Everett
 4. Yes
 5. No
 6. Yes, he's inviting both of them.
17. Did (either) Elliot or Emery see the baby elephant at the zoo?
1. Elliot did.
 2. Yes, Elliot did.
 3. It was Elliot.
 4. Yes
 5. No
 6. Yes, they both did.
18. Did (either) Madeline or Mallory enjoy going to the opera?
1. Mallory did.
 2. Yes, Mallory did
 3. It was Mallory
 4. Yes
 5. No
 6. Yes, they both did.
19. Did (either) Gregory or Romeo bring dessert?
1. Gregory did
 2. Yes, Gregory did
 3. It was Gregory
 4. Yes
 5. No
 6. Yes, they both did.

20. Did Leslie overhear (either) Jeremy or Oliver ordering pizza?
1. She overheard Jeremy
 2. Yes, she overheard Jeremy
 3. It was Jeremy
 4. Yes
 5. No
 6. Yes, she overheard them both.
21. Did Amanda see (either) Ariel or Miriam at a city council meeting?
1. She saw Miriam
 2. Yes, she saw Miriam
 3. It was Miriam
 4. Yes
 5. No
 6. Yes, she saw both of them.
22. Do you see (either) Damian or Beverly on the beach?
1. I see Beverly
 2. Yes, I see Beverly
 3. It's Beverly
 4. Yes
 5. No
 6. Yes, I see both of them.
23. Did (either) Julian or Pamela fly to Florida for spring break?
1. Julian did
 2. Yes, Julian did
 3. It was Julian
 4. Yes
 5. No
 6. Yes, they both did.
24. Did Martin go with (either) Allison or Julius to the amusement park today?
1. He went with Allison
 2. Yes, he went with Allison
 3. It was Allison
 4. Yes
 5. No
 6. Yes, he went with them both.
25. Did (either) Angela or Julio stop by to drop off their keys?
1. Julio stopped by
 2. Yes, Julio stopped by
 3. It was Julio
 4. Yes
 5. No
 6. Yes, they both stopped by.
26. Did Dan go with (either) Emerson or Benjamin to the new restaurant in town?
1. He went with Benjamin
 2. Yes, he went with Benjamin

3. It was Benjamin
 4. Yes
 5. No
 6. Yes, he went with both of them.
27. Is Mary allergic to (either) dairy or soy?
1. She's allergic to dairy.
 2. Yes, she's allergic to dairy
 3. It's dairy
 4. Yes
 5. No
 6. Yes, she's allergic to both.
28. Was Samantha going to feed (either) the turtle or the lizard?
1. She was going to feed the turtle
 2. Yes, she was going to feed the turtle
 3. It was the turtle
 4. Yes
 5. No
 6. Yes, she was going to feed them both.
29. Did William send (either) an email or a letter?
1. He sent a letter
 2. Yes, he sent a letter
 3. It was a letter.
 4. Yes
 5. No
 6. Yes, he sent both.
30. Is Frank finishing (either) homework or the online quiz?
1. He's finishing homework.
 2. Yes, he's finishing homework
 3. It's homework.
 4. Yes
 5. No
 6. Yes, he's finishing both.
31. Did Alan buy her (either) chocolates or flowers?
1. He bought her flowers
 2. Yes, he bought her flowers
 3. It was flowers
 4. Yes
 5. No
 6. Yes, he bought her both.
32. Does Amy use (either) Hulu or Netflix?
1. She uses Netflix
 2. Yes, she uses Netflix
 3. It's Netflix
 4. Yes
 5. No

6. Yes, she uses both.
33. Do the kids need (either) a snack or a bathroom break?
 1. They need a snack
 2. Yes, they need a snack
 3. It's a snack
 4. Yes
 5. No
 6. Yes, they need both.
34. Did you order (either) mocha or French vanilla?
 1. I ordered mocha
 2. Yes, I ordered mocha
 3. It was mocha
 4. Yes
 5. No
 6. Yes, I ordered both.
35. Does Dave plan to take (either) sick days or vacation days?
 1. He plans to take vacation days.
 2. Yes, he plans to take vacation days.
 3. It's vacation days
 4. Yes
 5. No
 6. Yes, he plans to take both.
36. Would Jordi like to visit (either) the zoo or the aquarium?
 1. He wants to visit the zoo
 2. Yes, he wants to visit the zoo
 3. It's the zoo
 4. Yes
 5. No
 6. Yes, he wants to visit both.
37. Does Heather play (either) soccer or softball?
 1. She plays soccer
 2. Yes, she plays soccer
 3. It's soccer.
 4. Yes
 5. No
 6. Yes, she plays both.
38. Did Mike wear (either) a hat or sunglasses?
 1. He wore sunglasses
 2. Yes, he wore sunglasses
 3. It was sunglasses
 4. Yes
 5. No
 6. Yes, he wore both.
39. Would you like (either) Ranch dressing or Thousand Island?
 1. I would like Ranch dressing

2. Yes, I would like Ranch dressing
 3. It's Ranch dressing
 4. Yes
 5. No
 6. Yes, I would like both.
40. Did Travis bring (either) drinks or dessert?
1. He brought dessert
 2. Yes, he brought dessert
 3. It was dessert
 4. Yes
 5. No
 6. Yes, he brought both.
41. Did Megan order (either) calzones or stromboli?
1. She ordered calzones
 2. Yes, she ordered calzones
 3. It was calzones
 4. Yes
 5. No
 6. Yes, she ordered both.
42. Did Renee clean (either) the kitchen or the bathroom?
1. She cleaned the bathroom
 2. Yes, she cleaned the bathroom
 3. It was the bathroom
 4. Yes
 5. No
 6. Yes, she cleaned them both.
43. Did Suzanne want to go to (either) the movies or the bars on Friday night?
1. She wants to go to the movies
 2. Yes, she wants to go to the movies
 3. It's the movies
 4. Yes
 5. No
 6. Yes, she wants to go to both of them.
44. Is Pamela going to knit (either) a blanket or an outfit for the baby shower?
1. She's knitting a blanket
 2. Yes, she's knitting a blanket
 3. It's a blanket
 4. Yes
 5. No
 6. Yes, she's knitting both.
45. Did the professor assign Bill (either) a poster or powerpoint for Tuesday?
1. She assigned a powerpoint.
 2. Yes, she assigned a powerpoint.
 3. It was a powerpoint.
 4. Yes

5. No
 6. Yes, she assigned both.
46. Did (either) Alyssa or Natalie take organic chemistry yet?
1. Natalie did.
 2. Yes, Natalie did.
 3. It was Natalie.
 4. Yes
 5. No
 6. Yes, they both did.
47. Were (either) hamburgers or hot dogs served at the orientation picnic?
1. Hamburgers were served
 2. Yes, hamburgers were served.
 3. It was hamburgers
 4. Yes
 5. No
 6. Yes, they both were served.
48. Do (either) pandas or kangaroos live in Australia?
1. Kangaroos do.
 2. Yes, kangaroos do.
 3. It's kangaroos.
 4. Yes
 5. No
 6. Yes, they both do.

Fillers

Infelicitous (12)

- (49) 1. Are you having a boy or a girl?
*A. A dragon
- (50) 2. Do you want ketchup and mustard on your hot dog?
*A. I hate ketchup
- (51) 3. Would you prefer an aisle or a window seat?
*A. Chicken
- (52) 4. How old is your turtle?
*A. Yes, Rutherford B. Hayes
- (53) 5. Was Pat going to wash the dishes and mop the floor?
*A. My mom used to have a cat named Pat.
- (54) 6. Did Eddie spend all night watching movies and playing video games?
*A. Tony did.
- (55) 7. Is Bruce buying us a map or drawing us a picture?
*A. Bruce is left-handed.
- (56) 8. Does Roger plan to mow the grass and take out the recycling?
*A. I always recycle.
- (57) 9. Did the teacher ask Charlie to come early and stay late?
*A. I am not Charlie.
- (58) 10. Do you have a pencil and pad of paper?
*A. Yes, I am literate.

(59) 11. Are all penguins found in the Southern hemisphere?
*A. Penguins can't fly.

(60) 12. Do you have a dollar and twenty-five cents?
*A. I have a nose and twenty-five freckles.

Felicitous (4)

Good

(61) 1. Did Amber fly to Europe and visit her grandparents?
A. Yes, but her grandparents didn't know she was coming.

(62) 2. Did Bob and Sue get married?
A. Yes, but not to each other.

(63) 3. Does Paula sing and dance?
A. Depends on how you define dancing.

OK

(64) 4. Is Petra going to file her taxes and pay the phone bill?
A. Petra is broke.

Appendix B: Perception Experiment Stimuli Transcription (all stimuli were heard aurally by participants. Each stimulus had four different question versions: Alt tune and Y/N tune without *either*, and Alt tune and Y/N tune with *either*)

1. Did (either) Melanie or Emily go to the party tonight?
 1. Melanie did (fall)
 2. Melanie did (rise)
 3. Yes, Melanie did.
 4. It was Melanie.
2. Will (either) Isabel or Andrea go to the baseball game today?
 - 1: Andrea will go. (fall)
 - 2: Andrea will go. (rise)
 - 2: Yes, Andrea will go
 3. It will be Andrea
3. Did you invite (either) Annabelle or Adrian out to the movies?
 1. I invited Adrian. (fall)
 2. I invited Adrian. (rise)
 3. Yes, I invited Adrian.
 4. It was Adrian.
4. Are (either) podiatrists or ophthalmologists prejudiced against nurses?
 1. Podiatrists are. (fall)
 2. Podiatrists are. (rise)
 3. Yes, podiatrists are.
 4. It's podiatrists.
5. Did Lauren see (either) Angelo or Julia perform at the tap dancing recital?
 1. She saw Julia. (fall)
 2. She saw Julia. (rise)
 3. Yes, she saw Julia.
 4. It was Julia.
6. Do (either) botanists or pharmacists attend the college of Arts and Sciences?
 1. Botanists do. (fall)
 2. Botanists do. (rise)
 3. Yes, botanists do.
 4. It's botanists.
7. Did (either) Dario or Gabriel get a puppy this summer?
 1. Gabriel did. (fall)
 2. Gabriel did. (rise)
 3. Yes, Gabriel did.
 4. It was Gabriel.
8. Did (either) journalists or scientists write the article?
 1. Scientists wrote it. (fall)
 2. Scientists wrote it. (rise)
 3. Yes, scientists wrote it.
 4. It was scientists.
9. Did (either) Chelsey meet with urologists or neurologists during her hospital visit?
 1. She met with neurologists. (fall)

2. She met with neurologists. (rise)
 3. Yes, she met with neurologists.
 4. It was neurologists.
10. Did (either) the new school principal meet with Mario or Darius before class?
1. He met with Mario.
 2. He met with Mario.
 3. Yes, he met with Mario.
 4. It was Mario.
11. Did (either) Solomon or Sullivan vandalize the old warehouse last night?
1. Sullivan did. (fall)
 2. Sullivan did. (rise)
 2. Yes, Sullivan did
 3. It was Sullivan.
12. Does the senator like (either) lobbyists or activists knocking on his door?
1. He likes lobbyists (fall)
 2. He likes lobbyists (rise)
 3. Yes, he likes lobbyists
 4. It's lobbyists.
13. Is Leah listening to (either) philosophers or psychologists at the conference this weekend?
1. She's listening to psychologists. (fall)
 2. She's listening to psychologists. (rise)
 3. Yes, she's listening to psychologists.
 4. It's psychologists.
14. Is Martha videotaping (either) Lillian or Dorian in the school play?
1. She's videotaping Lillian (fall)
 2. She's videotaping Lillian (rise)
 3. Yes, she's videotaping Lillian.
 4. It's Lillian
15. Did (either) Cameron or Jillian feed the dog yesterday?
1. Cameron did (fall)
 2. Cameron did (rise)
 3. Yes, Cameron did
 4. It was Cameron
16. Is Alex inviting (either) Everett or Ellery to spend the night?
1. He's inviting Everett (fall)
 2. He's inviting Everett (rise)
 3. Yes, he's inviting Everett
 4. It's Everett
17. Did (either) Elliot or Emery see the baby elephant at the zoo?
1. Elliot did. (fall)
 2. Elliot did. (rise)
 3. Yes, Elliot did.
 4. It was Elliot.
18. Did (either) Madeline or Mallory enjoy going to the opera?
1. Mallory did. (fall)

2. Mallory did. (rise)
 3. Yes, Mallory did
 4. It was Mallory
19. Did (either) Gregory or Romeo bring dessert?
1. Gregory did (fall)
 2. Gregory did (rise)
 3. Yes, Gregory did
 4. It was Gregory
20. Did Leslie overhear (either) Jeremy or Oliver ordering pizza?
1. She overheard Jeremy (fall)
 2. She overheard Jeremy (rise)
 3. Yes, she overheard Jeremy
 4. It was Jeremy
21. Did Amanda see (either) Ariel or Miriam at a city council meeting?
1. She saw Miriam (fall)
 2. She saw Miriam (rise)
 3. Yes, she saw Miriam
 4. It was Miriam
22. Do you see (either) Damian or Beverly on the beach?
1. I see Beverly (fall)
 2. I see Beverly (rise)
 3. Yes, I see Beverly
 4. It's Beverly
23. Did (either) Julian or Pamela fly to Florida for spring break?
1. Julian did (fall)
 2. Julian did (rise)
 3. Yes, Julian did
 4. It was Julian
24. Did Martin go with (either) Allison or Julius to the amusement park today?
1. He went with Allison (fall)
 2. He went with Allison (rise)
 3. Yes, he went with Allison
 4. It was Allison
25. Did (either) Angela or Julio stop by to drop off their keys?
1. Julio stopped by (fall)
 2. Julio stopped by (rise)
 3. Yes, Julio stopped by
 4. It was Julio
26. Did Dan go with (either) Emerson or Benjamin to the new restaurant in town?
1. He went with Benjamin (fall)
 2. He went with Benjamin (rise)
 3. Yes, he went with Benjamin
 4. It was Benjamin
27. Is Mary allergic to (either) dairy or soy?
1. She's allergic to dairy. (fall)

2. She's allergic to dairy. (rise)
 3. Yes, she's allergic to dairy
 4. It's dairy
28. Was Samantha going to feed (either) the turtle or the lizard?
1. She was going to feed the turtle
 2. She was going to feed the turtle
 3. Yes, she was going to feed the turtle
 4. It was the turtle
29. Did William send (either) an email or a letter?
1. He sent a letter (fall)
 2. He sent a letter (rise)
 3. Yes, he sent a letter
 4. It was a letter.
30. Is Frank finishing (either) homework or the online quiz?
1. He's finishing homework. (fall)
 2. He's finishing homework. (rise)
 3. Yes, he's finishing homework
 4. It's homework.
31. Did Alan buy her (either) chocolates or flowers?
1. He bought her flowers (fall)
 2. He bought her flowers (rise)
 3. Yes, he bought her flowers
 4. It was flowers
32. Does Amy use (either) Hulu or Netflix?
1. She uses Netflix (fall)
 2. She uses Netflix (rise)
 3. Yes, she uses Netflix
 4. It's Netflix
33. Do the kids need (either) a snack or a bathroom break?
1. They need a snack (fall)
 2. They need a snack (rise)
 3. Yes, they need a snack
 4. It's a snack
34. Did you order (either) mocha or French vanilla?
1. I ordered mocha (fall)
 2. I ordered mocha (rise)
 3. Yes, I ordered mocha
 4. It was mocha
35. Does Dave plan to take (either) sick days or vacation days?
1. He plans to take vacation days.
 2. He plans to take vacation days.
 3. Yes, he plans to take vacation days.
 4. It's vacation days
36. Would Jordi like to visit (either) the zoo or the aquarium?
1. He wants to visit the zoo

2. He wants to visit the zoo
 3. Yes, he wants to visit the zoo
 4. It's the zoo
37. Does Heather play (either) soccer or softball?
1. She plays soccer
 2. She plays soccer
 3. Yes, she plays soccer
 4. It's soccer.
38. Did Mike wear (either) a hat or sunglasses?
1. He wore sunglasses
 2. He wore sunglasses
 3. Yes, he wore sunglasses
 4. It was sunglasses
39. Would you like (either) Ranch dressing or Thousand Island?
1. I would like Ranch
 2. I would like Ranch
 3. Yes, I would like Ranch
 4. It's Ranch
40. Did Travis bring (either) drinks or dessert?
1. He brought dessert
 2. He brought dessert
 3. Yes, he brought dessert
 4. It was dessert
41. Did Megan order (either) calzones or stromboli?
1. She ordered calzones
 2. She ordered calzones
 3. Yes, she ordered calzones
 4. It was calzones
42. Did Renee clean (either) the kitchen or the bathroom?
1. She cleaned the bathroom
 2. She cleaned the bathroom
 3. Yes, she cleaned the bathroom
 4. It was the bathroom
43. Did Suzanne want to go to (either) the movies or the bars on Friday night?
1. She wants to go to the movies
 2. She wants to go to the movies
 3. Yes, she wants to go to the movies
 4. It's the movies
44. Is Pamela going to knit (either) a blanket or an outfit for the baby shower?
1. She knitting a blanket
 2. She knitting a blanket
 3. Yes, she's knitting a blanket
 4. It's a blanket
45. Did the professor assign Bill (either) a poster or powerpoint for Tuesday?
1. She asked him to make a powerpoint.

2. She asked him to make a powerpoint.
 3. Yes, she asked him to make a powerpoint.
 4. It was a powerpoint.
46. Did (either) Alyssa or Natalie take organic chemistry yet?
1. Natalie did.
 2. Natalie did.
 3. Yes, Natalie did.
 4. It was Natalie.
47. Were (either) hamburgers or hot dogs served at the orientation picnic?
1. Hamburgers were served.
 2. Hamburgers were served.
 3. Yes, hamburgers were served.
 4. It was hamburgers
48. Do (either) pandas or kangaroos live in Australia?
1. Kangaroos do.
 2. Kangaroos do.
 3. Yes, kangaroos do.
 4. It's kangaroos.

Fillers (infelicitous)

Capitalization indicates accented/contrastive prosody

- (49)1. Did Jill bring the BLUE ball?
A: Jill brought the BLUE jacket.
- (50)2. Is Sam wearing a red tie tonight?
A: He is wearing a RED shirt tonight.
- (51)3. Is your Mom baking an apple pie?
A: She is baking a APPLE cake.
- (52)4. What color dress is your daughter wearing to the dance this Saturday?
A: She is wearing a green DRESS this Saturday.
- (53)5. Does Scott have two laptops?
A: He has TWO cell phones.
- (54)6. Does Astrid have on a long dress?
A: She has on a LONG skirt.
- (55)7. Is Jake going to Veronica's party tonight?
A: He is going to Jerry's PARTY tonight.
- (56)8. Does Roger have a dentist appointment tomorrow?
A: He has a DENTIST appointment today.
- (57)9. Does Hank love hamburgers?
A: Hank LOVES hot dogs.
- (58)10. How many pairs of shoes does Felicia have?
A: Felicia has ten PAIRS of shoes.
- (59)11. Does Nicole have four tattoos?
A: She has FOUR piercings.
- (60)12. Is Ray bringing six wine glasses?
A: He is bringing SIX water bottles.

Fillers (felicitous)

- (61) 1. Did Joe come to the football game?
1. No, JANE came to the football game.
- (62) 2. Did Dad go to the 5 o'clock movie showing today?
Condition: He went to the 1 o'clock showing today.
- (63) 3. How many MARBLES did Sally lose yesterday?
Condition: She lost TWO marbles.
- (64) 4. What is Mark's favorite COLOR?
Condition: It's BLUE.
- (65) 5. Does Anthony like pancakes?
Condition: He LOVES pancakes.
- (66) 6. What is Billy's favorite TV show?
Condition: It's GAME OF THRONES.
- (67) 7. Do you walk your dog TWO times a day?
Condition: I walk her THREE times a day.
- (68) 8. Did Isaac wear his glasses today?
Condition: No, he wore his CONTACTS today.
- (69) 9. Does Adam like to write with BLUE pens?
Condition: No, he likes BLACK pens.
- (70) 10. Is Drew throwing a BASEBALL with Rory?
Condition: They are throwing a FOOTBALL together.
- (71) 11. Is Phillip going to IRELAND for vacation?
Condition: No, he is going to SCOTLAND for his vacation.
- (72) 12. Does Amelia have TWO parties to go to this weekend?
Condition: She has FOUR parties to go to!