Prosodic disambiguation of scopally ambiguous sentences*

Kristen Syrett[a], Georgia Simon[a] & Kirsten Nisula[b]

[a]Rutgers University – New Brunswick, [b]The Ohio State University

1. Introduction and background

Quantificational sentences such as the one in (1) are ambiguous. Under one interpretation, all of the men are such that they did not go (that is, none of the men went). Under another, it is not the case that all the men went; only some (and perhaps none) did.

(1) <u>All</u> the men did<u>n't</u> go.

These competing interpretations arise from the interaction of the universal quantifier in subject position (*all*) and the negation later in the sentence. These interpretations are formally captured in the logic provided in (2), with the 'all>negation' interpretation in (a) and the 'negation>all' interpretation in (b).

(2) a. $\forall x.\ \text{man}(x) \rightarrow \neg \text{go}$

b. $\neg \forall x.\ \text{man}(x) \rightarrow \text{go}$

These two interpretations may be teased apart, of course, given a discourse context in which it is clear that none or some of the men went. However, it has also been claimed that the way in which these sentences (and other scopally ambiguous sentences) are uttered also serves to tip the speaker's hand and indicate which of the two interpretations was intended. Specifically, (1) when uttered with a sentence-final falling intonational contour is said to favor interpretation (2a), while (1) when uttered with a rising (or fall-rise) contour is said to favor interpretation (2b).

Jackendoff (1972) argued that this pattern arises from how negation is encoded in the logical representation – whether it is associated with the presupposition and not the

* This work was supported by a Rutgers University startup grant to the first author and a small grant from the Aresty Research Center at Rutgers University. We are grateful for technical assistance from the Rutgers Phonology and Field Research Laboratory and discussions with Shigeto Kawahara.

assertion, as in (3), or with the focus, as in (4).

(3) A accent (falling contour): negation is part of the presupposition, not the focus
Presupposition: λQ [Q of the men did**n't go**] is well-formed/under discussion
Assertion: all $\in \lambda Q$ [Q of the men didn't go]

(4) B accent (rising contour): negation is associated with the focus (i.e., the assertion), not presupposition
Presupposition: λQ [Q of the men **went**] is well-formed/under discussion
Assertion: all $\notin \lambda Q$ [Q of the men went]

This connection was subsequently recast under a pragmatic approach by Ladd (1980) and Kadmon & Roberts (1986), and the sources of the connection between the contour and interpretation questioned by and Ward & Hirschberg (1985). To date, we lack empirical evidence to corroborate Jackendoff's original claim, and determine whether such sentences are produced in a reliable manner, and if so, whether the sentence-final contour is what houses this distinction. Moreover, the few studies that have investigated this phenomenon in English, which have come up largely empty-handed, have not incorporated into their experiments pragmatic apsects of the discourse that may affect interpretation and prosody, such as the question under discussion, the role of negation in presupposition/assertion relation, and the presence and type of scalar alternatives. This research aims to fill this gap. We show that speakers can signal sentence interpretation with auditory cues, although they are not always in the contour, and what cues there are may be unpredictable, since they are also tied to the information structure.

2. Experiment

2.1 Participants

19 undergraduates (all native speakers of American English) receiving course credit participated. Six additional non-native participants were excluded, and one additional speaker's sound files were damaged and excluded from analysis.

2.2 Experimental design and procedure

The aim of this experiment was to elicit productions of ambiguous sentences in a discourse context with the purpose of identifying auditory correlates of sentence interpretation. Test items were scopally ambiguous sentences involving a quantifier (either the universal quantifier *all* in subject position, or *many* or *most* in object position) and negation. Each test sentence type was embedded at the end of multiple discourse contexts varying the information structure and scopal relation. There were four contexts for *all* and three for *many* and *most*, for a total of 16 *all* and 12 *many*/*most* sentences. These 28 test items were then pseudorandomized with 28 control items, for a total of 56 items in the test session. In constructing the sentences, we controlled for sonorance, particularly at the end of the sentence, in order to elicit a smooth pitch track.

Control items were ambiguous sentences that have been shown in previous research to be able to be disambiguated through prosody, among which were five pairs of sentences with negation and a *because* clause, five pairs containing a focus-sensitive operator (three with *only*, two with *even*), and four pairs with pronominal reference. Each control sentence was presented in two different contexts, each favoring one of the competing interpretations. An example of each item type is included in (5)-(11).

(5) All the magnolias won't bloom. (*all*, negation)

(6) Liam doesn't know many alumni. (*many*, negation)

(7) Neil doesn't enjoy most musicals. (*most*, negation)

(8) Georgia isn't singing because she's preparing for an audition. (*because*, negation)

(9) Warren only likes the Orioles. (focus sensitive operator: *only*)

(10) She even painted the garage. (focus sensitive operator: *even*)

(11) Alan punched Owen and then he kicked him. (pronominal reference)

For all items (with the exception of those with pronominal reference), the target sentence was followed by an additional sentence. The reason for this was that sentences that seemed most naturally produced with a rising intonation ended abruptly without such a continuation.

We note here a key contrast with the two types of test items. There is a quantifiable difference between the two interpretations of the sentences containing the universal quantifier *all* and negation. When *all* takes wide scope over negation, negation is at the VP level, and none of the discourse entities mentioned in the sentence have the property (e.g., none of the magnolias bloom). When negation takes wide scope over the quantifier, negation is propositional, and the quantity of discourse entities mentioned in the sentence that have the property is 'not all'. With the *many/most* sentences, the situation is different. No matter what the scopal relation is (whether negation takes wide or narrow scope), the quantity does not vary: what varies is the *focus* on the quantity. For example, in (6), the number of alumni that Liam knows is always small. When *many* takes wide scope over negation, emphasis is placed is on the number of alumni that Liam *does not know* (*many*). However, when negation takes wide scope over *many*, emphasis is placed on how many alumni he *does* know, which is *few*. (See Baltazani (2002, 2003).)

For the test items, two of the contexts varied the scopal relation between the quantifier and negation (quantifier > negation, negation > quantifier). In one context, negation was associated with the presupposition and the Question Under Discussion (QUD) was negative, while in a second context, the QUD was positive, and negation was associated with the assertion or focus. In addition to the first two contexts, we presented these test items in additional contexts, allowing us to further evaluate how prosody can vary across discourse contexts, even when a scopal relation remains constant. For *many* and *most*, we created a third context (modeled after an example discussed by Kadmon & Roberts (1986)) in which we favored an interpretation where negation would take wide scope over the quantifier, but in which we predicted we might elicit a falling contour, since the target sentence appeared as an embedded clause.

For the universal quantifier items, we created two additional contexts that allowed us to achieve a fully crossed design for these items (negation in presupposition or assertion x prosodic contour), also varying the type of scalar alternatives (quantity or object-based). In the third context, we created a salient scalar alternative to the DP (e.g., magnolias), thereby inducing a fall-rise on this lexical item (cf. Ward & Hirschberg, 1985). In the fourth context, we manipulated the information structure to favor negation taking wide scope over the universal quantifier, as in the second context. However, here, we attempted to create a scenario that favored a falling contour. In (12), we present four representative contexts for the test sentence *All the magnolias won't bloom*.

(12) a. Context 1:
The township decided to plant magnolia saplings a number of years ago to line a path through the park. They have experienced lovely blossoms every year. However, this year the area is experiencing less-than-standard rainfall, which means that they expect the magnolias to struggle this year, with only a few surviving. In fact, I think the situation is much more dire than that. *All the magnolias won't bloom.* They'll just have to wait till next year.
Presupposition: Some magnolias may not bloom. (QUD contains negation.)
Assertion: None of them will.
Scopal relation favored: all > negation
Contour predicted: falling
Scalar alternatives: Quantity (How many trees)

b. Context 2:
A few years ago, the township decided to plant magnolia saplings to line a path through the park. The saplings on the north side were planted mainly in sand, and haven't been getting nearly enough nutrients. However, the soil near the south side is rich, and the magnolias are thriving there. *All the magnolias won't bloom*. But I bet the ones on the south side will.
Presupposition: All magnolias will bloom. (QUD does not contain negation.)
Assertion: Some will.
Scopal relation favored: negation > all
Contour predicted: fall-rise
Scalar alternatives: Object-based (which trees)

c. Context 3:
An aggressive beetle that targets magnolia trees is spreading through our area, and the magnolias are doomed. The township has been planning to take pictures for their website next month. The official photographer is concerned that there won't be beautiful rows of trees in the background for his pictures. I think he's worrying too much. *All the magnolias won't bloom*. However, there will still be other trees that will look just as lovely.
Presupposition: Some of the trees will not bloom. (QUD contains negation.)
Assertion: All of the *magnolias* will not bloom; other trees will.
Scopal relation favored: all > negation
Contour predicted: fall-rise
Scalar alternatives: Object-based (which trees)

d. Context 4:
   The weather recently has been conducive to plant growth, and all the trees are looking healthy. Some optimistic members of the township are predicting that each of the magnolia trees will give us lovely, fragrant blossoms to enjoy all season. But I think they're being rather unrealistic, and I keep telling them this. *All the magnolias won't bloom*. The odds of each of them blooming are pretty slim.
   Presupposition: All magnolias will bloom. (QUD does not contain negation.)
   Assertion: Not all of the magnolias will bloom.
   Scopal relation favored: negation > all
   Contour predicted: falling
   Scalar alternatives: Quantity (How many trees)

Participants were recorded one at a time using an AT4040 Cardioid Capacitor microphone with a pop filter in a sound-attenuated recording booth, and amplified through an ART Digital MPA Gold microphone pre-amplifier. Stimuli were presented to participants using SuperLab stimulus presentation software (Cedrus Corporation, 2012) on a Macbook. For each trial, the participant first read the entire paragraph silently. They then answered a comprehension question that tested for their understanding of the target sentence in the context. Finally, they read the entire paragraph out loud, this time recorded. Items for which participants did not answer the comprehension question correctly were not included in the analysis, since we wanted to be sure of interpretation.

Participants were told to read the items as naturally as possible. Stimuli were divided into two blocks, so that members of a minimal pair were separated between blocks; each block contained a token from each test or control item type. Test and control items were pseudorandomized within each block. Participants were run in two conditions. In the first condition, participants completed all 56 items within one session, which lasted approximately 45-60 minutes. They were told to read the items as naturally as possible. The second condition was constructed to alleviate some of the burden of the task. Participants were presented with only one of the two blocks, and were again encouraged to read the items as naturally as possible, as though they were recording them for an audiobook or reading to children. They were also provided with an example of a written discourse beforehand (a section from a soap opera transcript) to model expressive reading.

2.3 Analysis

Target sentences were excised from the surrounding context using Praat speech analysis software (Boersma & Weenink, 2011). They were then annotated with segments delineated from the onset and offset of each lexical item. For the control items containing a *because* clause and negation or a focus sensitive operator, we collected the value of the maximum F0 on the relevant lexical items. For the *because* sentences, this was the verb following negation and the final word in the sentence (e.g., *singing* and *audition* in the sentence *Georgia isn't singing because she's preparing for an audition*). For the focus sensitive operators, this was the head of the VP and DP that could host the focus (e.g., *painted* and *garage* in the sentence *She only painted the garage*).

For the test items, we conducted two main analyses. First, each file was coded for the type of sentence-final contour that was observed. To do this, two experimenters (the second and third authors) independently listened to each file (with all identifiers of context removed) and blindly coded it as either a falling or non-falling (fall-rise) contour. The initial rate of agreement was 80%. Any and all discrepancies were easily reconciled blindly afterwards with the assistance of a third coder (the first author), using as a comparison other clear exemplars from the participant whose items were in question.

We then conducted an analysis of the test items in order to look for any acoustic signatures in the speech signal that accompanied a difference in interpretation. Scripts in Praat (Boersma & Weenink, 1999-2011) were written and run on the annotated files to extract the relevant acoustic information. We excluded from analysis tokens that had a comprehension score of 0, as well as tokens that had mis-starts, errors in pronunciation, use of the partitive in the quantificational phrase, glottalization in key lexical items, and/or errors in subject-verb agreement. Participants for whom there were not enough data points to perform an acoustic analysis across items after this filtering process were excluded from analysis (n=4). In addition, one set of the *all* sentences and one set of the *many/most* sentences were excluded from analysis, since the comprehension scores for one or all of the items in each set were consistently at or below chance level across speakers, and speakers either reported difficulty in accessing the correct interpretations for these items, or experienced difficulty producing the items while being recorded. This conservative filtering process still left us with a rather large sample to work with across and test and control items and speakers, given our design.

2.4 Results

Pronominal reference: Participants did not mark pronominal reference with pitch accents. In the scenarios favoring a default, unstressed pronoun, participants produced the target pronoun in this manner 96.3% of the time. However, in the context supporting a stressed pronoun, with reversed reference, participants also produced the target pronoun as unstressed 95.6% of the time. This may be in part due to the fact that participants had much lower comprehension scores for the 'reversed' cases than for the 'default' cases (an average of 56.3% v. 90.6%), however, even for the most successful 'reversed pronominal reference' item in which comprehension scores were quite high (87.5%), participants were more likely than not to produce the pronoun as unstressed.

***Focus-sensitive operators* only *and* even**: We first recorded whether the main accent was on the head verb of the VP or on the DP for each type of target sentence. We predicted that this pattern would be correlated with the focus pattern supported by the discourse context. The results are presented in (13).

(13) *Table 1: Pitch accenting patterns for the focus sensitive items* only *and* even

	focus	verb	DO
even	VP	6	24
	DO	6	24
only	VP	25	22
	DO	5	37

Participants preferred to place the nuclear pitch accent on a syllable in the direct object for both sets of items. For *even*, there was absolutely no difference in accenting patterns for the individual items (p=1), regardless of the preceding context and the favored interpretation. This pattern held, despite the fact that participants' comprehension scores were the same for both context: 90.6% in each. For *only*, the pattern was different, and participants' accenting pattern varied with the interpretation favored by the preceding context. As with *even*, a syllable in the direct object was more likely to be accented when the previous context favored a scalar contrast of the direct object (e.g., *Warren only likes the ORIOLES*), However, when the preceding context favored a scalar contrast of the verb (e.g., *Warren only LIKES the Orioles*), participants were pulled away from the default pattern (Pearson χ^2 (1) = 16.92, p<.0001, ϕ = -.44), but were no more likely to place the accent on the verb than on the direct object (binomial probability p=.77). As before, participants' comprehension scores were quite high: 81.3% for contexts favoring a verb contrast, and 100% for contexts favoring a contrast of the direct object.

We then calculated the maximum F0 within the target lexical item in the sentence, then calculated its location within the duration of that item, and conducted a 2 x 2 ANOVA with word (V, DO) and context (focus on VP or focus on DO) as within-subject factors on items. Values greater than 1.5 SD away from the mean were excluded. When the focus was on the VP, the ratio of the Max F0 location in the V was .30, and .36 in the DO. When the focus was on the DO, the ratio of the Max FO location in the V was .26, but .47 in the DO (the largest value observed). Accordingly, there was a significant main effect of word (F(1, 181) = 12.2, MSE = .78, p=.0006), no main effect of context (F(1, 181) = .16, MSE = .01, p=.70), and a significant interaction (F(1, 181) = 4.38, MSE = .06, p=.038). Post-hoc Tukey's HSD tests revealed that the location of Max F0 was larger (at the .01 level of significance) in the DO in the context in which this focus pattern was predicted.

Because-*negation sentences*: For these sentences, we calculated the maximum F0 in two lexical items: the negated verb or predicate preceding the word *because* and the final word in the sentence (e.g., *They're not <u>late</u> because of his <u>driving</u>*). We then compared these two values in both contexts. As before, values greater than 1.5 SD away from the mean were excluded. For sentences where *because* took scope over negation, the value of F0 was 229.0 Hz for word 1, and 193.4 Hz for word 2. For sentences where negation took wide scope, the F0 for word 1 was 213.9 Hz and 206.6 Hz for word 2. A 2 x 2 (word x context) ANOVA revealed a main effect of word (F(1, 170) = 8.98 , MSE = 19403.79, p=.003), no effect of context (F(1, 170) = 0, MSE = 2.32, p=1), but a significant interaction (F(1, 170) = 4.03, MSE = 8709.38, p=.046). Post-hoc Tukey's HSD tests revealed that in the *because*>negation contexts, there was a significantly higher maximum F0 on word 1 than on word 2 (at the .05 level of significance), and that the F0 of word 1 was significantly higher in the *because*>negation context than in the negation>*because* context (.05 significance).

We then compared the location of the maximum F0 in the key words in both contexts, conducting two-tailed independent t tests assuming unequal variance. The ratio of the F0 location was .40 for word 1 in both contexts. However, the location of Max F0 for word 2 varied: it was .34 in the *because*>negation context, but .45 (more delayed) in the negation>*because* context. Accordingly, there was no significant difference for word 1 for the two contexts (t(82)=-.13, p=.90), but there was a significant difference in the Max F0 placement between the two contexts on word 2 (t(73)=-2.51, p=.01). It has also been claimed that these contexts are more likely to exhibit a rising (or fall-rise) contour (cf. Koizumi, 2009).

Quantifier-*negation sentences*: The quantificational test sentences underwent two types of analysis. First, the excised sentences were coded for falling/non-falling contour. The percentage of observed falling contour for each context is presented in Table 2 (14).

(14) *Table 2: Scopal relation favored, scalar alternative, and % observed falling contour for target sentences in each context*

quantifier	context	scopal relation	scalar alternative	% falling contour
all	1	*all* > negation	quantity	93.4
	2	negation > *all*	object-based	89.1
	3	*all* > negation	object-based	71.1
	4	negation > *all*	quantity	95.5
many/most	1	*M* > negation	quantity	91.3
	2	negation > *M*	object-based	65.1
	3	negation > *M*	object-based	63.0

A binomial probability analysis for the *all* sentences revealed that the frequency of observed falling contour is greater than chance for all contexts (contexts 1, 2, 4: $p<.0001$; context 3: $p<.01$). For the *many/most* sentences, the frequency of observed falling contour is greater than chance for context 1 ($p<.0001$), marginally significant for context 2 at $p=.07$, and not significant for context 3 ($p=.10$). Thus, the sentence-final contour categorized as either falling or not falling is not at all a good indicator of scope for the *all*-negation sentences. However, in spite of the overall trend for a falling contour in the *many*-negation sentences, the difference in the sentence-final contour is correlated with scopal relation: when *many* or *most* scopes over negation, a falling contour is much more likely than when negation takes wide scope. Thus, we do observe an intonational difference between the types of quantificational sentences in the direction predicted.

We then conducted a second, acoustic analysis of the test sentences. For each of the sentences, we focused on two key lexical items: the quantifier (*all*, *many*, or *most*) and the sentence-final word. For the two types of test items, we conducted a two-way ANOVA on the items, targeting the quantifier and the final word, comparing the factors of context and acoustic measure: (a) maximum pitch (F0) in the word, (b) the F0 standard deviation within the word, and (d) the duration of the word. We take each of these analyses in turn. The results for sentences containing *all* and negation are presented in Table 3 (15). Significant differences are highlighted with dark border; marginally significant differences indicated with a dashed line.

Turning first to the analysis *all*, we found no main effects when directing our attention to the quantifier itself: max F0: $F(3, 66) = .34$, $p=.80$; F0 standard deviation: $F(3,66) = .15$, $p=.93$; word duration: $F(3, 66) = .19$, $p=.90$. The sentence-final word also exhibited a similar pattern of null results for two of the acoustic measures: max F0: $F(3, 66) = .87$, $p=.46$; F0 standard deviation: $F(3,66) = 1.27$, $p=.29$. However, there was a significant main effect of word duration: $F(3, 66) = 6.20$, $p=.001$. Planned pairwise comparisons between each of the contexts within each of the acoustic measures revealed that the significant main effect of final word duration was driven by a highly significant difference between contexts 3 and 4 ($p=<.0001$), and a significant difference between contexts 1 and 3 and 2 and 3 (both *p*=.01). There was a marginally significant difference in the final word duration between contexts 1 and 4 and between contexts 2 and 4 (both *p*=.08), and no significant difference in the final word duration between contexts 1 and 2 (*p*=1.0). Thus, duration of the final word was longest in context 3, where the scopal relation favored was *all*>negation, but there was a contrast in object-based scalar alternatives (e.g., magnolias v. other trees), predicting a fall-rise contour would be exhibited. The duration was shortest in context 4, where the scopal relation favored was negation>*all*, and the predicted contour was falling.

(15) *Table 3: Values of three acoustic measures in four discourse contexts for the quantifier* all *and the sentence-final word*

quantifier: *all*				sentence-final word			
Acoustic measure	Context	Mean	Std. Error	Acoustic measure	Context	Mean	Std. Error
maximum F0 (Hz)	1 ∀>¬	230.90	15.13	maximum F0 (Hz)	1	192.56	85.67
	2 ¬>∀	245.23	15.13		2	162.84	51.40
	3 ∀>¬	247.30	15.13		3	181.73	39.85
	4 ¬>∀	230.78	16.11		4	190.17	47.39
F0 standard deviation (Hz)	1 ∀>¬	26.06	3.94	F0 standard deviation (Hz)	1	29.45	17.07
	2 ¬>∀	28.73	3.94		2	21.01	14.36
	3 ∀>¬	25.83	3.94		3	28.43	13.13
	4 ¬>∀	28.40	4.19		4	30.43	17.25
word duration (ms)	1 ∀>¬	170	10	word duration (ms)	1	391	65
	2 ¬>∀	160	10		2	391	65
	3 ∀>¬	170	10		3	448	59
	4 ¬>∀	168	11		4	349	75

We turn next to the results for the sentences containing *many/most* and negation (Table 4). Note that we collapse over *many* and *most*, but are never comparing values for the two quantifiers. The ANOVA run on the quantifier uncovered no main effects for two of the measures: maximum F0: $F(2, 44) = .20$, $p=.82$; F0 standard deviation: $F(2,44) = .61$, $p=.55$. However, there was a significant main effect of word duration: $F(2, 44) = 3.68$, $p=.03$. Planned pairwise comparisons revealed significant differences among the contexts for quantifier duration. While contexts 2 and 3 were not significantly different from each other ($p=.52$), there was a significant difference between contexts 1 and 2 ($p=.01$) and a marginally significant difference between contexts 1 and 3 ($p=.06$). The ANOVA on the final word revealed no main effect for the first two acoustic measures: maximum F0: $F(2, 44) = .86$, $p=.43$; F0 standard deviation: $F(2,44) = .30$, $p=.74$. However, as before, there was a significant main effect of word duration: $F(2, 44) = 4.03$, $p=.03$. Planned pairwise comparisons between each of the contexts within each of the acoustic measures revealed significant differences between the three contexts for final word duration: context 1 v. 2 p=.009, 1 v. 3 p=.05, but contexts 2 v. 3 p=.45. Combining these two analyses for *many/most*, we find that the length of the quantifier was shortest in context 1, with the duration in contexts 2 and 3 not differing from each other. At the same time, the duration of the final word was longest in context 1, with the duration in contexts 2 and 3 not differing from each other.

(16) *Table 4: Values of three acoustic measures observed in three discourse contexts for the quantifiers* many/most *and the sentence-final word*

quantifier: *many* or *most*				sentence-final word			
Acoustic measure	Context	Mean	Std. Error	Acoustic measure	Context	Mean	Std. Error
maximum F0 (Hz)	1 M>¬	186.54	38.64	maximum F0 (Hz)	1	189.97	72.32
	2 ¬>M	195.40	42.32		2	189.15	39.36
	3 ¬>M	187.19	44.78		3	167.32	40.57
F0 standard deviation (Hz)	1 M >¬	10.68	6.21	F0 standard deviation (Hz)	1	24.79	14.70
	2 ¬>M	10.59	7.25		2	25.33	10.88
	3 ¬>M	13.69	11.59		3	21.72	14.68
word duration (ms)	1 M >¬	277	24	word duration (ms)	1	520	51
	2 ¬>M	308	38		2	462	51
	3 ¬>M	300	34		3	478	69

Focusing only on the final word for our test sentences, we observe an interesting comparison between the *all* and the *many/most* cases: the longest final word duration for each set of target sentences was observed in a context in which the quantifier>negation scopal relation was favored (context 3 for *all* and context 1 for *many/most*). The duration was shortest for the *all* sentences in a context favoring a negation>quantifier relation (context 4). These durations do not, however, entirely correlate with the predicted or observed contour, since all four contexts of the *all* sentences exhibited a robust falling contour, but the first context for the *many/most* sentences was the contour that had the highest percentage of falling contour. Thus, auditory correlates to scopal relation and contextual alternatives can be retrieved, but they are not entirely consistent or predictable, and vary among quantifiers.

3. General discussion and conclusions

In this research, we sought to obtain empirical evidence relevant to seminal claims by Jackendoff (1972) that differing interpretations of scopally ambiguous quantificational sentences in English are flagged by distinct and reliable prosodic cues. By incorporating into our experimental design additional information tied to the discourse context, we were also able to evaluate the role of pragmatic factors such as the type of scalar alternatives (cf. Kadmon & Roberts, 1986; Ladd, 1980; Ward & Hirschberg, 1985).

For our control sentences containing the focus-sensitive operators *only* and *even* scoping over a VP or the DO, and for sentences containing a *because* clause scopally interacting with negation, speakers varied their production of such sentences according to the scopal relation and the interpretation favored by the discourse context. Our findings are consistent with findings from previous researchers, including Cooper & Paccia-Cooper (1980), Frazier & Clifton (1996), Koizumi (2009), Hirschberg & Avesani (2007). However, while speakers produced these ambiguous control sentences in a way that could disambiguate them for the hearer, the results were less clear-cut for our scopally ambiguous quantificational test sentences. Contrary to previous claims in the theoretical literature that speakers reliably signal their interpretation with the sentence-final contour, we found no such pattern for sentences with the universal quantifier *all* and negation. And while we found a difference along those lines with the *many/most* sentences, in that a falling contour was far more likely with sentences in which negation took narrow scope, speakers more often than not exhibited a falling contour, and did not robustly signal negation taking wide scope with a fall-rise contour. Instead, the differences appear to occur at a lower level—most often with a difference in word duration. However, even here, such acoustic correlates are hard to pin down as a definite signal to interpretation. We find this difference between the *all* and the *many/most* sentences intriguing. The current work cannot say whether the difference stems from the quantifiers themselves or their syntactic position, or a combination of these aspects. We leave the resolution of this open question to future research.

We also note that it is possible is that an omnibus analysis of the data may mask intra- and interspeaker variability in the production of these sentences. That is, speakers may be recruiting a variety of cues (such as word duration and intonational contour) differentially from utterance to utterance, and/or each speaker may have his/her own strategy for disambiguation. It is not possible to tease apart this possibility without an extensive item-by-item and speaker-by-speaker analysis. For now, we leave it as open possibility that individual speakers are actually better at consistently identifying strategies for signaling interpretation with auditory cues than our current work (averaging over speakers and items) suggests. We present our contribution as showing that sentence-final contour does not appear to be *the* indicator of interpretation, and that there are other cues, such as duration, which may be informative.

In a complementary project (Syrett, Simon, & Nisula, *under revision*), we took a subset of sound files from this production study produced by speakers with clear productions consistent with the theoretical claims, and fed them into two complementary perception experiments. There, we asked if hearers could recruit such auditory cues to assign a sentence interpretation, and whether they could use the information provided by the discourse to select one prosodic rendition or another. Our findings revealed that despite the variability observed in the current production experiment, hearers do associate certain interpretations (scopal relations) with certain prosodic versions of the sentences, in the direction predicted. Taken together, the production and perception experiments indicate that there is a correlation between sound and meaning in the case of scopally ambiguous sentences, and that psycholinguistic studies investigating participants' ability to access multiple interpretations of such sentences should incorporate prosody into the design.

References

Baltazani, Mary. 2002. The prosodic structure of quantificational sentences in Greek. In

Proceedings from the 38th Meeting of the Chicago Linguistic Society, ed. by M. Andronis, E. Debenport, A. Pycha, and K. Yoshimura, 63-78). Chicago, IL: Chicago Linguistic Society.

Baltazani, Mary. 2003. Quantifier scope and the role of intonation in Greek. Doctoral dissertation, UCLA.

Boersma, Paul, & Weenink, David. 1999-2011. Praat: doing phonetics by computer. [Computer program]. Retrieved from http://www.fon.hum.uva.nl/praat/.

Cooper, William E., & Paccia-Cooper, Jeanne. 1980. *Syntax and speech.* Cambridge, MA: Harvard.

Frazier, Lyn, & Clifton, Charles. 1996. *Construal*. Cambridge, MA.: MIT Press.

Hirschberg, Julia, & Avesani, Cinzia. (1997). The role of prosody in disambiguating potentially ambiguous utterances in English and Italian. In *Intonation: Theory, Models, and Applications*, ed. by A. Botinis, G. Kouroupetroglou, and G. Carayiannis, 189-192. European Speech Communication Association and University of Athens Department of Informatics.

Jackendoff, Ray. S. 1972. *Semantic interpretation in generative grammar.* Cambridge, MA: MIT Press.

Kadmon, Nirit, & Roberts, Craige. 1986. Prosody and scope: The role of discourse structure. In *Proceeding from the 22nd Meeting of the Chicago Linguistics Society*, ed. by A. M. Farley, P. T. Farley, and K.-E. McCullough, 16-28. Chicago, IL: Chicago Linguistic Society.

Koizumi, Yukiko. 2009. Processing the not-because ambiguity in English: The role of pragmatics and prosody. Doctoral dissertation, City University of New York.

Ladd, D. Robert. 1980. *Structure of intonational meaning: Evidence from English*. Bloomington, IN: Indiana University Press.

Syrett, Kristen, Simon, Georgia, & Nisula, Kirsten. *under revision*. Prosodic disambiguation of scopally ambiguous sentences in a discourse context. Ms., Rutgers, The State University of New Jersey – New Brunswick.

Ward, Gregory, & Hirschberg, Julia. 1985. Implicating uncertainty: The pragmatics of Fall-Rise intonation. *Language* 61: 747-776.

Corresponding author:
Kristen Syrett
Department of Linguistics
Rutgers, The State University of New Jersey – New Brunswick
18 Seminary Place
New Brunswick, NJ 08901

k-syrett@ruccs.rutgers.edu