Question Tags and Sentential Negativity

Adrian Brasoveanu   Karen De Clercq   Donka Farkas
UC Santa Cruz   Ghent University   UC Santa Cruz

Floris Roelofsen
ILLC Amsterdam

March 20, 2014

Abstract

This paper presents an experiment that is designed to quantify the negativity of sentences with different types of negative operators (n-words like *never* and downward entailing operators like *rarely*) in different syntactic positions (adverb, subject, and direct object). In the experiment, participants were provided with a minimal context, then asked to choose one tag-question out of two; one of questions had a positive tag and the other had a negative tag. Clearly positive sentences (i.e., sentences without any negative operators) and clearly negative sentences (i.e., sentences with overt sentential negation and no other relevant operators present) were used as controls. The relative frequency of positive and negative tags was then taken as a measure of the sentential negativity of each experimental item. Our main finding is that sentential negativity is a graded notion, sensitive to both semantic and syntactic factors. With respect to semantics, we find that n-words contribute more negativity than downward entailing operators, confirming the logical distinction between anti-additivity and downward entailment identified in the previous semantic literature on NPI licensing. With respect to syntactic position, we find that negative items in subject or adverbial position contribute more negativity than negative items in direct object position.

1 Introduction

The most straightforward way to mark sentential negativity is by means of an explicit, dedicated sentential negation operator. This is exemplified in (1-a), whose positive counterpart is given in (1-b).

(1)   a.   Susan will not leave.   NEGATIVE
      b.   Susan will leave.   POSITIVE

However, it has been well-known at least since Klima (1964) that there are many sentence types that do not contain an overt, dedicated sentential negation operator, but may still be called 'negative' in light of the fact that they exhibit striking similarities with clear-cut negative sentences like (1-a) and dissimilarities with clear-cut positive sentences like (1-b). Such sentence types are exemplified in (2):

(2) a. Susan never left. (n-word adverb)
     b. Nobody left. (n-word subject)
     c. Susan saw nobody. (n-word object)
     d. Susan rarely left. (downward entailing adverb)
     e. Few students saw Susan. (downward entailing subject)
     f. Susan saw few students. (downward entailing object)

A large body of literature since Klima (1964) and Ladusaw (1979) has concentrated on the fact that the sentences in (2) pattern with (1-a) rather than (1-b) in licensing Negative Polarity Items (NPIs) like *any*. Upon closer examination, the category of NPIs has turned out to be quite complex, leading researchers to posit finer grained classifications of NPIs and of negative contexts – see van der Wouden (1994), Krifka (1995), Zwarts (1995), Zwarts (1998), Levinson (2008), Giannakidou (2011), Gajewski (2011) and Hoeksema (2012) among others.

For instance, a basic distinction has been made between *weak* and *strong* NPIs, where weak NPIs, like *any*, are licensed in any downward entailing environment, while strong NPIs, like *until* or *lift a finger*, are only licensed in environments that are not just downward entailing, but have the stronger property of being anti-additive.[1] Sentential negation and n-words like *never* and *nobody*, featuring in examples (2-a)-(2-c), are anti-additive, while items like *rarely* and *few*, featuring in examples (2-d)-(2-f) are downward entailing but not anti-additive.[2] This is reflected by the potential of these items to license weak and strong NPI, as exemplified in (3) and (4).

(3) a. Nobody said anything. / Susan never said anything.
     b. Nobody lifted a finger. / Susan never lifted a finger.

(4) a. Few students said anything. / Susan rarely said anything.
     b. *Few students lifted a finger. / *Susan rarely lifted a finger.

---

[1]For a more detailed discussion of the licensing conditions for different subclasses of strong NPIs, see Krifka (1995) and Gajewski (2011) among others.

[2]A function $F(\cdot)$ from sets to truth-values, i.e., of quantifier type $\langle\langle et\rangle t\rangle$, is downward entailing iff $X \subseteq Y$ implies that $F(Y) \to F(X)$. For example:

(i)   *Few students are tall* $\to$ *Few students are tall and blond*

because if less than half (or whatever other contextually salient proportion we choose) of the students are tall, then definitely less than half of the students are tall and blond.
Similarly:

(ii)   *No students are tall* $\to$ *No students are tall and blond*

A function $F(\cdot)$ of quantifier type $\langle\langle et\rangle t\rangle$ is anti-additive iff $F(X \cup Y) \leftrightarrow F(X) \wedge F(Y)$. Anti-additivity is stronger than downward entailment. For example:

(iii) *No students are tall or blond* $\leftrightarrow$ (*No students are tall* $\wedge$ *No students are blond*)

In contrast, the same equivalence does not hold if we replace *no* by *few*:

(iv)   *Few students are tall or blond* $\leftrightarrow$ (*Few students are tall* $\wedge$ *Few students are blond*)

Consider for example a set of 7 students, 2 of which are tall but brunette, 2 of which are blond but short, and the remaining 3 of which are brunette and short. In this situation, it is false to say that few students are tall or blond: 4 out 7 students are either tall or blond. But it is true to say both that few students are tall (only 2 out of 7) and that few students are blonde (only 2 out of 7). So *no students* and *few students* are both downward entailing, but only the former is anti-additive.

Based on such observations, it is tempting to think of n-words like *never* and *nobody* as being, in some sense, *more* negative than items like *rarely* and *few*. This paper addresses a number of issues that arise from this line of thought, listed in (5):

(5) a. Is it possible to *quantify* the degree of negativity of the sentences in (2)?
     b. Can some of the sentences in (2) be said to be more negative than others?
     c. If so, which factors affect the degree of negativity?

In order to address these questions, we will identify a phenomenon that clearly distinguishes the two simple cases in (1), and then investigate to what extent the sentences in (2) pattern like (1-a) or like (1-b). The closer the behavior of a sentence is to the behavior of (1-a), the more 'negative' we will take that sentence to be.

The phenomenon that we will focus on in this paper is the formation of *tag questions*. Even though Klima (1964) already noted that tag questions based on positive sentences are markedly different from tag questions based on negative sentences, this construction has received relatively little attention as a negativity probe (as compared, for instance, to the licensing of NPIs). The main idea is to rely on the polarity of a question tag to probe the polarity of the sentence that anchors the tag.

The literature on tag questions in English has distinguished two types of tag questions, exemplified in (6) below, which we call here reverse (REV) tag questions and reduplicative (RED) tag questions, respectively:

(6) a. Anne left, didn't she? / Anne didn't leave, did she?   REV tag
     b. Anne left, did she?   RED tag

In (6-a) the polarity of the tag question is the reverse of that of the sentence it attaches to, which we call the *anchor* of the tag. In (6-b), the polarity of the tag is the same as that of the anchor, which is why they are often called same-polarity tags in the literature (Klima 1964; Quirk *et al.* 1985; McCawley 1998; Swan 2005).

Tag questions are a good negativity probe for two reasons: (i) the polarity of the tag is clearly marked, and (ii) the polarity of the tag is sensitive to the polarity of the anchor.[3] With REV tags, unquestionably positive sentences anchor negative tags, while unquestionably negative sentences anchor positive tags. With RED tags, the polarity of the tag must be the same as that of the anchor, and we expect that this polarity has to be positive, given that it is often assumed in the literature that tag questions whose anchor and tag are both negative do not occur (Quirk *et al.* 1985; McCawley 1998).

Turning now to unclear cases, the closer the behavior of a sentence is to an unquestionably negative sentence with respect to the polarity of the tag it anchors, the more negative we may take the sentence to be. The q-tag test for negativity then is based on gauging the negativity of the anchor from the polarity of the tag. In this paper we report on an experiment that uses the q-tag test in order to quantify the negativity of the sentence types illustrated in (2) based on a comparison with those in (1).[4]

---

[3]It is possible, of course, that different probes for negativity would lead to different results than those obtained here with q-tags. Indeed, McCawley (1998) already pointed out that the different tests for sentential negation developed by Klima (1964) may not all converge. However, the important task of experimentally comparing the q-tags with other negativity probes is beyond the scope of this paper.

[4]As an anonymous reviewer emphasizes, Payne (1985) and Penka (2011) point out that q-tags are sensitive to the scope of the negative operator and not just its presence, as Klima (1964) had assumed. This is illustrated by the examples below, showing a contrast in the ability to license a reverse positive tag depending on whether the negative operator in the anchor has *widest* scope or not:

However, before we can begin using tag questions to measure negativity, it is crucial to be able to differentiate the two types of tag questions just mentioned. Fortunately, this has been done in the previous literature, where it has been noted that the two types of tag questions serve different discourse functions. Thus, the function of REV tags is to ask confirmation from the Addressee for a proposition that the Speaker commits herself to, albeit somewhat tentatively (Cattell 1973; McCawley 1998; Malamud and Stephenson 2011; Farkas and Roelofsen 2012). RED tags, on the other hand, are used by the Speaker to challenge or question a proposition that she feels the interlocutor committed herself to (Cattell 1973; McCawley 1998). The tone of a RED tag may be one of sarcastic suspicion (Quirk *et al.* 1985, 812). In order to turn tag questions into a useful test for negativity then, we attempt to differentiate contexts that encourage the use of REV tags from those that encourage the use of RED tags. A context is successful in inviting REV tags if an unquestionably positive sentence anchors a negative tag in that context, while an unquestionable negative sentence anchors a positive tag in the same context; a context is successful in inviting RED tags if unquestionably positive sentences anchor positive tags in that context.

Our strategy therefore is the following: we design an experiment that investigates the type of tags anchored by sentences such as (2) and compare to what extent they are similar to those anchored by the clear-cut cases in (1). We will see that the contexts we use to invite REV tags are indeed successful in doing so, while the contexts we use to invite RED tags are more problematic. We will therefore focus our attention on tag-question patterns in REV contexts when we discuss our results.

The primary question we ask then is to what extent the sentences in (2) behave like (1-a) in contexts that invite REV tags. In such contexts, the more likely a sentence is to anchor a positive tag the more negative it is. We break this question down into the following more specific questions:

- Do n-words affect the negativity of the sentence they occur in?

- Do DE-items affect the negativity of the sentence they occur in?

- If the answer to both of the above questions is *yes*, is the effect equally strong?

- Does the syntactic position of an n-word/DE-item affect the negativity it induces?

- If the answer to the last question is *yes*, how does the syntactic parameter (syntactic position) interact with the semantic one (n-word vs DE-item)?

We will find that the negativity of a sentence as measured by the q-tag test is influenced by two interacting scales, given in (7):

(7) a. n-words ≫ DE-items   SEMANTIC SCALE
     b. ADV, SUBJ ≫ OBJ   SYNTACTIC SCALE

(i) a. John doesn't often pay taxes, does he/*doesn't he?
     b. John often doesn't pay taxes, *does he/doesn't he? (Payne 1985:200)

However, the issue of the scope of the negative element relative to other operators in the anchor does not arise in our experimental setting, given that our test-examples only involve definite DPs.

The semantic scale is the primary one, and each level of the semantic scale is further refined by the syntactic scale. This means that both n-words and DE-items cause a sentence to pattern more closely to unquestionably negative sentences with respect to REV question tags, but this effect is always stronger in the case of n-words than in the case of DE-items, no matter what syntactic position n-words or DE-items are found in. As such, the experiment based on tag questions confirms the difference in negative strength between anti-additive and downward entailing operators argued for in the NPI-licensing literature (van der Wouden 1994; Zwarts 1995, 1998; Levinson 2008; Giannakidou 2011; Gajewski 2011; Hoeksema 2012).

Although subordinate to the semantic scale, the syntactic scale has a clear effect: the higher an n-word or a DE-item is on the syntactic scale, the stronger its effect on the negativity of the sentence in which it occurs (when compared to n-words or DE-items, respectively). However, even the least negative n-words, i.e., n-words in an OBJ position, are clearly more negative than the most negative DE-items, which are found in ADV or SUBJ positions.[5]

In Section 2, we present an experiment designed to use the q-tag test to measure the negativity of the sentence types in (2). Section 3 presents the data analysis and resulting generalizations. Section 4 discusses the theoretical implications of these results and Section 5 concludes with some directions for further research.

2 Experimental method and participants

2.1 Experimental method

We used a two-alternative forced choice task to test the negativity of sentences like the ones in (2) above when they are anchors for q-tags. We manipulated two features of the sentences / anchors:

    i. the nature of the negative element: N-WORDS vs DE-quantifiers;

    ii. the syntactic position of this negative element: SUBJECT vs OBJECT vs ADVERB.

In addition, we used POSITIVE and NEGATIVE sentences as controls: they provided a way to measure that the behavior elicited by the task was the one relevant to our theoretical goals, and they also provided reference levels relative to which we could measure the effects of our experimental manipulations.

These 8 conditions (2 types of negative elements $\times$ 3 syntactic positions, plus 2 controls) were tested in both REV and RED contexts, for a total number of $8 \times 2 = 16$ conditions.

We were interested in whether the participants chose:

- either a positive q-tag, henceforth coded as YES

---

[5]Note that throughout this paper we talk about DE-items but in the experiment we only used the adverb *rarely* and the determiner *few*. The reason we chose *few* and *rarely* as DE QPs is because other DE QPs, like *at most N* and *not all*, consist of a superlative and an explicit negative marker, respectively, which might influence the results. Nevertheless, it would be interesting to also test the behavior of these DE QPs. The question of how typical the behavior of *few* and *rarely* is for the whole class of DE-items is left open here. For relevant discussion, see Gajewski (2011) and Horn (1989), among others.

- or a negative q-tag, henceforth coded as NO

As mentioned in section §1, we expected the two contexts REV and RED to trigger different kinds of tags given the generalizations in the previous literature:

- REV contexts: these were contexts in which the interlocutor knew and was an authority for the content of the anchor, so the q-tag was most obviously used to ask for confirmation; we therefore expected the polarity of the q-tag to be the reverse of the polarity of the anchor;

- RED contexts: these were contexts in which the interlocutor finds the content of the anchor hard to believe, so the q-tag was most obviously used to express skepticism / suspicion / sarcasm; given the previous literature, we expected the polarity of the q-tag to be the same as the polarity of the anchor.

Every experimental item contained a subject, a direct object and an adverb. These were all referential (definites) unless they were experimentally manipulated, in which case they were switched to an N-WORD or DE quantifier. Thus, the control conditions POSITIVE and NEGATIVE had a referential subject, a referential object and a referential adverb and they differed only with respect to the absence vs presence of an overt sentential negation operator. Example items for each of the 16 conditions are provided below. For each item, the participants were required to choose one response out of the two options by clicking on it with their mouse; both responses were available for every stimulus and they were displayed in random order (randomized for every participants and every stimulus).

(8) POSITIVE

   a. REV: Mary has just told Jane something about government representatives. Jane knows it already and says:

   "Aha, I was right! The government representative visited the colonies this year, $\left\{ \begin{array}{l} \text{did he?} \\ \text{didn't he?} \end{array} \right\}$"

   b. RED: Mary has just told Jane something about government representatives. Jane finds it hard to believe and says:

   "You don't say! The government representative visited the colonies this year, $\left\{ \begin{array}{l} \text{did he?} \\ \text{didn't he?} \end{array} \right\}$"

(9) NEGATIVE

   a. REV: Geoff has just told Ryan something about composers. Ryan knows it already and says:

   "Aha, I was right! The composer did not use the cello in his late period, $\left\{ \begin{array}{l} \text{did he?} \\ \text{didn't he?} \end{array} \right\}$"

   b. RED: Geoff has just told Ryan something about composers. Ryan finds it hard to believe and says:

   "You don't say! The composer did not use the cello in his late period, $\left\{ \begin{array}{l} \text{did he} \\ \text{didn't he?} \end{array} \right\}$"

(10) N-WORD-ADV

   a. REV: Lilly has just told Adrian something about the children. Adrian knows it already and says:

   "Aha, I was right! The children never understood the science experiments, $\left\{ \begin{array}{l} \text{did they} \\ \text{didn't they?} \end{array} \right\}$"

b. RED: Lilly has just told Adrian something about the children. Adrian finds it hard to believe and says:

You don't say! The children never understood the science experiments, $\left\{ \begin{array}{l} \text{did they} \\ \text{didn't they?} \end{array} \right\}$"

(11) DE-ADV

a. REV: Bob has just told Justine something about the athletes. Justine knows it already and says:

"Aha, I was right! The athletes rarely skipped the practice games, $\left\{ \begin{array}{l} \text{did they} \\ \text{didn't they?} \end{array} \right\}$"

b. RED: Bob has just told Justine something about the athletes. Justine finds it hard to believe and says:

You don't say! The athletes rarely skipped the practice games, $\left\{ \begin{array}{l} \text{did they} \\ \text{didn't they?} \end{array} \right\}$"

(12) N-WORD-SUBJ

a. REV: Lindsey has just told Betty something about the anthropologists. Betty knows it already and says:

"Aha, I was right! No anthropologists discussed the social concerns at the January conference, $\left\{ \begin{array}{l} \text{did they} \\ \text{didn't they?} \end{array} \right\}$"

b. RED: Lindsey has just told Betty something about the anthropologists. Betty finds it hard to believe and says:

"You don't say! No anthropologists discussed the social concerns at the January conference, $\left\{ \begin{array}{l} \text{did they} \\ \text{didn't they?} \end{array} \right\}$"

(13) DE-SUBJ

a. REV: Josephine has just told Ward something about the hostages. Ward knows it already and says:

"Aha, I was right! Few hostages met the envoys today, $\left\{ \begin{array}{l} \text{did they} \\ \text{didn't they?} \end{array} \right\}$"

b. RED: Josephine has just told Ward something about the hostages. Ward finds it hard to believe and says:

"You don't say! Few hostages met the envoys today, $\left\{ \begin{array}{l} \text{did they} \\ \text{didn't they?} \end{array} \right\}$"

(14) N-WORD-OBJ

a. REV: Rick has just told Ann something about the travel agents. Ann knows it already and says:

"Aha, I was right! The travel agents visited no Greek islands last summer, $\left\{ \begin{array}{l} \text{did they} \\ \text{didn't they?} \end{array} \right\}$"

b. RED: Rick has just told Ann something about the travel agents. Ann finds it hard to believe and says:

"You don't say! The travel agents visited no Greek islands last summer, $\left\{ \begin{array}{l} \text{did they} \\ \text{didn't they?} \end{array} \right\}$"

(15) DE-OBJ

a. REV: Brad has just told Tom something about the managers. Tom knows it already and says:

"Aha, I was right! The managers revised few contracts this month, $\left\{ \begin{array}{l} \text{did they} \\ \text{didn't they?} \end{array} \right\}$"

b. RED: Brad has just told Tom something about the managers. Tom finds it hard to believe and says:

"You don't say! The managers revised few contracts this month, $\left\{\begin{array}{l}\text{did they}\\\text{didn't they?}\end{array}\right\}$"

As we already observed, we expect POSITIVE sentences like (8) above to discriminate very well between the REV and RED contexts: POSITIVE sentences should be overwhelmingly associated with NO (i.e., negative tag) responses in REV contexts and with YES (i.e., positive tag) responses in RED contexts. Furthermore, if the generalizations in the literature are correct (see, e.g., Quirk *et al.* 1985), we expect NEGATIVE sentences as in (9) above to not discriminate between these REV and RED contexts.

Our overall goal then is to locate N-WORDS and DE quantifiers in various syntactic positions within the interval whose limits are provided by POSITIVE and NEGATIVE sentences. We used 40 experimental items, provided in appendix A. In addition, we used 60 fillers, 10 in each of the six categories listed and exemplified below.

(16) Fillers:

    a. referential, positive and 'surprised':

        Jane has just told Lisa something about the plants in the front yard. Lisa is surprised by what Jane said and says:

        "Oh, how unexpected! So the plants in the front yard blossomed yesterday morning, $\left\{\begin{array}{l}\text{did they}\\\text{didn't they?}\end{array}\right\}$"

    b. referential, positive and 'obvious':

        Bob has just told Jacob something about the surgeon. Jacob thinks that Bob is stating the obvious and says:

        "That's obvious! The surgeon did the operation early in the morning, $\left\{\begin{array}{l}\text{did he}\\\text{didn't he?}\end{array}\right\}$"

    c. referential, negative and 'surprised':

        Bea has just told Ulrike something about the amusement park. Bea is surprised by what Ulrike said and says:

        "Oh, how unexpected! So the amusement park did not reopen the rollercoaster after the recent accident, $\left\{\begin{array}{l}\text{did it}\\\text{didn't it?}\end{array}\right\}$"

    d. referential, negative and 'obvious':

        Parsifal has just told Theo something about the online bookstores. Theo thinks that Parsifal is stating the obvious and says: "That's obvious! The online bookstores did not ship the books free of any charge, $\left\{\begin{array}{l}\text{did they}\\\text{didn't they?}\end{array}\right\}$"

    e. quantificational and 'surprised':

        Leo has just told Dan something about the violins in the orchestra. Dan is surprised by what Leo said and says: "Oh, how unexpected! So every violin in the orchestra is made by European craftsmen, $\left\{\begin{array}{l}\text{is it}\\\text{isn't it?}\end{array}\right\}$"

    f. quantificational and 'obvious':

        Louis has just told Marcel something about coffees sold by the pound. Marcel thinks that Louis is stating the obvious and says: "That's obvious! Most coffees sold by the pound come from some Latin American country, $\left\{\begin{array}{l}\text{do they}\\\text{don't they?}\end{array}\right\}$"

Every participant saw each of the 40 items in one of the 16 conditions and all 60 fillers, for a total of 100 stimuli. The condition in which an item appeared was initially randomly chosen for every item and then rotated for every participant (following closely a Latin square design). Every participant saw every item exactly once in one of the 16

RED	NO	YES
NEGATIVE	33	254
N-WORD-ADV	36	252
N-WORD-OBJ	38	248
N-WORD-SUBJ	43	244
DE-ADV	81	207
DE-SUBJ	116	171
DE-OBJ	126	162
POSITIVE	155	134

REV	NO	YES
N-WORD-SUBJ	60	223
N-WORD-ADV	71	221
NEGATIVE	75	214
N-WORD-OBJ	100	186
DE-ADV	170	119
DE-SUBJ	183	103
DE-OBJ	241	45
POSITIVE	275	14

Figure 1: Cross-tabulation of responses and conditions in increasing order of NO (i.e., negative tag) responses: RED context in the left table and REV context in the right table.

conditions. The order of the 100 stimuli was randomized for every participant. The order in which the two choices containing the two tags (positive and negative) appeared was randomized for every participant and every stimulus the participant saw.

2.2 Participants

There were 117 participants in the experiment, all UC Santa Cruz undergraduate students who completed the experiment for credit or extra-credit. The participants completed the experiment online. As already indicated, 40 observations were collected from every participant, for a total of 4680.

In order to filter out the participants who did not complete the experiment properly and simply clicked through it, we examined which participants took less than 1000 ms to read the entire context and decide between the two alternative replies. There were 10 participants out of 117 who had at least 2 answer times less than 1000 ms out of 40.

We also examined the pattern of responses to the POSITIVE & REV condition for all participants: participants should have exclusively selected NO answers (i.e., negative tags) in this condition. And indeed, the overwhelming majority of the participants (106 out of 117) had only NO responses to this condition; the remaining 11 participants had at least one YES response. But since those responses might simply be mistakes (hence should be subsumed under measurement / observation-level error), we only excluded those participants who appeared on both lists, i.e., participants who both rushed through the items (i.e., they took less than 1000 ms to respond), and incorrectly answered YES to the POSITIVE & REV control condition. There were two such participants. Thus, the number of participants included in the final analysis is 115.

3 Data analysis and resulting generalizations

3.1 Descriptive summaries

The cross-tabulation of responses and fixed effects (the experimental conditions) in ascending order of NO responses is provided in Figure 1. The corresponding barplots, which are easier to examine at a glance, are provided in Figure 2. The generalizations that can be drawn from these results are the following:

Yes (white) / No (black) counts in RED context

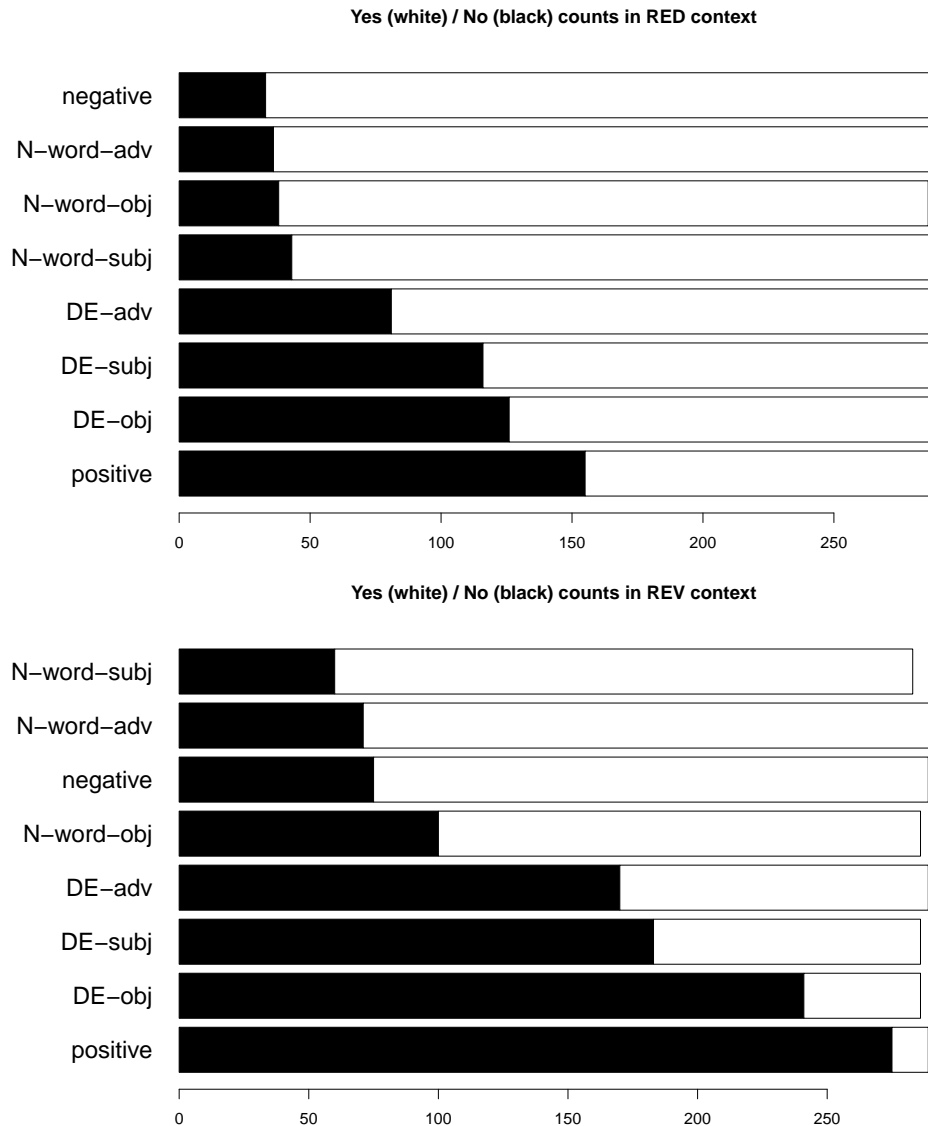**Yes (white) / No (black) counts in REV context**

Figure 2: Barplots for responses in the 16 conditions (YES: white; NO: black) grouped by context: RED in the top panel and REV in the bottom panel.

- N-WORDS are clearly more negative than DE quantifiers; what we mean here by 'more negative' is that they had a clearly higher proportion of positive responses; in the interest of brevity, we formulate all our generalizations in terms of being 'more / less / just as negative', but the reader should keep in mind that the actual, behavioral correlate of the 'degree of negativity' is the relative proportion of YES/NO responses;[6]

- within N-WORDS, N-OBJECTS are less negative than N-SUBJ and N-ADV and also less negative than the NEGATIVE controls; this contrast is seen in the REV context but is neutralized in the RED context;

- overall, there is no clear difference between N-ADVERBS and N-SUBJECTS in either the RED or the REV context;

---

[6]We are indebted to an anonymous reviewer for emphasizing this point.

- within DE quantifiers, we see two distinct patterns in the RED and REV contexts;

- DE quantifiers in REV: DE-OBJECTS are clearly less negative than all the others (except for the POSITIVE controls); DE-ADVERBS and DE-SUBJECTS are basically the same and they are clearly less negative than any N-WORDS and clearly more negative than DE-OBJECTS;

- DE quantifiers in RED are clearly less negative than N-WORDS and more negative than POSITIVE controls, just as in REV; but within DE, there is no difference between DE-SUBJ and DE-OBJ: both are less negative than anything else except POSITIVE controls; however, DE-ADVERBS in RED stand out: they are clearly more negative than DE-SUBJECTS or DE-OBJECTS;

- there is a difference between DE-ADVERBS and DE-SUBJECTS in RED contexts, but not in REV contexts.

Given that both the NEGATIVE and the POSITIVE controls behaved as expected in the REV context but less so in the RED context, we take the former to be less noisy than the latter. So whenever there are discrepancies between the two contexts, we will focus primarily on the generalizations supported by the REV context.

The generalizations that emerge based on the REV context are as follows:

- N-ADVERBS and N-SUBJECTS are as negative as the negative controls; therefore n-adverbs, n-subjects and negative controls form a single class with respect to their effect on the polarity of the q-tag;

- N-OBJECTS are less negative than the previous group; they are clearly more negative than everything else, but not as negative as NEGATIVE controls, N-SUBJECTS and N-ADVERBS;

- thus, N-WORDS and NEGATIVE controls form a clearly separate class with two subclasses;

- DE quantifiers also form a clearly separate class: they are less negative than N-WORDS and more negative than POSITIVE controls;

- within the DE class, we can also clearly identify the subclass of DE-ADVERBS and DE-SUBJECTS, which are more negative than DE-OBJECTS.

For completeness, the generalizations supported by the data for the RED context are the following:

- the major separation into classes is the same as for REV: N-WORDS and NEGATIVE controls are the most negative, DE quantifiers are less negative and POSITIVE controls are the least negative;

- but the POSITIVE controls are definitely not identified as clearly positive, unlike in the REV context: the POSITIVE controls seem to be identified as positive half the time and as negative the other half;

- this indicates that the RED context was much noisier and did not work as intended; therefore, it might well preserve the major distinctions between classes, but not the finer grained ones;

- we see that for the N-WORD class: there are no differences between N-OBJECTS on one hand and N-ADVERBS / N-SUBJECTS on the other;

- we also partially see that for the DE class: DE-OBJECTS are not distinct from DE-SUBJECTS;

- but DE-ADVERBS are now distinct from both DE-SUBJECTS and DE-OBJECTS; this might very well be a chance difference in a noisy environment due to sampling 'error', but it might also be a real difference that somehow became obscured in the REV context;

- clarifying the situation here can only be done in a follow-up investigation in which RED contexts are independently tested and shown to work as intended; we leave this for future research.

To summarize, the major distinctions in RED contexts replicate and reinforce the distinctions in the more trustworthy REV environment. There is one distinction that is visible only in RED contexts, but we leave the investigation of this difference for a future occasion. In the rest of the paper we focus on REV contexts.

3.2   Statistical analysis

Statistical analysis of the data can give us a measure of (un)certainty for the generalizations suggested by the above descriptive summaries. Given that the response variable is binary (categorical), we use mixed-effects logistic regression models to do the data analysis. The item random effects account for practically no variance, so we will henceforth omit them.

The model that seems to strike the best balance between data fit and simplicity[7] is the one with full fixed-effect structure (main effects for 8 conditions and the 2 contexts plus their interactions) and subject random effects for the intercept and the context slopes. More details about this model are provided in appendix B.1.

We can quantify our (un)certainty about the generalizations listed above by examining the full posterior distributions of the model parameters in a Bayesian model with vague / low-information priors. More details about this model and how the posterior distributions were estimated are provided in appendix B.2.

We will simply focus here on the plot of median probabilities and 95% credible intervals (CRIs) provided in Figure 3 below. This plot closely mirrors the barplots in Figure 2 above. The points in the center of the bars represent the median probabilities of a YES, i.e., a positive tag, response in the corresponding condition and context. For example, if we look at the leftmost condition, i.e., the POSITIVE controls, the position of the point inside the continuous bar indicates that the probability of a positive tag in REV contexts for positive controls is very close to 0. In contrast, if we look at the NEGATIVE controls, i.e., the 6th condition from the left, the point inside the dotted bar for NEGATIVE controls indicates that the probability of a positive tag in RED contexts is very close to 1 (equivalently, very close to 100%).

The bars themselves indicate our uncertainty in these estimated probabilities. Looking again at POSITIVE controls in REV contexts, we see that we are 95% confident that the probability of a positive polarity tag is anywhere between basically 0 and 0.05, i.e.,

---

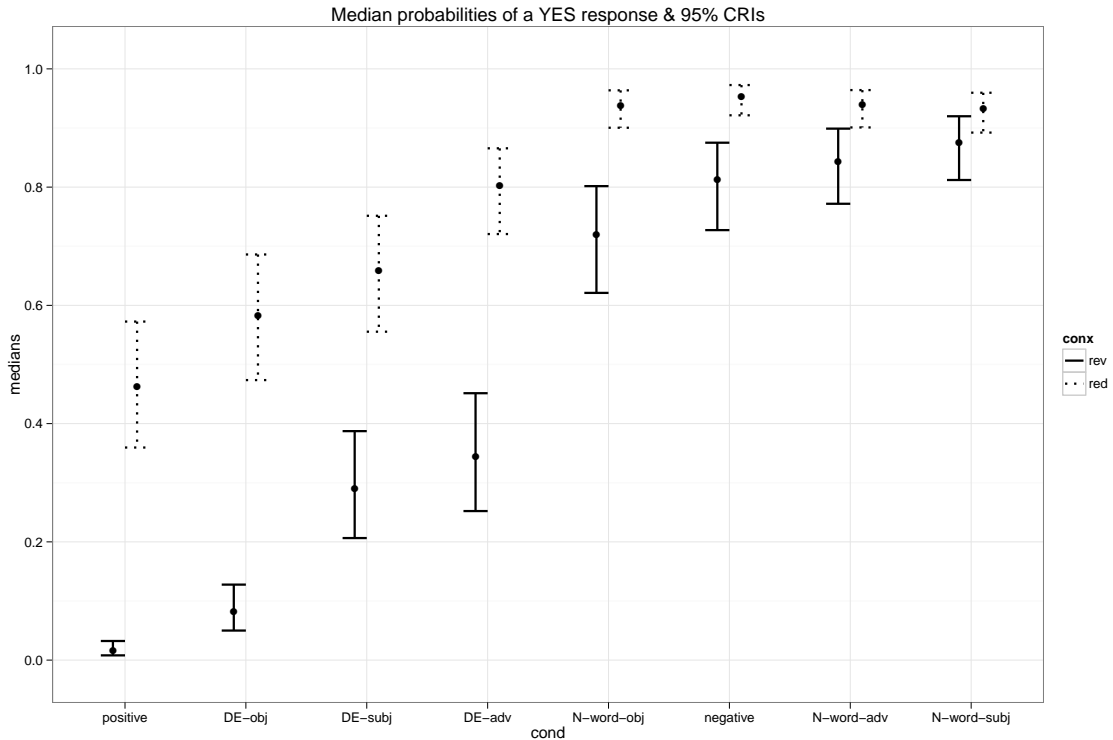[7]We used the Bayesian Information Criterion (BIC) as a guide for this.

Figure 3: Median probabilities and 95% credible intervals (CRIs) for YES (i.e., positive tag) responses: horizontal axis – the 8 conditions; vertical axis – probabilities; REV estimates – continuous lines; RED estimates – dotted lines. The 8 conditions on the horizontal axis are listed in ascending order of median probability of YES in the REV context.

anywhere between 0% and 5%. Having 95% confidence in this interval – or equivalently, taking this interval to be 95% credible – means that we believe probabilities outside the 0-to-0.05 interval to be very unlikely, although we do not completely rule them out. For our purposes, however, we are willing to disregard anything that is so unlikely. The points together with the bars / intervals around them are very similar to the 'probability of being elected' estimates $+/-$ some error that we often hear during electoral campaigns.

This plot of median probabilities and 95% CRIs provides a bird's-eye view of our experimental results. We will focus on various specific issues in the remainder of this subsection, but the broad view provided by this plot is useful in several respects.

First, we see that the 8 conditions in the REV context, i.e., the continuous bars, are much more clearly separated into distinct groups than the corresponding ones in the RED context, i.e., the dotted bars. Thus, as we suspected, the RED context is much more noisy, which is why we concentrate on the REV context.

Focusing now on the REV context, we see four clear groupings / subclasses:

a. the POSITIVE controls are clearly the most positive kind of sentences; their probability of getting a YES (positive tag) response is very close to 0 and their 95% CRI excludes all the other 7 CRIs, including the one for DE-OBJECTS, which are the most positive sentences after the controls; that is, we are very confident that we can generalize the results from the sample we obtained in our particular experiment to the underlying phenomena;

13

b. DE-OBJECTS are the next most positive sentences; their 95% CRI excludes all the 6 CRIs for the remaining conditions;

c. DE-SUBJECTS and DE-ADVERBS are more negative than the previous two conditions but more positive than the N-WORDS and the NEGATIVE controls; just by looking at this plot, we cannot tell whether DE-SUBJECTS are more positive than DE-ADVERBS; to see whether there is a difference between them we will have to examine the distribution of the difference in probability between the two of them;

d. the final subclass includes the N-WORDS and the NEGATIVE controls, which are clearly more negative than all the other four conditions; once again, we cannot determine yet if there are differences within this subclass just by looking at this plot.

Thus, the two outstanding questions (for the REV context) that the plot in Figure 3 above does not answer are (*i*) whether there is any difference between DE-SUBJECTS and DE-ADVERBS and (*ii*) whether there are any differences between conditions in the N-WORD and NEGATIVE subclass.

These questions are answered by the seven plots in Figure 4 below.[8] These plots provide the posterior distributions of the differences in probability that are of interest to us and they should be read as follows. The vertical dotted line on top of 0 in each plot indicates a difference of 0 between the two probabilities whose difference is plotted there.

- if this dotted line falls within the thick black horizontal segment in each plot, then a difference of 0 between those two probabilities is credible, i.e., we really cannot rule out the possibility that the two underlying probabilities are essentially the same;[9]

- if the dotted line falls outside the horizontal segment, then a difference of 0 between the two probabilities under consideration is very unlikely and we should take the underlying probabilities to be different.

The single plot in the top row of Figure 4 indicates that DE-SUBJECTS and DE-ADVERBS are not really different. Numerically, DE-ADVERBS are more negative than DE-SUBJECTS, but it is very probable that this numerical difference is just an artifact of the particular sample of observations we collected. That is, it is likely that this numerical difference does not reflect a true underlying difference in sentential negativity between DE-SUBJECTS and DE-ADVERBS.

The ability to draw this kind of conclusions is one of the main reasons for statiscal modeling and inference over and above the inspection of the raw numbers and the corresponding barplots: apparent differences in the raw numbers might simply be an artifact of the particular sample we collected. Correspondingly, differences that do not surface in the raw numbers might actually be uncovered when suitably realistic statistical models (e.g., our mixed-effects logistic regression models) are estimated.

---

[8]Code based on Kruschke (2011).

[9]The thick black horizontal segment provides the highest density, i.e., the smallest possible, 95% CRI for the difference in probabilities.
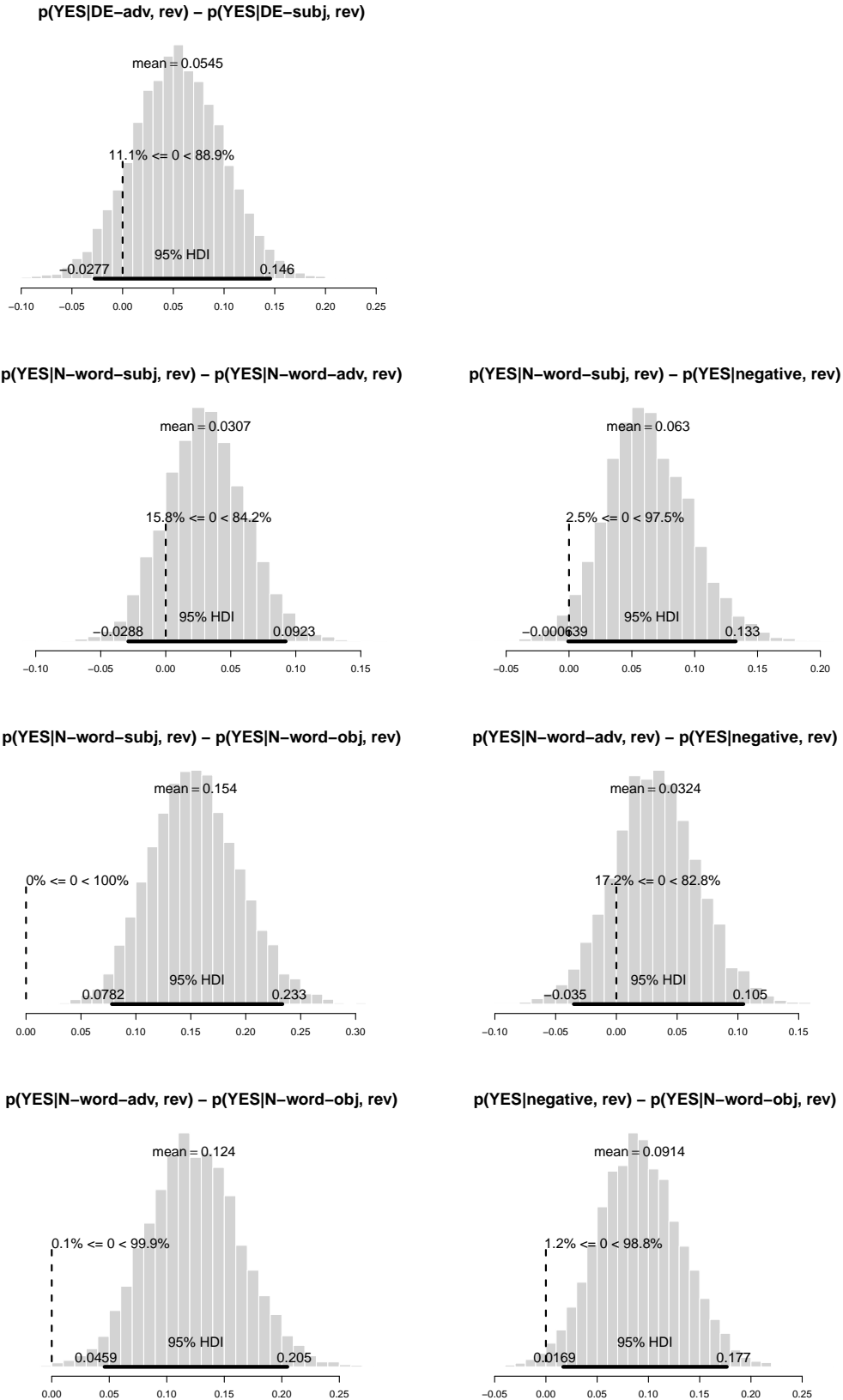
Figure 4: Differences in probabilities of YES responses between the specific conditions in the REV context.

Moving on to the remaining six plots of Figure 4, we see that NEGATIVE controls,

N-SUBJECTS and N-ADVERBS are not really distinct. But N-OBJECTS are clearly less negative than any of these three conditions.

3.3 Summary

The descriptive summaries and subsequent statistical analysis of the experimental results indicate that the REV context elicited the expected kind of behavior from the participants, while the RED context is much more noisy.

Focusing on the REV context, then, the emerging hierarchy of sentential negativity is the following (from more negative to less negative):

(17)   {NEG, N-ADV, N-SUBJ} ≫ N-OBJ ≫ {DE-ADV, DE-SUBJ} ≫ DE-OBJ ≫ POS

4   Discussion and theoretical implications

If we omit the two control conditions, the hierarchy of sentential negativity that emerges from our experimental results is the following:

(18)   {N-ADV, N-SUBJ} ≫ N-OBJ ≫ {DE-ADV, DE-SUBJ} ≫ DE-OBJ

A natural way of making sense of this hierarchy is to see it as the result of a primarily semantics-based ordering N-WORD ≫ DE-item which is further refined within each of the two ordered strata by a syntactic ordering {ADV, SUBJ} ≫ OBJ. More formally, the experimental results can be derived if we assume a lexicographic order of the 8 experimental conditions according to the semantic order in (19-a) first, followed by the syntactic order in (19-b).

(19)   a.   Semantic hierarchy of negativity: N-WORD ≫ DE-item
        b.   Syntactic hierarchy of negativity: {ADV, SUBJ} ≫ OBJ

The fact that the degree of negativity of a sentence is influenced by these two interacting scales is a new result. However, the details of each scale are not surprising. The fact that n-words are more negative than DE-items is in line with findings in the literature on NPI licensing, which we already briefly illustrated in the introduction. In particular, n-words are anti-additive, just like sentential negation, and as such license both strong and weak NPIs, while items like *few* and *rarely* are downward entailing but not anti-additive, and as such only license weak NPIs. Assuming that anti-additivity renders an item more negative from a semantic point of view, n-words and sentential negation are semantically 'stronger' negatives than other DE-items.

Whether this semantic distinction is indeed the operative notion here depends on whether the effects we found with the DE-items *few* and *rarely* generalize to other DE-items. If the results do not generalize to other DE-items, the behavior of *few* and *rarely* will have to be explained in terms of the special properties they have (see, for instance, Gajewski 2011).

With respect to the syntactic scale, the accounts in McCawley (1998) and Moscati (2006) distinguish between the strength of the influence of N-WORD SUBJECTS vs. that of N-WORD OBJECTS based precisely on the type of REV tags anchored to sentences involving them. Our experiments confirm this distinction and extend the empirical domain to adverbs. Thus, our results partly confirm previous observations but also go

beyond them. Overall, they point towards accounts where both syntactic and semantic factors are taken into consideration and in which the negativity of a sentence is treated as a matter of degree.

Various approaches that could account for the syntactic scale can be imagined. For example, we might want to attribute the entire syntactic scale to processing effects rather than to syntax *per se*, broadly following the view of expectation-based syntactic processing in Hale (2001, 2011); Levy (2008).[10] For simplicity, let's focus on the SUBJ ≫ OBJ part of the syntactic scale, and on n-words only. We could assume that a default assumption made by hearers as they incrementally process a sentence is that the sentence is positive unless signaled otherwise. For a sentence with an n-word in subject position like *No representatives visited the colonies*, the signal that the default positivity assumption is incorrect comes very early and the revision of this assumption is inexpensive. In contrast, for a sentence with a direct-object n-word like *The representatives visited no colonies*, the signal that the sentence is in fact negative comes late, after the 'commitment to positivity' is relatively entrenched, so revising the default positivity assumption is more expensive / less effective. The gradient effects we found would then be connected to the cost of the revision: the cheaper the revision, the more likely it is to be activated, resulting in a positive REV tag; the more costly the revision, the less likely it is to be activated, resulting in a negative REV tag.

A more 'competence-driven' syntactic account may also be pursued (e.g., building on De Clercq, 2011; De Clercq *et al.*, 2012). For example, it might be that every anchor sentence has an abstract functional projection that hosts a 'polarity' feature whose negativity is activated by negative heads or phrases that are high in the clausal structure. We could then derive the syntactic scale by assuming that clausal negation and n-words in adverbial and subject positions are always sufficiently high in the structure to activate the negativity feature. In contrast, n-words in direct object position could optionally remain VP-internal, in which case the negativity feature would not be activated; or they could optionally move to the left periphery, in which case they would activate the negativity feature. A possibility that does not involve optional movement on the part of the direct objects is to assume that the higher a negative head or phrase is in the structure, the more likely it is that it activates the negativity feature. Such an approach would predict further distinctions on the syntactic scale, e.g., between adverbs and subjects, possibly with the adverb more negative than the subject, as the RED context suggests in our experiment.

Further empirical investigation is needed before these suggestions can be made concrete and before we can distinguish between various accounts, be they processing accounts or competence-driven syntactic accounts or mixtures thereof. An important question that needs to be addressed in further work concerns the extent to which the results we obtained here with q-tags can be replicated with other diagnostics for sentential negativity. The accounts we have sketched above rely solely on the properties of the anchor and therefore we would expect other diagnostics to lead to parallel results. If this expectation is not borne out, the account of the data obtained here has to involve the tag question formation process as well.

The status of DE-items has to be investigated in more detail too. If the contrast we found is confirmed, purely syntactic accounts of the phenomena discussed here become more problematic since such accounts would need independent evidence for a DE functional projection and a DE syntactic feature along the lines of the above clausal

---

[10]We are indebted to an anonymous reviewer for emphasizing this point.

polarity projection and associated negativity feature.

Turning now to the semantic aspects of our findings, the generalizations we arrived at are relevant for various existing treatments of the relation between n-words and negation. One popular approach is to treat n-words like *nobody, nothing*, and *never* as indefinite expressions within the scope of a negative operator (see Laka 1990; Ladusaw 1992, 1995; Brown 1999; Zeijlstra 2004; Kratzer 2005; Penka 2012; Homer 2012 for different approaches in this vein). Under this approach, often referred to as the N(egative) I(ndefinite) approach, there is a simple connection between negative sentences and sentences involving n-words: they both involve a negative operator. A simple characterization of negative clauses could then connect negativity to the presence of the negative operator. Such an approach explains the strong negativity of n-words but it would need to be further elaborated in order to account for the difference in negativity between n-words in subject and adverb position versus n-words in object position. The reduced but still significant effect of DE-items on the negativity of a sentence would also stand in need of explanation.

In de Swart and Sag (2002), n-words and sentential negation are treated as involving the same polyadic negative quantifier. Connecting the negativity of a sentence to the presence of this negative quantifier gives us essentially the same predictions and therefore the same problems as the simple NI approach sketched above.

A different approach to n-words, originating in Barwise and Cooper (1981), treats them as negative quantifiers (NQs). An NQ approach to n-words accounts directly for the strong negativity of sentences with n-words and the weaker negativity of sentences with DE-items under the assumption that the negativity of a sentence is sensitive to the strength of the negative items within it – and taking the additional anti-additive property of n-words to make them stronger negatives. Such an account on its own, however, does not explain the effect of the syntactic scale we found above.

A syntactic treatment of the NQ approach has been proposed in Haegeman and Zanutinni (1991); Haegeman (1995); Beghelli (1995); Laka (1990), among others. In these analyses, NQs move at LF to the locus of sentential negation in order to check negative features. On this account, then, just as in the NI approach, sentences with n-words have an active negative node, which can be linked to their high degree of negativity. But just as the other approaches sketched above, this view has to be further enriched in order to account for the effects of the syntactic scale and for the negativity effect of DE-items.

We conclude that there is no approach in the existing literature that directly predicts the pattern we found. As mentioned above, before taking steps to remedy this situation further empirical work is needed to ascertain whether the behavior of *rarely* and *few* generalizes to other DE-items as well as to determine what precisely lies behind the syntactic scale we found. Finally, one has to determine whether our results are confirmed by other negativity diagnostics. For the time being we conclude, based on the data we have, that the polarity of a REV question tag is determined by a complex negativity calculus that takes into account both semantic and syntactic considerations. This result will, we expect, influence the further development of theoretical accounts of negativity and its role in determining the form of the question tags.[11]

---

[11]An anonymous reviewer notes that some of the burden placed here on theories of sentential negativity may actually be taken care of by mechanisms involved in the formation of tag-questions (e.g., ellipsis). Whether this strategy is suitable depends on whether the results obtained here can be replicated using other ways of diagnosing sentential negativity, not involving tag-questions. Further experimental

5 Conclusion

We now return to the basic questions we started out with, namely what factors influence the negativity of a sentence as reflected by the q-tag test, and what is their relative strength. In particular, we were interested in whether both n-words and DE-items contribute negativity and whether they do so in equal measure. We were also interested in finding out whether the syntactic position of a negative element influences the strength of the negativity it induces. Finally, if both parameters (type of negative element and syntactic position) would turn out to be relevant, we were interested in the interaction between them.

The experimental results reported here lead to the following series of generalizations. First, our results show that both n-words and DE-items have a significant impact on the negativity of the sentence they occur in. Abstracting away from other factors, sentences involving such items pattern more like ordinary negative sentences with respect to q-tags than like ordinary positive sentences.

Second, our results indicate that n-words contribute more negativity than DE-items. That is, sentences involving n-words behave more like ordinary negative sentences with respect to q-tags than sentences involving DE-items.

Third, our results point to the conclusion that the strength of negativity induced by both n-words and DE-items is sensitive to the same syntactic factors. In particular, subjects and adverbs are stronger negative contributors than direct objects: a negative expression in subject or adverb position renders the sentence it occurs in more negative with respect to the q-tag test than the same expression in direct object position, independently of whether the expression in question is an n-word or a DE-item. Thus, we found evidence for the hypothesis that the negativity of a sentence is sensitive to the two scales repeated below:

(20)   a.   n-words ≫ DE-items   SEMANTIC SCALE
           b.   ADV, SUBJ ≫ OBJ   SYNTACTIC SCALE

Our results also indicate that the semantic nature of the source of negativity (n-word vs. DE-item) is primary relative to the syntactic dimension it interacts with. Even the least negative n-words, i.e., n-words in direct object position, are more negative than the most negative DE-items, i.e., DE-items in adverbial or subject position.

These generalizations do not follow directly from any theoretical approach we know of. This is not to say, of course, that existing approaches cannot be modified and/or extended to account for them. But before we embark on the task of accounting for these generalizations more empirical work is needed to test and refine them. In our view, the most urgent open questions are following:

(i) Is it possible to replicate the results we obtained here using other ways of diagnosing sentential negativity, not involving tag-questions? Addressing this issue is essential in order to determine whether the results we obtained here really tell us something about sentential negativity in general, or rather something more specifically about the interplay between sentential negativity and the formation of tag-questions.

(ii) Do the results we obtained here for the DE-items *few* and *rarely* generalize to other DE-items? Addressing this issue is relevant to establishing the nature of the semantic parameter involved.

---

work is needed to resolve this important issue.

(ii) On the syntactic side, there are two empirical issues that need to be resolved. First, our results indicate that the q-tag diagnostic does not distinguish between adverbs and subjects with respect to their effect on negativity. The question that arises is whether the results of other diagnostics still conflate adverbs and subjects or make finer grained distinctions.

(iii) Relatedly, further empirical work needs to be carried out to test whether word order plays a role in negativity effects and if it does, how it interacts with grammatical relations. One hypothesis consistent with our results is that what matters for negativity is not grammatical relation or hierarchical syntactic structure, but simply whether the negative item is in the post- or pre-verbal field. In fact, the processing account of the syntactic scale we outlined above is very much along these lines. Thus, further empirical work is needed to establish the factors underlying the syntactic scale we identified using the q-tag diagnostic. For example, we want to know whether preverbal direct objects or post verbal subjects behave differently than canonical cases we considered here with respect to negativity effects.

(iv) Finally, it is important to investigate to what extent the results we obtained here can be replicated in languages other than English, both using tag-questions (if available in the language) as well as other ways to diagnose sentential negativity. We hope that the work reported here will spur research on these empirical issues, which in turn will lead to progress on the theoretical front as well.

References

Barwise, J. and Cooper, R. (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy*, **4**(2), 159–219.

Beghelli, F. (1995). *The Phrase Structure of Quantifier Scope*. Ph.D. thesis, UCLA.

Brown, S. (1999). *The Syntax of Negation in Russian: A Minimalist Approach*. CSLI Publications, Stanford.

Cattell, R. (1973). Negative transportation and tag questions. *Language*, **49**, 612–639.

De Clercq, K. (2011). Squat, zero and no/nothing: Syntactic negation vs. Semantic negation. In M. Elenbaas and R. Nouwen, editors, *Linguistics in the Netherlands 2011*, Amsterdam. John Benjamins.

De Clercq, K., Haegeman, L., and Lohndal, T. (2012). Medial adjunct PPs in English: Implications for the syntax of sentential negation. *Nordic Journal of Linguistics*, **35**(1), 5–26.

de Swart, H. and Sag, I. (2002). Negation and Negative Concord in Romance. *Linguistics and Philosophy*, **25**(4), 373–417.

Farkas, D. and Roelofsen, F. (2012). Polar initiatives and polarity particle responses in an inquisitive discourse model. Presented at the Amsterdam Colloquium, December 2011, and at the Newcastle Workshop on the Syntax of Answers to Polar Questions, June 2012.

Gajewski, J. (2011). Licensing strong NPIs. *Natural Language Semantics*, **19**(2), 109–148.

Giannakidou, A. (2011). Negative and positive polarity items. In K. v. Heusinger, C. Maienborn, and P. Portner, editors, *Semantics: An International Handbook of Natural Language Meaning*, pages 1660–1712. Mouton de Gruyter, La Hague.

Haegeman, L. (1995). *The Syntax of Negation*. Cambridge University Press, Cambridge.

Haegeman, L. and Zanutinni, R. (1991). Negative heads and the neg criterion. *The Linguistic Review*, **8**, 233–51.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the 2nd Meeting of the North American Asssociation for Computational Linguistics*, pages 159–166.

Hale, J. (2011). What a rational parser would do. *Cognitive Science*, **35**, 399–443.

Hoeksema, J. (2012). On the natural history of negative polarity items. *Linguistic Analysis*, **38**, 3–33.

Homer, V. (2012). Domains of polarity items. *To appear in the Journal of Semantics*.

Horn, L. R. (1989). *A Natural History of Negation*. University of Chicago Press, Chicago/London.

Klima, E. (1964). Negation in English. In J. Fodor and J. Katz, editors, *The structure of language*, pages 246–32. Prentice Hall, New Jersey.

Kratzer, A. (2005). Indefinites and the operators they depend on: from Japanese to Salish'. In G. N. Carlson and F. J. Pelletier, editors, *Reference and Quantification: The Partee Effect*, pages 113–41. CSLI Publications, Stanford, CA.

Krifka, M. (1995). The semantics and pragmatics of polarity items. *Linguistic Analysis*, **25**, 1–49.

Kruschke, J. K. (2011). *Doing Bayesian Data Analysis: A Tutorial with R and BUGS*. Academic Press/Elsevier.

Ladusaw, W. (1979). *Polarity sensitivity as inherent scope relations*. Ph.D. thesis, University of Texas.

Ladusaw, W. A. (1992). Expressing negation. In C. Barker and D. Dowty, editors, *Proceedings of SALT II*, volume Ohio State Working Papers in Linguistics 40, pages 237–59. Ohio State University.

Ladusaw, W. A. (1995). Configurational expression of negation. In J. van der Does and J. van Eijck, editors, *Quantifiers, Logic and Language Logic and Language*, volume Logic and Language, pages 203–23. CSLI Publications, Stanford, CA.

Laka, I. (1990). *Negation in Syntax: On the Nature of Functional Categories and Projections*. Ph.D. thesis, MIT.

Levinson, D. (2008). *Licensing of negative polarity particles* yet, anymore, either *and* neither: *Combining downward monotonicity and assertivity*. Ph.D. thesis, Stanford.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, **106**, 1126–1177.

Malamud, S. and Stephenson, T. (2011). Three ways to avoid commitments: Three Ways to Avoid Commitments: Declarative Force Modifiers in the Conversational Scoreboard. In A. Ron, M. Core, D. DeVault, K. Georgila, E. Kaiser, and A. Stent, editors, *SemDial 2011: Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue.*, pages 74–83., Los Angeles, California.

McCawley, J. D. (1998). *The Syntactic Phenomena of English*. University of Chicago Press, Chicago/London, 2nd edition.

Moscati, V. (2006). *The Scope of Negation*. Ph.D. thesis, University Siena.

Payne, J. R. (1985). Negation. language typology and syntactic description. In T. Shopen, editor, *Clause structure*, pages 197–242. Cambridge University Press, Cambridge.

Penka, D. (2011). *Negative indefinites*. Oxford University Press, Oxford.

Penka, D. (2012). Negative Features on Negative Indefinites: Evidence from Split Scope. *Journal of Semantics*, **29**, 373–402.

Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. Longman, New York.

Swan, M. (2005). *Practical English Usage*. Oxford University Press, 3rd edition.

van der Wouden, T. (1994). *Negative Contexts*. Routledge, London.

Zeijlstra, H. (2004). *Sentential Negation and Negative Concord*. LOT.

Zwarts, F. (1995). Nonveridical contexts. . *Linguistic Analysis*, **25**(1), 286–32.

Zwarts, F. (1998). Three Types of Polarity. In F. Hamm and E. Hinrichs, editors, *Plurality and Quantification*, pages 177–238. Kluwer Academic Press.

A   Experimental items

We only list the first item in all 16 conditions. All the other items are listed only in the POSITIVE & REV condition.

(1)   a.Mary has just told Jane something about government representatives. Jane knows it already and says:

    i.Aha, I was right! The government representative visited the colonies this year, did he?
    ii.Aha, I was right! The government representative visited the colonies this year, didn't he?

b.Mary has just told Jane something about government representatives. Jane finds it hard to believe and says:

    i.You don't say! The government representative visited the colonies this year, did he?
    ii.You don't say! The government representative visited the colonies this year, didn't he?

c.Mary has just told Jane something about government representatives. Jane knows it already and says:

    i.Aha, I was right! The government representative did not visit the colonies this year, did he?
    ii.Aha, I was right! The government representative did not visit the colonies this year, didn't he?

d.Mary has just told Jane something about government representatives. Jane finds it hard to believe and says:

    i.You don't say! The government representative did not visit the colonies this year, did he?
    ii.You don't say! The government representative did not visit the colonies this year, didn't he?

e.Mary has just told Jane something about government representatives. Jane knows it already and says:

    i.Aha, I was right! No government representative visited the colonies this year, did he?
    ii.Aha, I was right! No government representative visited the colonies this year, didn't he?

f.Mary has just told Jane something about government representatives. Jane finds it hard to believe and says:

    i.You don't say! No government representative visited the colonies this year, didn't he?
    ii.You don't say! No government representative visited the colonies this year, didn't he?

g.Mary has just told Jane something about government representatives. Jane knows it already and says:

    i.Aha, I was right! Few government representatives visited the colonies this year, did they?
    ii.Aha, I was right! Few government representatives visited the colonies this year, didn't they?

h.Mary has just told Jane something about government representatives. Jane finds it hard to believe and says:

    i.You don't say! Few government representatives visited the colonies this year, did they?
    ii.You don't say! Few government representatives visited the colonies this year, didn't they?

i.Mary has just told Jane something about government representatives. Jane knows it already and says:

    i.Aha, I was right! The government representative visited no colonies this year, did he?
    ii.Aha, I was right! The government representative visited no colonies this year, didn't he?

j.Mary has just told Jane something about government representatives. Jane finds it hard to believe and says:

    i.You don't say! The government representative visited no colonies this year, did he?
    ii.You don't say! The government representative visited no colonies this year, didn't he?

k.Mary has just told Jane something about government representatives. Jane knows it already and says:

    i.Aha, I was right! The government representative visited few colonies this year, did he?
    ii.Aha, I was right! The government representative visited few colonies this year, didn't he?

l.Mary has just told Jane something about government representatives. Jane finds it hard to believe and says:

    i.You don't say! The government representative visited few colonies this year, did he?
    ii.You don't say! The government representative visited few colonies this year, didn't he?

m.Mary has just told Jane something about government representatives. Jane knows it already and says:

    i.Aha, I was right! The government representative never visited the colonies this year, did he?
    ii.Aha, I was right! The government representative never visited the colonies this year, didn't he?

n.Mary has just told Jane something about government representatives. Jane finds it hard to believe and says:

    i.You don't say! The government representative never visited the colonies this year, did he?
    ii.You don't say! The government representative never visited the colonies this year, didn't he?

o.Mary has just told Jane something about government representatives. Jane knows it already and says:

    i.Aha, I was right! The government representative rarely visited the colonies this year, did he?
    ii.Aha, I was right! The government representative rarely visited the colonies this year, didn't he?

p.Mary has just told Jane something about government representatives. Jane finds it hard to believe and says:

    i.You don't say! The government representative rarely visited the colonies this year, did he?
    ii.You don't say! The government representative rarely visited the colonies this year, didn't he?

(2)   Geoff has just told Ryan something about composers. Ryan knows it already and says:

   a.   Aha, I was right! The composer used the cello in his late period, did he?
   b.   Aha, I was right! The composer used the cello in his late period, didn't he?

(3)   Lindsey has just told Betty something about the anthropologists. Betty knows it already and says:

   a.   Aha, I was right! The anthropologists discussed the social concerns at the January conference, did they?
   b.   Aha, I was right! The anthropologists discussed the social concerns at the January conference, didn't they?

(4)   Josephine has just told Ward something about the hostages. Ward knows it already and says:

   a.   Aha, I was right! The hostages met the envoys today, did they?
   b.   Aha, I was right! The hostages met the envoys today, didn't they?

(5)   Rick has just told Ann something about the travel agents. Ann knows it already and says:

   a.   Aha, I was right! The travel agents visited the Greek islands last summer, did they?
   b.   Aha, I was right! The travel agents visited the Greek islands last summer, didn't they?

(6)   Brad has just told Tom something about the managers. Tom knows it already and says:

   a.   Aha, I was right! The managers revised the contracts this month, did they?
   b.   Aha, I was right! The managers revised the contracts this month, didn't they?

(7)   Lilly has just told Adrian something about the children. Adrian knows it already and says:

   a.   Aha, I was right! The children understood the science experiments this semester, did they?
   b.   Aha, I was right! The children understood the science experiments this semester, didn't they?

(8)   Bob has just told Justine something about the athletes. Justine knows it already and says:

   a.   Aha, I was right! The athletes skipped the practice games this month, did they?
   b.   Aha, I was right! The athletes skipped the practice games this month, didn't they?

(9)   Jerry has just told Robert something about the publishers. Robert knows it already and says:

   a.   Aha, I was right! The publishers approved the book proposals last year, did they?
   b.   Aha, I was right! The publishers approved the book proposals last year, didn't they?

(10)   Annie has just told Matthew something about the juveniles. Matthew knows it already and says:

   a.   Aha, I was right! The juveniles returned the questionnaires this week, did they?
   b.   Aha, I was right! The juveniles returned the questionnaires this week, didn't they?

(11)   Richard has just told Jim something about the participants. Jim knows it already and says:

   a.   Aha, I was right! The participants received the special awards at the summit, did they?
   b.   Aha, I was right! The participants received the special awards at the summit, didn't they?

(12)   Violet has just told George something about the miners. George knows it already and says:

   a.   Aha, I was right! The miners contacted the reporters this week, did they?
   b.   Aha, I was right! The miners contacted the reporters this week, didn't they?

(13)   Fred has just told Marc something about the Neanderthals. Fred knows it already and says:

   a.   Aha, I was right! The Neanderthals crossed the mountains in that period, did they?
   b.   Aha, I was right! The Neanderthals crossed the mountains in that period, didn't they?

(14)   Lena has just told Maggie something about the farmers. Maggie knows it already and says:

   a.   Aha, I was right! The farmers sold their crops in January, did they?
   b.   Aha, I was right! The farmers sold their crops in January, didn't they?

(15)   Edith has just told Sara something about the junior lawyers. Sara knows it already and says:

   a.   Aha, I was right! The junior lawyers wrote the briefs this month, did they?
   b.   Aha, I was right! The junior lawyers wrote the briefs this month, didn't they?

(16)   Derek has just told Jeremy something about the lawyers. Jeremy knows it already and says:

   a.   Aha, I was right! The lawyers ignored the objections in this case, did they?
   b.   Aha, I was right! The lawyers ignored the objections in this case, didn't they?

(17)   Linda has just told Gary something about the bars at the beachfront. Gary knows it already and says:

   a.   Aha, I was right! The bars at the beachfront attracted the American tourists this summer, did they?
   b.   Aha, I was right! The bars at the beachfront attracted the American tourists this summer, didn't they?

(18)   Ashley has just told Clara something about the brokers. Clara knows it already and says:

   a.   Aha, I was right! The brokers mentioned the obvious drawbacks this morning, did they?
   b.   Aha, I was right! The brokers mentioned the obvious drawbacks this morning, didn't they?

(19)   Robin has just told Tommy something about the candidates. Tommy knows it already and says:
a.   Aha, I was right! The candidates answered the questions during the last debate, did they?
b.   Aha, I was right! The candidates answered the questions during the last debate, didn't they?

(20)   Karen has just told Brigit something about the researchers. Brigit knows it already and says:
a.   Aha, I was right! The researchers interviewed the subjects this morning, did they?
b.   Aha, I was right! The researchers interviewed the subjects this morning, didn't they?

(21)   Michal has just told Mia something about the children. Mia knows it already and says:
a.   Aha, I was right! The children learned the difficult vocabulary items this week, did they?
b.   Aha, I was right! The children learned the difficult vocabulary items this week, didn't they?

(22)   Reiko has just told Will something about the authors. Will knows it already and says:
a.   Aha, I was right! The authors provided the crucial references during the meeting, did they?
b.   Aha, I was right! The authors provided the crucial references during the meeting, didn't they?

(23)   Rachel has just told Alexandra something about the member states. Alexandra knows it already and says:
a.   Aha, I was right! The member states announced the target dates at the summit, did they?
b.   Aha, I was right! The member states announced the target dates at the summit, didn't they?

(24)   Philippe has just told Ellen something about the debaters. Ellen knows it already and says:
a.   Aha, I was right! The debaters overlooked the national issues at the last event, did they?
b.   Aha, I was right! The debaters overlooked the national issues at the last event, didn't they?

(25)   Tina has just told Stephen something about the employees. Stephen knows it already and says:
a.   Aha, I was right! The employees received the benefits in this period, did they?
b.   Aha, I was right! The employees received the benefits in this period, didn't they?

(26)   Bert has just told Kim something about the villagers. Kim knows it already and says:
a.   Aha, I was right! The villagers respected the national traditions in the twenties, did they?
b.   Aha, I was right! The villagers respected the national traditions in the twenties, didn't they?

(27)   Sheila has just told Kendra something about the militants. Kendra knows it already and says:
a.   Aha, I was right! The militants attacked the American ships during that war, did they?
b.   Aha, I was right! The militants attacked the American ships during that war, didn't they?

(28)   Keanu has just told Maria something about bishops. Maria knows it already and says:
a.   Aha, I was right! The answers troubled the bishops in the sixteenth century, did they?
b.   Aha, I was right! The answers troubled the bishops in the sixteenth century, didn't they?

(29)   Angelina has just told Martin something about the caregivers. Martin knows it already and says:
a.   Aha, I was right! The caregivers comforted the children during the night, did they?
b.   Aha, I was right! The caregivers comforted the children during the night, didn't they?

(30)   Johan has just told Janna something about the chimps. Janna knows it already and says:
a.   Aha, I was right! The chimps ate the apples during the experiment, did they?
b.   Aha, I was right! The chimps ate the apples during the experiment, didn't they?

(31)   Paul has just told Leon something about the women. Leon knows it already and says:
a.   Aha, I was right! The women tried the products at the fair, did they?
b.   Aha, I was right! The women tried the products at the fair, didn't they?

(32)   Nora has just told Francine something about the visitors. Francine knows it already and says:
a.   Aha, I was right! The visitors attended the meetings this week, did they?
b.   Aha, I was right! The visitors attended the meetings this week, didn't they?

(33)   Peter has just told Matt something about the pianists. Matt knows it already and says:
a.   Aha, I was right! The pianists played the Beethoven sonatas at the competition, did they?
b.   Aha, I was right! The pianists played the Beethoven sonatas at the competition, didn't they?

(34)   Gwendolyn has just told David something about the bookstores. David knows it already and says:
a.   Aha, I was right! The bookstores sold the poetry collections at the book fair, did they?
b.   Aha, I was right! The bookstores sold the poetry collections at the book fair, didn't they?

(35)   Jorge has just told Andrew something about the critics. Andrew knows it already and says:
a.   Aha, I was right! The critics praised these movies at the festival, did they?
b.   Aha, I was right! The critics praised these movies at the festival, didn't they?

(36)   Ryan has just told Guido something about the published communications. Guido knows it already and says:

    a.   Aha, I was right! The published communications discussed the pressing problems this week, did they?
    b.   Aha, I was right! The published communications discussed the pressing problems this week, didn't they?

(37)   Megan has just told Ronald something about the volunteers. Ronald knows it already and says:

    a.   Aha, I was right! The volunteers performed the hard tasks this month, did they?
    b.   Aha, I was right! The volunteers performed the hard tasks this month, didn't they?

(38)   Sheila has just told Amelia something about the managers. Amelia knows it already and says:

    a.   Aha, I was right! The managers met the promising applicants this morning, did they?
    b.   Aha, I was right! The managers met the promising applicants this morning, didn't they?

(39)   Jennifer has just told Diane something about the doctors. Diane knows it already and says:

    a.   Aha, I was right! The doctors read the charts this morning, did they?
    b.   Aha, I was right! The doctors read the charts this morning, didn't they?

(40)   Danny has just told Larry something about the monks. Larry knows it already and says:

    a.   Aha, I was right! The monks said their prayers at noon, did they?
    b.   Aha, I was right! The monks said their prayers at noon, didn't they?

B   Statistical modeling details

B.1   Frequentist analysis

Our final mixed-effects logistic regression model is as follows. Fixed effects: COND (reference level: POSITIVE), CONX (reference level: RED) and their interaction. Random effects: subject random effects for the intercept and the CONX slope.

    The maximum likelihood estimates (MLEs) for this model are provided below:

RANDOM EFFECTS		std.dev.	corr.	
	INTERCEPT	1.74		
	CONX-REV	2.57	-0.74	
FIXED EFFECTS		estimate	std.error	p-value
	INTERCEPT	-0.15	0.22	0.496
	COND-DE-ADV	1.54	0.21	4×10^{-13}
	COND-DE-OBJ	0.48	0.20	0.018
	COND-DE-SUBJ	0.80	0.20	9×10^{-5}
	COND-NEGATIVE	3.13	0.26	$< 2 \times 10^{-16}$
	COND-N-WORD-ADV	2.88	0.26	$< 2 \times 10^{-16}$
	COND-N-WORD-OBJ	2.86	0.26	$< 2 \times 10^{-16}$
	COND-N-WORD-SUBJ	2.76	0.25	$< 2 \times 10^{-16}$
	CONX-REV	-3.88	0.42	$< 2 \times 10^{-16}$
	COND-DE-ADV : CONX-REV	1.84	0.41	6×10^{-6}
	COND-DE-OBJ : CONX-REV	1.14	0.42	0.006
	COND-DE-SUBJ : CONX-REV	2.33	0.40	8×10^{-9}
	COND-NEGATIVE : CONX-REV	2.35	0.45	2×10^{-7}
	COND-N-WORD-ADV : CONX-REV	2.83	0.45	3×10^{-10}
	COND-N-WORD-OBJ : CONX-REV	2.10	0.44	2×10^{-6}
	COND-N-WORD-SUBJ : CONX-REV	3.19	0.45	1×10^{-12}

B.2   Bayesian analysis

The priors for the fixed effects, i.e., for the intercept and the non-reference levels of COND, CONX and their interaction, are all independent normals $N(\mu = 0, \sigma^2 = 1000)$. These priors place most of their probability on the interval $(-65, 65)$, which is an extremely wide interval on the standard logit scale. Therefore, the priors contribute very little information and the posterior estimates will overwhelmingly reflect the data.

The prior for the random effects will be similarly vague / low information: we assume a bivariate normal distribution for the intercept and CONX-REV random effects with correlation ρ between the two random effects $N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & \rho\sigma\tau \\ \rho\sigma\tau & \tau^2 \end{bmatrix}\right)$. The priors for the intercept standard deviation σ and the CONX-REV standard deviation τ are independent uniforms $Unif(0, 100)$ and the prior for ρ is $Unif(-1, 1)$.

The table below provides the estimated means, standard deviations and 95% credible intervals (CRIs) for the posterior distributions of the random and fixed effects (all on the logit scale, i.e., the same scale as the MLEs above).[12]

RANDOM EFFECTS		mean	std.dev.	95% CRI
	σ	1.794	0.169	$(1.484, 2.149)$
	τ	2.655	0.227	$(2.243, 3.143)$
	ρ	-0.734	0.055	$(-0.828, -0.615)$

FIXED EFFECTS		mean	std.dev.	95% CRI
	INTERCEPT	-0.147	0.223	$(-0.577, 0.292)$
	COND-DE-ADV	1.548	0.219	$(1.127, 1.979)$
	COND-DE-OBJ	0.482	0.207	$(0.078, 0.888)$
	COND-DE-SUBJ	0.807	0.209	$(0.396, 1.226)$
	COND-NEGATIVE	3.152	0.268	$(2.630, 3.689)$
	COND-N-WORD-ADV	2.887	0.266	$(2.379, 3.425)$
	COND-N-WORD-OBJ	2.880	0.265	$(2.359, 3.394)$
	COND-N-WORD-SUBJ	2.775	0.253	$(2.290, 3.281)$
	CONX-REV	-3.929	0.431	$(-4.797, -3.125)$
	COND-DE-ADV : CONX-REV	1.882	0.407	$(1.110, 2.694)$
	COND-DE-OBJ : CONX-REV	1.171	0.413	$(0.376, 2.010)$
	COND-DE-SUBJ : CONX-REV	2.369	0.402	$(1.595, 3.167)$
	COND-NEGATIVE : CONX-REV	2.386	0.457	$(1.507, 3.312)$
	COND-N-WORD-ADV : CONX-REV	2.878	0.458	$(2.002, 3.804)$
	COND-N-WORD-OBJ : CONX-REV	2.141	0.452	$(1.291, 3.035)$
	COND-N-WORD-SUBJ : CONX-REV	3.244	0.454	$(2.380, 4.157)$

The median probabilities and 95% CRIs (also on prob. scale) plotted in Figure 3 above are as follows:

COND & CONX	median prob.	95% CRI
POSITIVE & REV	0.017	$(0.008, 0.032)$
DE-OBJ & REV	0.082	$(0.050, 0.128)$
DE-SUBJ & REV	0.289	$(0.206, 0.387)$
DE-ADV & REV	0.344	$(0.252, 0.451)$
N-WORD-OBJ & REV	0.720	$(0.621, 0.802)$
NEGATIVE & REV	0.812	$(0.727, 0.875)$
N-WORD-ADV & REV	0.844	$(0.772, 0.899)$
N-WORD-SUBJ & REV	0.875	$(0.812, 0.920)$
POSITIVE & RED	0.463	$(0.360, 0.573)$
DE-OBJ & RED	0.582	$(0.474, 0.686)$
DE-SUBJ & RED	0.659	$(0.555, 0.751)$
DE-ADV & RED	0.802	$(0.721, 0.866)$
N-WORD-OBJ & RED	0.939	$(0.900, 0.964)$
NEGATIVE & RED	0.953	$(0.921, 0.973)$
N-WORD-ADV & RED	0.939	$(0.901, 0.964)$
N-WORD-SUBJ & RED	0.932	$(0.892, 0.960)$

Finally, there is no clear evidence of inter-speaker variation, i.e., no evidence for grouping subjects into dialects / systematic response patterns. The pattern that we see for pretty much all the conditions is that there is a big group of subjects that patterns in a uniform way and a long tail of subjects patterning in somewhat different ways from that big group and from each other.

---

[12]Details of the MCMC estimation: 3 chains, 30,000 iterations per chain, 5,000 burn-in, 10 thinning, total 7,500 iterations saved.