

MODULARITY AND INTUITIONS IN FORMAL SEMANTICS: THE CASE OF POLARITY ITEMS*

Emmanuel Chemla, Vincent Homer, Daniel Rothschild

October 25, 2011

Abstract

Linguists often sharply distinguish the different modules that support linguistics competence, e.g., syntax, semantics, pragmatics. However, recent work has identified phenomena in syntax (polarity sensitivity) and pragmatics (implicatures), which seem to rely on semantic properties (monotonicity). We propose to investigate these phenomena and their connections as a window into the modularity of our linguistic knowledge. We conducted a series of experiments to gather the relevant syntactic, semantic and pragmatic judgments within a single paradigm. The comparison between these quantitative data leads us to four main results. (i) Our results support a departure from one element of the classical Gricean approach, thus helping to clarify and settle an empirical debate. This first outcome also confirms the soundness of the methodology, as the results align with standard contemporary accounts of scalar implicature. (ii) We confirm that the formal semantic notion of monotonicity underlies negative polarity item syntactic acceptability, but (iii) our results indicate that the notion needed is *perceived* monotonicity. We see results (ii) and (iii) as the main contribution of this study: (ii) provides an empirical interpretation and confirmation of one of the insights of the model-theoretic approach to semantics, while (iii) calls for an incremental, cognitive implementation of the current generalizations. (iv) Finally, our results do not indicate that the relationship between negative polarity item acceptability and monotonicity is mediated by pragmatic features related to scalar implicatures: this tells against elegant attempts to unify polarity sensitivity and scalar implicatures (pioneered by Krifka and Chierchia). These results illustrate a new methodology for integrating theoretically rigorous work in formal semantics with an experimentally-grounded cognitively-oriented view of linguistic competence.

Keywords: modularity; polarity; scalar implicatures; monotonicity; intuitions; experiment.

1 Architecture of Grammar

Language use is constituted by a variety of different abilities. Much research in linguistics and psychology concentrates on either linking different sorts of abilities, or, more often, distinguishing them. To give a hackneyed example, Chomsky (1959) argued that syntactic

*We wish to thank Isabelle Brunet, Anne-Caroline Fievet, Christophe Lalanne, Jérôme Sackur, Philippe Schlenker and Benjamin Spector. We benefitted greatly from the comments of two referees for *Linguistics and Philosophy*, particularly Michael Israel, and the associate editor, Chris Potts. This work was supported by a 'Euryi' grant from the European Science Foundation ("Presupposition: A Formal Pragmatic Approach") and by a travel grant from the ESF attributed to the Euro-Xprag network.

ill-formedness can be clearly distinguished from certain forms of semantic deviance, such as the category errors in “colorless green ideas.” Such distinctions lead many researchers to posit some degree of modularity between different sorts of linguistic processes.

Hence, linguists tend to make sharp distinctions between different domains, e.g., syntax, semantics, pragmatics. However, recent work in syntax and semantics has found surprising connections between syntactic properties and semantic ones, linking the two domains. Investigating these connections allows us to directly assess the modularity of our linguistic knowledge. We hope to show that by investigating the right phenomena with the right methods, one can address such ambitious questions of relevance to central debates in cognitive science. We conducted a series of experiments to assess the links between logical inferences, pragmatic reasoning, and syntactic judgments.

Two strands of influential work in pragmatics and syntax have correlated *logical* properties (*monotonicity* properties) of intra-sentential environments with both the syntactic acceptability of a class of terms (*negative polarity items*, henceforth NPIs) in those environments (Fauconnier 1975b, Ladusaw 1979) and the presence of pragmatic inferences (*scalar implicatures*) involving certain terms in those environments (Horn 1972, Grice 1967/1989).¹ It is not that surprising that pragmatic inferences are sensitive to logical, semantic properties such as monotonicity, since such inferences are usually thought to occur after the calculation of sentence meaning. However, it is surprising that grammaticality judgments should be sensitive to judgments of monotonicity, since most researchers assume that grammaticality is determined prior to (or, at least, independently of) the calculation of sentence meaning. These connections — between NPIs and monotonicity on the one hand, and between scalar implicatures and monotonicity on the other hand — have suggested to many, such as Chierchia (2004), that the role of the logical monotonicity property for both the syntactic NPI phenomenon and the pragmatic phenomenon might be the result of a single system.

We tested speakers’ judgments of inferences, both logical and pragmatic, and of grammaticality, to assess these connections. As expected, we found robust correlations between logical properties and pragmatic inferences, on the one hand, and logical properties and grammaticality, on the other hand. But our results also challenge some assumptions normally made about these generalizations. Before we explain how, we need to give some background about these two generalizations.

¹We label the acceptability of the negative polarity items a *syntactic* matter, because this is a standard, almost pre-theoretical assumption. However, part of the point of this study is to examine the lines between syntax and semantics. We discuss this issue in section 2.3.

2 Scalar implicatures and negative polarity items: the role of monotonicity

In this section, we introduce the basics about three phenomena we will be interested in: monotonicity (§2.1), scalar implicatures (§2.2) and negative polarity (§2.3). Section 2.4 summarizes the traditional links that have been established between the last two linguistic phenomena and monotonicity.

2.1 Monotonicity: preliminary definitions

By definition, a predicate appears in an *upward-entailing* environment if replacement of the predicate by a weaker one (one with a wider extension) produces a logically weaker sentence. For example, the schematic environment **John ate** — is upward-entailing. To test this claim, we first check that **smoked salmon** is a stronger predicate than **salmon** (if something is smoked salmon, it has to be salmon); and second that (1a), where the environment is filled with the strong predicate, entails (1b).

- (1) a. John ate smoked salmon.
b. John ate salmon.

Downward-entailingness is the reverse property. For instance, one can see that the environment **John didn't eat** — is downward-entailing because the entailment pattern is reversed: the weaker predicate **salmon** leads to the stronger sentence. Concretely, (2b) entails (2a).

- (2) a. John didn't eat smoked salmon.
b. John didn't eat salmon.

By checking the entailment patterns in (3) and (4), one can check that the restrictor of **every** is downward-entailing, while the nuclear scope of **every** is upward-entailing.

- (3) a. Every alien that ate smoked salmon is obese.
b. Every alien that ate salmon is obese.
(4) a. Every obese alien ate smoked salmon.
b. Every obese alien ate salmon.

These definitions capture general notions of positivity and negativity, as defined by the class of entailments licensed in a given environment. Environments which satisfy none of them, i.e. environments which are neither upward nor downward-entailing, are called *non-monotonic*. These notions are based on formal definitions, and they are of an all-or-nothing kind. It is helpful to note that these formal notions are not easily accessible to introspection, and some claim in particular that downward-entailingness is rather difficult to detect (see, e.g., Geurts and van der Slik 2005).

2.2 Scalar Implicatures (SIs)

Grice (1967/1989) developed the concept of a *conversational implicature* to explain certain inferences that audiences (in a speech-act situation) derive from the fact that the speaker chose certain words to express his thought. One class of implicatures, *scalar implicatures*, involves the use of terms that can be arranged on a scale by strength, e.g., <**some, all**>, <**or, and**> (Horn 1972). In many cases, uses of the weaker item in the scale can lead to the implication that the speaker does not believe the sentence formed by replacing the weaker item with the stronger item. For example the inference in (5) seems reasonable when the first sentence is asserted.

- (5) The alien studied some African languages.
 → The speaker does not believe: The alien studied all African languages.

When we assume that the speaker has full knowledge of the situation we can strengthen this inference to the inference that the speaker believes it is false that the sentence with the stronger scalar item is true, rather than merely lacking belief in the stronger sentence. So the following inference is also commonly found:

- (6) The alien studied some African languages.
 → The alien did not study all African languages.

If this inference was merely a result of an ambiguity in the word *some*, possibly meaning **some but not all**, we would expect the **not all** part of the meaning to appear in almost any environment in which *some* is used. However, one can show that this is not the case.

- (7) Every alien must learn some African languages.
 ↗ Every alien must learn some but not all African languages.

It is also not the case that the alternative sentence formed with the stronger item (**all**) systematically gives rise to a negative inference, which can be shown by the fact that the inference in (8) is not possible in normal circumstances.

- (8) Every alien that studied some African languages is obese.
 ↗ It's not the case that every alien that studied all African languages is obese.

Rather it is well-known that while scalar implicatures, such as the **some but not all** reading of **some** in (5) and (6), systematically arise in some environments, they do not arise in others.

The standard Gricean story easily explains the facts above. To see this, we can capitalize on the definitions introduced in §2.1. Monotonicity properties naturally relate explanations of the derivation of implicatures, because these properties determine which of the scalar term leads to the strongest sentence. Suppose for instance that a speaker uses a weak scalar term in an upward-monotonic environment. The hearer can now reason

as follows, as suggested by Grice (1967/1989) and Horn (1972): if the speaker had used the stronger scalar term, s/he would have said something stronger (by upward monotonicity). Since the speaker is trying to be maximally informative, s/he would have used the stronger scalar term if s/he had believed it to be true. So, we can conclude that s/he does not believe it to be true, which is the inference in (5). Assuming the speaker is well-informed, we can conclude that s/he believes it to be false, the inference in (6). Thus, the Gricean view makes the following prediction: the use of a weak scalar term in an upward-monotonic environment licenses the inference that the speaker does not believe that the sentence formed by replacing the weak scalar term with a stronger one is true.

In sentence (8), however, **some** appears in the restrictor of **every**, which is not upward-entailing but downward-entailing (see (3)). Hence, because the entailment relations are different, the reasoning rehearsed above cannot be replicated and so we do not derive the scalar implicature. When a strong term (e.g., **all**) is uttered in a downward-entailing environment, as in (9), we get what are called *indirect* scalar implicatures (as opposed to the *direct* implicature, discussed in (8) above).²

- (9) Every alien that studied all African languages is obese.
 → It's not the case that every alien that studied some African languages is obese.

The explanation of why indirect implicatures occur in downward-entailing environments is exactly parallel to the explanation of why direct implicatures occur in upward-entailing environments. In short: an alternative sentence with a weaker scalar term in a downward-entailing environment is stronger, hence, the implication that this alternative, if not uttered, is false.

In recent years there has been much discussion about the validity of the Gricean account of scalar implicatures. Some recent work has challenged the Gricean view, suggesting that the phenomena illustrated above may have a grammatical component (Chierchia 2004, Fox 2006, Chierchia et al. forthcoming). Other work, which we call *neo-Gricean*, has advanced modifications of the Gricean approach to make it more predictively adequate (Levinson 2000, Spector 2003, 2006, van Rooij and Schulz 2004). The present study is not directed to the debate between grammatical accounts and Gricean accounts of scalar implicatures.

Any account of scalar implicatures owes us an explanation of the relationship between monotonicity and the presence or absence of scalar implicatures. One important empirical issue that arises in this context is the question of what implicatures arise in non-monotonic environments. Consider, for example, this sentence which uses the scalar term **some**.

- (10) Exactly three aliens studied some African Languages.

²Not all authors distinguish between direct and indirect implicatures and some use different terminology; we do not take this terminology to bear theoretical weight.

Note that (10) is not entailed by the sentence found by replacing **some** with **all**:

(11) Exactly three aliens studied all African Languages.

Nonetheless it is plausible that a statement of (10) suggests that the speaker thinks (11) is false. A classical Gricean understanding of scalar implicatures does not naturally provide an account of this inference, as nothing in the Gricean story explains why an assertion of (10) should lead to the inference that a sentence that is neither stronger nor weaker, (11), is not believed by the speaker. However, some grammatical and neo-Gricean accounts of implicatures predict these sorts of implicatures.³ Although they were interested in a different question, Chemla and Spector (in press) tested cases of this sort and found that people derive this kind of implicature. This shows that some departure from pure Gricean view is necessary.

2.3 Negative Polarity Items

Negative polarity items are a diverse class of expressions found in many languages. In English, they include **any**, **ever**, **at all**, **a red cent**, **give a damn**, **the least bit**, etc. They are called *negative* polarity items (NPIs) because they are almost always acceptable when they appear under negation, as with **ever** in (12), but unacceptable otherwise, as in (13).

(12) John didn't ever eat.

(13) *John ever ate.

However, NPIs are not merely sensitive to negation, as they are sometimes grammatical without any negation, as with **ever** in (14).

(14) Everyone who ever ate this kind of food may develop problems.

Ladusaw (1979), building on previous work by Fauconnier (1975b, 1978), proposed that negative polarity items are licensed by operators which have in common a certain semantic property, downward-entailingness (in at least one of their arguments). According to the licensing condition he offered, NPIs are only acceptable in the (immediate) scope of downward-entailing expressions.⁴ A variant of this conceptualization — which we adopt in this article — retains the central role of downward monotonicity, but holds that an NPI

³Scalar implicatures in non-monotonic contexts are discussed in Spector (2007a), Chemla (2008), and Chierchia et al. (forthcoming).

⁴Downward-entailingness is a necessary condition for the licensing of all NPIs, weak (e.g., *any* and *ever*) and strict (e.g., *in years*, *punctual until*). The two categories differ in the width of their distribution: weak NPIs are acceptable in a proper superset of the environments where strict NPIs can appear. To account for the difference, some authors have proposed that strict NPIs require more strongly negative environments, and hence offered a more restrictive licensing condition, by adding further criteria to downward monotonicity (such as Anti-additivity, cf. Zwarts 1998). Throughout this study, we are only concerned with downward-entailingness.

is not licensed by an operator *per se* but by an intra-sentential environment, which has the property of supporting downward-entailing inferences, i.e. substitution from sets to subsets, in the position where the NPI appears (Zwarts 1996, Gajewski 2005, Homer 2010); DE expressions are necessary, but they are not directly responsible, for licensing.

This hypothesis can explain the pattern of acceptability of negative polarity items in the restrictor and nuclear scope of different quantifiers. As discussed above, **every** is downward-entailing in its restrictor but upward-entailing in its nuclear scope (see (3) and (4)); accordingly, the NPI **any** is acceptable in the restrictor of **every** but not in its nuclear scope as illustrated in (15) and (16).

(15) Every alien with any talent for music is obese.

(16) * Every obese alien has any talent for music.

By contrast, the quantifier **few** is downward-entailing in both its restrictor and its nuclear scope, and, correspondingly, licenses NPIs in both positions:

(17) Few aliens with any talent for music are obese.

(18) Few obese aliens have any talent for music.

There are many well-known problems with the use of DEness to characterize the licensing conditions of NPIs. Much of the literature on NPIs consists of modifications of or departures from DEness (e.g., Linebarger 1987, 1980, Heim 1984, Kadmon and Landman 1993, Israel 1996, Giannakidou 1999). Nonetheless the basic generalization that NPIs are licensed in DE environments is widely accepted.

Unlike with scalar implicatures, there is relatively little consensus on *why* negative polarity items are sensitive to downward monotonicity. Accounts that try to explain this, such as Kadmon and Landman (1993) and Krifka (1995), make use of lexical properties that go beyond truth-conditional meaning. For instance, Kadmon and Landman (1993) treat **any** as an existential determiner truth-conditionally identical to **some**, but that has two special features: 1) it pragmatically demands a wider domain of quantification, and 2) it requires that the domain widening pragmatic requirement results in a stronger sentence. These two stipulations result in the prediction that **any** is only acceptable in downward-entailing environments.⁵

Before going on, a comment is in order on why we label sentences like (13) ungrammatical. We are not claiming that licensing is, or is solely, determined at the syntactic level (a view rejected by Fauconnier 1975a, 1978, Israel 2011 and others). However, judgments of NPI acceptability are widely classified as syntactic in the literature. One reason for this

⁵Their account is designed only to handle **any**, but accounts that depend on lexical stipulations can be expanded to a wide-range of lexical items, perhaps the most extensive such account is that of Israel (1996, 2011).

is that if they instead reflect some sort of semantic deviance it is not clear what that deviance is, as the sentences like (13) seem to have clear meanings. So there is some reason to think that the badness of sentence (13) is not purely a matter of semantic deviance, and thus it is natural, as a first assumption to label it as a grammatical infelicity. Of course, the purpose of our experiments is to shed light on the exact nature of NPI licensing, and we by no means assume *a priori* it is part of some sort of syntactic module.

Chierchia (2006) claims more explicitly that the link between NPIs and monotonicity is due to the fact that both NPI judgments and scalar implicatures are subserved by the same linguistic processes. Such a hypothesis is strongly supported by a minimal explanation type of approach, because it re-uses whatever explains the connection between monotonicity and scalar implicatures to explain the connection between monotonicity and NPIs. However, very few tests have been carried out to prove the mediation of SIs in NPI licensing.

2.4 Monotonicity in Pragmatics and Syntax

We described two important and robust generalizations in pragmatics and syntax respectively:

- A. Direct scalar implicatures arise from upward-monotonic environments, while indirect scalar implicatures arise from downward-entailing environments.
- B. Negative polarity items are licensed in downward-entailing environments.

The goal of our study is to empirically investigate these connections between monotonicity properties and pragmatic (A) and grammatical (B) ones. In addition, we wanted to see if there is a possibility that there might be a connection between scalar implicatures and NPIs unmediated by monotonicity. In the next section, we introduce some general issues with such generalizations, which our experiments are intended to address.

We also mentioned the possibility that the relation between NPIs and scalar implicatures could be subserved by the same processes. This would lead to the following generalization, which our study will investigate as well:

- C. Negative polarity items are licensed in environments where scalar implicatures arise.⁶

⁶This statement does not determine whether direct or indirect scalar implicatures should matter. In fact, a closer look at Chierchia's recent ideas may suggest that the relevant phenomenon to compare NPIs to are so-called free-choice inferences, yet another type of inferences that can be treated within the system for scalar implicatures. We will be content at a first approximation to stay at this level of details and investigate whether there is any detectable direct link between scalar implicatures and NPI licensing.

3 Status of Semantic Generalizations

As illustrated above, semantic models often uncover generalizations connecting intuitive linguistic judgments (presence of scalar implicatures, acceptability of NPIs) with formal properties (monotonicity). Such models are welcome because they lead to objective, testable theories of natural phenomena. In this section, we would like to reflect further on the status of such generalizations.

NPI licensing does not appear to be an all-or-nothing affair. In the literature it is often noted that environments which are non-monotonic (neither upward-entailing nor downward-entailing) give rise to intermediate judgments of NPI acceptability (see Rothschild 2006 for discussion):

(19) ? Exactly one student ate anything.

(20) * Some student ate anything.

Such intermediate judgments are surprising if we hold a simple picture of NPIs, suggested by the term *licensing*, according to which NPIs simply have some syntactic feature that needs to get satisfied in some way. On this view, NPIs should be like other syntactic phenomena such as gender agreement which are meant mostly to show binary deviant/non-deviant judgments, rather than to admit a wide spectrum of judgments.⁷ However, if NPIs do not behave like this, we need to conceptualize their licensing in a more nuanced way. In particular if, as we suspect, NPI judgments are graded, we would like to discover the source of this gradability and a binary formal property such as monotonicity does not qualify as a solid candidate.

Linguistic semantics has its origins in the twentieth-century project of developing mathematically rigorous theories for describing artificial languages (Tarski 1944, Montague 1974). The formal tools used in this project allow semanticists to state precise abstract properties of sentence meaning. Hence, it is no wonder that semanticists have found a number of generalizations relating such abstract semantic properties to linguistic intuitions, e.g., other polarity items (strong NPIs, Positive Polarity Items), Kanazawa's (1994) generalization about donkey sentences, the grammaticality of strong and weak quantifiers in *there*-constructions. In order to examine linguistic semantics as a cognitive science, it is important to reflect on these generalizations to determine what sort of psychological mechanisms sustain the observed connections between logical properties and linguistic intuitions. There are various different positions one can have.

One view that is perhaps implicit in much work in formal syntax and semantics goes as follows: there is a modular system keeping track of formal properties like monotonicity. This system provides outputs (correlating to the formal properties) which eventually result in grammatical judgments about NPIs (and possibly other things), but to which

⁷Though, of course, not all syntactic deviance comes as an all-or-nothing matter.

other processes remain blind (e.g., explicit monotonicity judgments *per se*, see §2.1). On this view, it is natural to use formal notions as the underlying source of intuitive judgments, and this comes with no commitment for these formal notions to be accessible to introspection. However, the source of the gradation of the linguistic judgments would then have to be found in independent factors, yet to be determined.

Another, probably less common, view is that in fact there is no modular detection of formal properties. Rather various processes provide continuous subjective representations that are *related* to these formal properties. These subjective representations would then be the real basis of subjective grammatical judgments. In this view, the generalization relating intuition about NPI and monotonicity is taken very seriously. Yet, by denying that the relevant notion of monotonicity is the textbook formal notion, this view is not committed to a modular model of the linguistic system. It also provides a testable basis to explain the gradience of intuitive syntactic judgments: in short, a speaker should find an NPI felicitous in an environment E to the extent that this speaker would also report that this environment E is downward-entailing.

4 Our Study: the role of experimental data

Much of the current discussion of both negative polarity items and scalar implicatures is marred by disagreement over the data. As we have seen, in many crucial cases, such as those with non-monotonic environments, there appear to be graded rather than all-or-nothing judgments for both the naturalness of a scalar implicature and the acceptability of negative polarity items. A first contribution of our study is to provide judgments acquired from controlled experiments which could help decide between competing theories.

More importantly, we wanted to investigate the connections between monotonicity and both negative polarity item acceptability and scalar implicature derivation. We also wanted to see if there was any evidence for the hypothesis that the pragmatics of scalar implicature and the grammaticality of negative polarity items result from one syntactic/pragmatic system. We thus collected intuitions of grammaticality (NPI) and strength of inferences (scalar implicatures or monotonicity inferences) in one setting, from the same speakers and using methodologies that were as comparable as possible.

We hoped that collecting these judgments could allow for novel observations about these phenomena and their relations. Syntactic and semantic theories often do not need experimental work when judgments are relatively clear (i.e. we don't need an experiment to tell us that 'John eat food' is an ill-formed sentence). However, when the data is murky, the relevant judgments consistently hard to make by introspective methods or informal testing, experimental methods are needed. By collecting systematic responses from several participants, we get a set of quantitative data to which we can then apply statistical methods to try to find patterns not accessible from the armchair.

Furthermore, we wanted to investigate the role of *perceived* monotonicity, rather than actual monotonicity, as described in the literature. This goal also led us to collect actual judgments, even for notions which may appear to be uncontroversial in textbooks.

This type of quantitative data is, moreover, something that seems to be required by the questions we are asking for yet another reason. We know from basic intuitions about core cases that both NPIs and SIs are sensitive to monotonicity. But to collect evidence on whether they are subserved by the same system, we need to go beyond this and look in detail at how the different kinds of judgments relate. Our data allows us to do this in a way armchair intuitions do not.

In sum, we tested at once the derivation of scalar implicatures, acceptability of negative polarity items, and perceptions of monotonicity to investigate these phenomena and to compare them with one another.

5 Description of the experiments

We collected judgments from naive speakers of French about three phenomena in two closely related experiments. In this section, we provide the details of the experimental setup. The results are presented and discussed in the following sections.

5.1 Participants

45 native speakers of French aged between 19 and 43 years participated to the experiments (23 for the ‘Scope’ experiment, 22 for the ‘Restrictor’ experiment). They were mostly university students, no participant had any background in formal linguistics.

5.2 Design and stimuli

5.2.1 Two experiments to test a wide range of environments

The two experiments differed in the range of syntactic environments in which the phenomena were tested, but were otherwise methodologically identical. For the first experiment, we tested the phenomena in the scope of various quantifiers. We were thus able to test environments obtained from quantifiers with minimal semantic differences (e.g., **Exactly 42** vs. **Only 12**). For the second experiment, we tested the phenomena with fewer quantifiers but more systematically in both the scope and in the restrictor of these quantifiers. The totality of environments tested between these two experiments were as follows:

S-Experiment (only scopes): **Aucun** (*No*), **Certains** (*Some*), **Chaque** (*All*), **Exactement 42** (*Exactly 42*), **Seulement 12** (*Only 12*), **Moins de 12** (*Fewer than 12*), **Au plus 12** (*At most 12*).

R-Experiment (scopes and restrictors): *Aucun (No)*, *Certains (Some)*, *Exactement 42 (Exactly 42)*, *Chaque (All)*.

As described below, we constructed stimuli to test our three target phenomena in each of these environments. Participants were given a cover story mentioning an alien invasion. This simple story has several advantages: it gives some flexibility for the content of the stimuli and it justifies the repeated uses of sentences about groups of aliens, since aliens would quickly become the focus of attention in case of an invasion. In addition, since most subjects would not have much knowledge of alien culture, sentences on these subjects minimized the effect of background knowledge which could, otherwise, be a confounding factor in NPI licensing,⁸ or in inference judgments (see e.g., Evans et al. 1983). The subject matter also made the task superficially entertaining.

5.3 Procedure and task

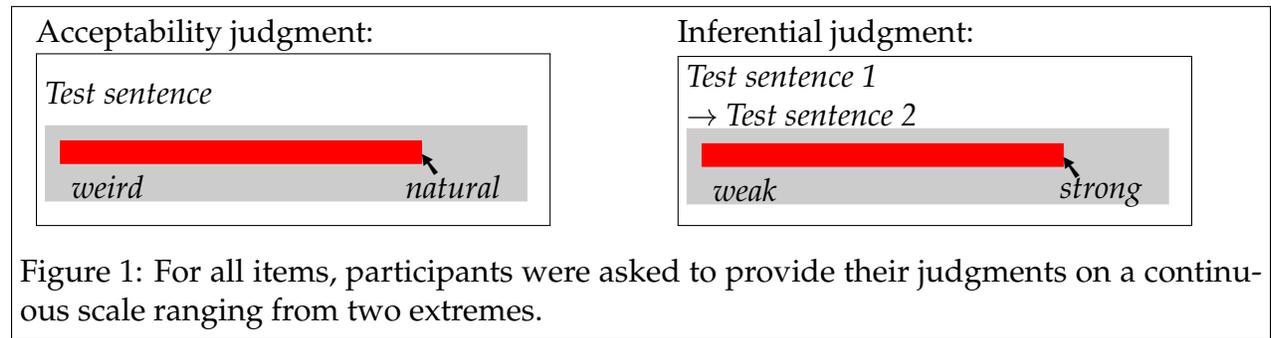
We tested 3 different phenomena: syntacticality of negative polarity items (NPIs), scalar implicatures (SI) and monotonicity inferences (MI). These three phenomena were tested in independent parts of each experiment. Upward- and downward-entailingness (UEness and DEness) were tested in the same MI block; direct and indirect scalar implicatures (dSI and iSI) were tested in the same SI block. The order of presentation of the three blocks was randomized, except that the NPI block was never the third (last) block. New instructions were given at the beginning of each new block, with a recommendation to take a short break.

We chose to prompt all the judgments on a continuous scale: participants were asked to provide their answer by setting the length of a red line at any point between two extreme anchors, which depend on the type of judgments concerned (see Fig. 1). This type of measure was chosen because it has proved to be successful to gather various types of subtle inferential judgments and it is close to other measures used in experimental studies of grammaticality (see Chemla 2009a,b for discussion).

For the NPI block, there were comparable versions of each example without and with an NPI (see examples (21-22)). The two versions were presented on the same screen in this order.⁹ Participants were asked to rate the naturalness of these sentences on a continuous scale ranging from ‘weird’ to ‘natural’. Examples were given to illustrate this grammaticality judgment task.

⁸Much theoretical research suggests that contextual information affects NPI licensing (e.g., Borkin 1971, Linebarger 1980, 1987, Fauconnier 1975a, 1978, Israel 1996, 2011).

⁹In a sense, the no-NPI sentence played the role of the so-called modulus in standard magnitude estimation task, already used for grammaticality judgments (see, e.g., Bard et al. 1996, Cowart 1997, Schütze 1996). The baseline role of this example, however, was implicit here which probably makes the task easier. See Sprouse (2011) for discussion.



For the inferences (SIs or MIs), participants were presented with one potential inference at a time. The instructions were the same in the two cases: participants were asked to answer along a similar continuous scale of judgments, although now one extreme corresponded to a rejection of the inference while the other corresponded to acceptance of the inference. Again, examples were given to explicate the task.

5.4 Material

In this section, we present examples of items that we used to test each phenomenon. A more extensive description of the experimental material is given in Appendix. All the items presented here are given in the same environment: the scope of the quantifier **Chaque** (*Every*). Other items were obtained by varying the environment as described in §5.2.1, and by varying the superficial content of the examples (6 contents for each phenomena). All these items were presented in random order within blocks.

5.4.1 Negative Polarity Items (NPI): experimental material

We used the French negative polarity item *le moindre* (literally *the least*).¹⁰ In this NPI part of the experiment, participants were prompted to give two judgments on each screen: one for a sentence without an NPI (this one always appeared on top of the screen), and one for a corresponding sentence with the NPI. Actual no-NPI and NPI examples were as follows:

- (21) no-NPI: Chaque alien qui a de l'intérêt pour la littérature est rouge.
Every alien who takes an interest in literature is red.
- (22) NPI: Chaque alien qui a le moindre intérêt pour la littérature est rouge.
Every alien who takes any interest in literature is red.

5.4.2 Scalar Implicatures (SI): experimental material

Direct scalar implicatures were tested using a sentence with the scalar item **plusieurs** (*several*) and asking whether (a paraphrase equivalent to) the negation of the alternative sentence with **tous** (*all*) instead of **plusieurs** followed. Participants were asked to consider that the first sentence was uttered, and to say whether they would endorse the second sentence. Here is an actual example:

- (23) "Chaque alien a étudié plusieurs langues africaines."
 → Certains n'ont pas étudié toutes les langues africaines.
"Every alien studied several African languages."
 → *Some did not study all the African languages.*

Indirect scalar implicatures were tested by giving a sentence with the scalar item **tous** (*all*) and asking subjects whether a paraphrase equivalent to the negation of the alter-

¹⁰The distribution of the French NPI *le moindre* seemed to us to be very similar to that of English NPI *any* (an item largely assumed to be sensitive to downward-entailingness). The experiment was, among other things, designed to put this intuitive judgment to the test. It did in fact confirm that *le moindre* is sensitive to DEness, despite its family resemblance with so-called minimizers (such as *the least bit*, *sleep a wink*, *lift a finger*, etc.), whose acceptability is sometimes claimed to be contingent on different conditions than that of *any* (see, e.g., Schmerling 1971, Borkin 1971, Heim 1984 Lahiri 1998, Guerzoni 2003). There is still a chance that *le moindre* is not a good representative of weak NPIs (the paradigm case of which is *any*) and may not serve to elucidate the licensing of its cross-linguistic counterparts, but, in any case, the existence a category of weak NPIs, stable across languages, has yet to be established. Given the high complexity of the landscape of polarity items, we do not intend to extrapolate our conclusions about *le moindre* to other NPIs within and outside of French. Another clarification is in order about the choice of *le moindre*: there is a well-known source of confound in the study of *any*, since alongside NPI *any*, English has a free choice item *any*, which can occur in non downward monotonic environments, e.g., in the scope of an existential modal. It is not clear to us that *le moindre* can be used as a free choice item in the first place; but to neutralize this potential confounding factor, in all our stimuli we used mass terms in the restrictor of *le moindre* (such as *interest*, *taste*, *sensibility*, etc.) which independently block free choice readings.

native sentence with an existential quantifier instead of **tous** followed from the original sentence:

- (24) “Chaque alien a étudié toutes les langues africaines.”
 → Certains n’ont pas étudié de langue africaine.
“Every alien studied all the African languages.”
 → *Some did not study African languages.*

5.4.3 Monotonicity Inferences (MI): experimental material

Monotonicity was tested using set/subset pairs like **dance/classical dance**. Participants, in different occasions spread through the MI block, were asked 1) whether a sentence formed with the first one entailed a similar sentence formed with the second one (downward-entailingness) or 2) whether the second one entailed the first one (upward-entailingness).

- (25) Set: Chaque alien a fait de la danse.
Every Alien practiced dance.
- (26) Subset: Chaque alien a fait de la danse classique.
Every Alien practiced classical dance.

MI items were presented in the same format as the SI items: the first sentence (the premise) appeared between quotation marks, while the second sentence (the conclusion) was preceded by a right arrow (see (23) and (24)).

6 Bare results

In this section, we present the mean ratings for each condition. The main conclusion we want to draw from these results is that the judgments we collected are intuitively reasonable and rich. This validates our methodology. In section 7, we will use these data to investigate our main purpose more directly and present analyses of the correlations between the different target phenomena.

6.1 Bare results for NPIs

The mean acceptability judgments we collected for NPIs in different configurations are reported in Fig. 2 (S-experiment) and Fig. 3 (R-experiment). We also report a composite measure which normalizes the NPI judgment with the judgments for the corresponding sentence without the NPI, schematically $\log(\frac{\text{NPI}}{\text{noNPI}})$. The idea is that the log of the quotient between these two measures would give an approximation of the particular contribution of the NPI to the acceptability of a sentence.

We do not intend to study these results in depth because we are mostly interested in the comparison between polarity phenomena and other phenomena. Yet, we would like to emphasize a couple of aspects of these first results.

(i) The control sentences without an NPI are overall judged pretty high and show less variability than the target sentences with an NPI (see the noNPI column in Figures 2 and 3). Furthermore, in both experiments again, NPIs are judged acceptable in the scope of clearly negative quantifiers (e.g., **No**), and not in positive contexts (e.g., in the scope of **Some** or **Every**). Although this generalization does not stand unchallenged, it seems that the clearest judgments on an introspective basis do come across in our participants' answers as well. This sensitivity suggests that responses were based on the right property of the sentences, i.e. the same property (acceptability) that leads linguists to claim that NPIs are acceptable in the scope of **No** and not in the scope of **Some** or **Every**.

(ii) In fact, some aspects of the acceptability ratings of NPI licensing may be surprising. For instance, it may seem that NPIs are only licensed in the scope of **No**, in which environment NPIs are judged much better than in all other environments. Notice however that there is no sharp line in the acceptability ratings across the licensing contexts; we cannot identify some set of contexts as licensing NPIs with the remainder not licensing them. In a setting like this, only differences are relevant (and as we discuss below there are finer differences to be discussed). Furthermore, this NPI data alone cannot speak to our general claim that NPIs are licensed in *perceived* DE environments. To evaluate this claim, we must assess the correlation between monotonicity judgments and NPI licensing. The differences found for NPI licensing, as surprising as they may seem, may be reflected in the monotonicity judgments. Such an outcome would support the idea that it is subjective perceptions of monotonicity that determine NPI-licensing rather than formal monotonicity.

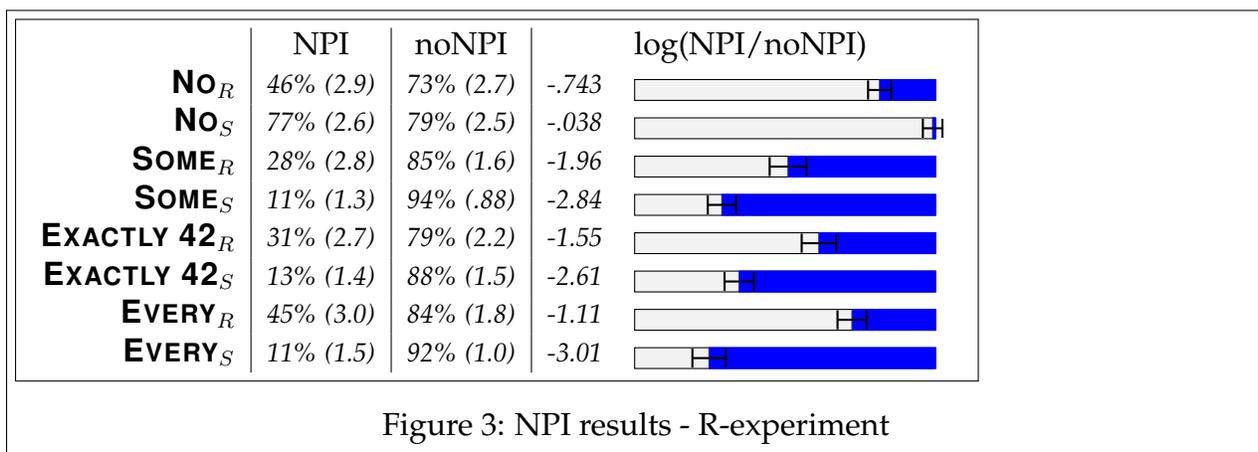
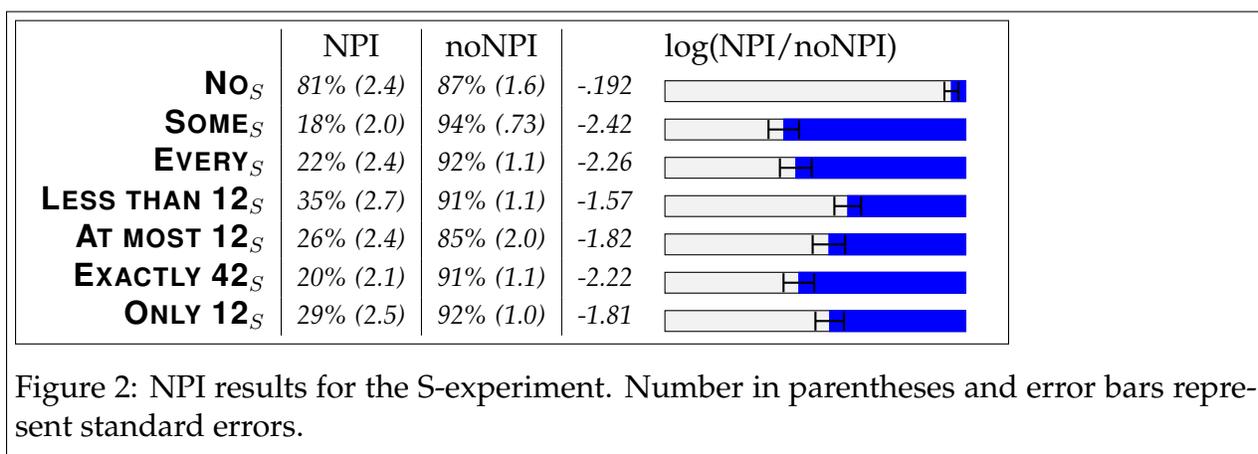
(iii) There are fine-grained differences in the judgments. For instance, one may see differences appearing in the scope of nearly synonymous environments (e.g., **Less than 12** and **At most 12** differ $M=35\%$ vs. $M=26\%$, $p = .013$).^{11,12} Here is another example. NPIs are judged more acceptable in the scope of **Only 12** than in the scope of **Exactly 42**: $M=29\%$ vs. $M=20\%$, $p = .003$. In that case, the difference remains statistically significant if we compare the normalized $\log(\text{NPI}/\text{noNPI})$ scores ($M=-.65$ vs. $-.45$, $p = .007$).¹³ We

¹¹Unless otherwise noted, p -values correspond to non-parametric bootstrap tests (2-tailed) computed from 5000 random samples. We systematically computed parametric t -tests and obtained entirely similar results.

¹²Notice however that a similar difference is found between the corresponding controls without an NPI. Consequently, the difference ceases to be significant when we compare the normalized $\log(\text{NPI}/\text{noNPI})$ score: $M = -1.8$ vs. -1.6 , $p = .14$. NB: Similarly, a 2-way ANOVA would reveal no significant interaction of the factors environments (**At most** vs. **Less than**) and presence/absence of the NPI.

¹³Similar results are obtained from a 2-way ANOVA: the interaction between quantifier (**Exactly 42** and **Only 12**) and presence/absence of NPI is significant ($F(1, 22) = 6.32$, $p < .05$, $\eta^2 = .183$).

do not want to scrutinize such details at length because, again, our focus is different but such results may prove to be of interest in the long run. The message we would like to highlight is that there is significant variation between environments. This variation is not random, as it can be linked to experts' intuitions and existing models, and it is important for our main purpose: investigating how NPI judgments vary with judgments about other phenomena.

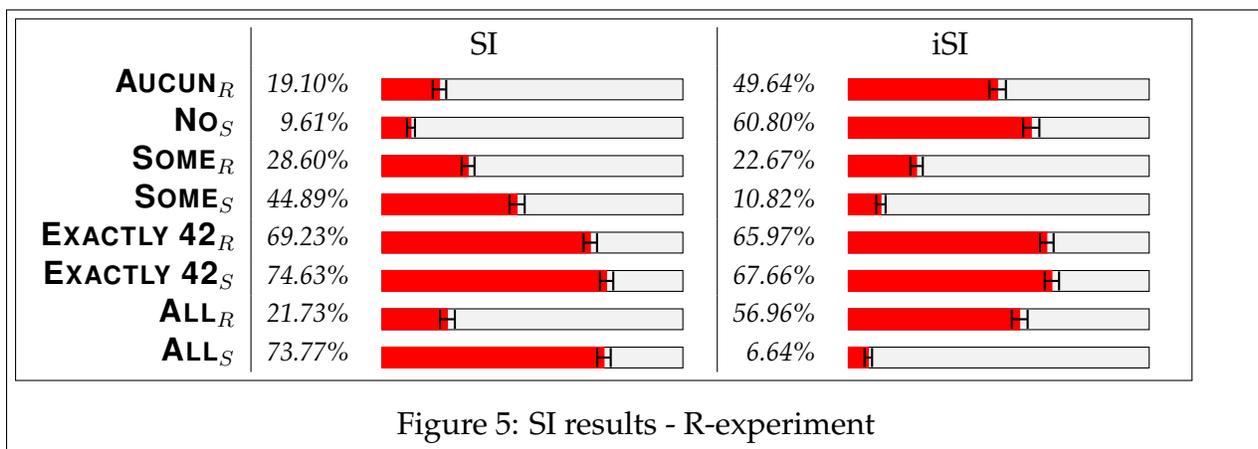
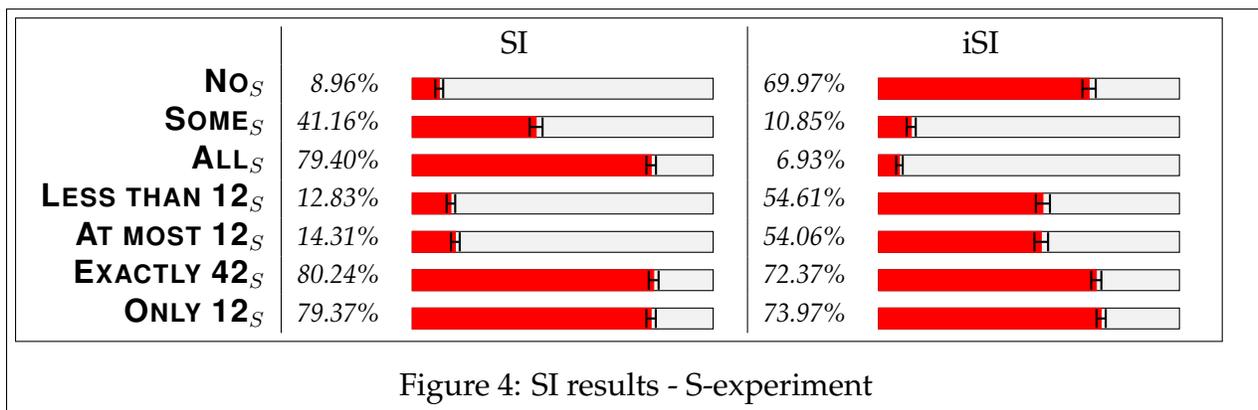


6.2 Bare results for SIs

The mean acceptability judgments we collected for SIs in different configurations are reported in Fig. 4 (S-experiment) and Fig. 5 (R-experiment). The generalization that seems to emerge from these results is that direct SIs arise in environments that are not DE, and indirect SIs arise in environments that are not UE.¹⁴ We will come back to this pattern

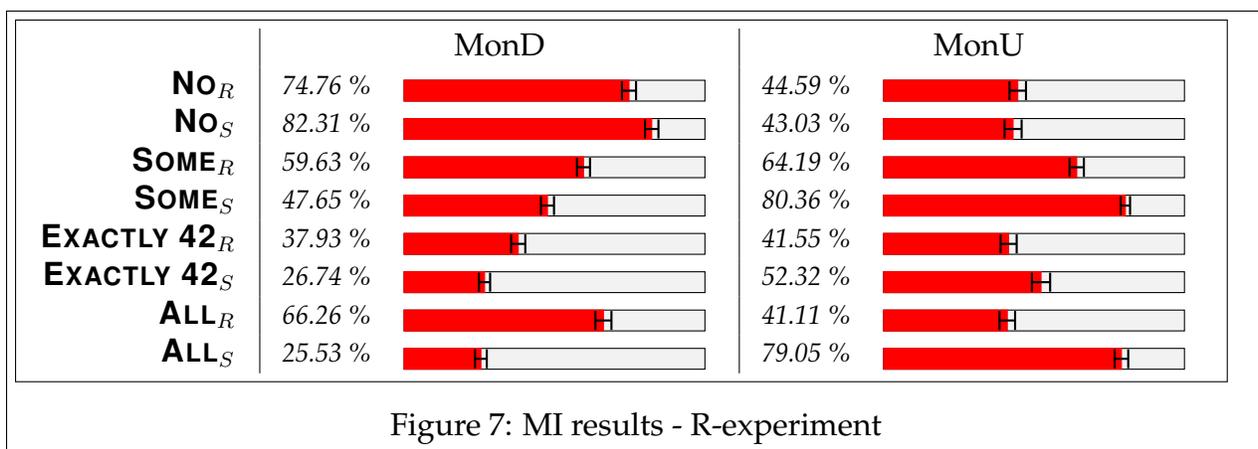
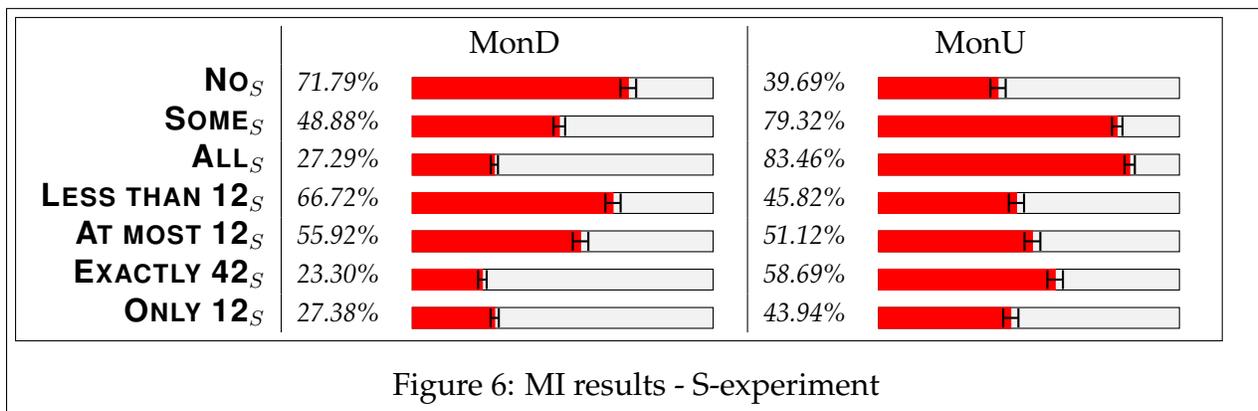
¹⁴The scope of the quantifier **Some** seems to be an exception, maybe because it is itself a scalar item.

of results when we explicitly compare SI judgments with judgments about monotonicity inferences.



6.3 Bare results for MIs

The mean acceptability judgments we collected for MIs in different configurations are reported in Fig. 6 (S-experiment) and Fig. 7 (R-experiment). Unsurprisingly, participants did not answer as would be prescribed in a formal semantics textbook and one can find variability between environments of the same monotonicity. Yet, overall, the results are in line with theoretical results and UE environments are judged UE more than DE environments ($M=78.0\%$ vs. $M=44.3\%$, $p < 1/5000$) and similarly DE environments are judged DE more than UE environments ($M=69.4\%$ vs. $M=41.2\%$, $p < 1/5000$). As before, the main conclusion we want to draw from these unanalyzed results is some confidence that the methodology captures the right type of judgments.



7 Relations between phenomena

We showed in the previous section that the judgments we collected for each phenomenon are both intuitively sound and interesting. In this section, we use these judgments to *compare* the different phenomena, e.g., the judgments of NPIs with the judgments about downward monotonicity. As we discussed earlier, formal semantic and pragmatic theories predict strong correlations between these judgments. In this section, we present our findings on the correlations between these different phenomena. A less technical summary of the main results can be found in the following section.

7.1 Scalar implicatures and monotonicity

In an upward-entailing environment, weak term (e.g., **some**) leads to a weaker statement than a stronger term (e.g., **all**). If logical strength (or logical weakness) is the source of scalar implicatures, direct SIs should arise in all and only UE environments. In downward-entailing environments, the entailment pattern is reversed, hence, indirect SIs should arise in all and only DE environments. To put it differently, the more a hearer

judges that an alternative is stronger than the utterance, the more she should derive a scalar implicature (see §2.2).

To investigate the plausibility of these Gricean ideas, we looked at the correlation between our participants' judgments about the monotonicity properties of an environment, and the robustness they attribute to an SI in this same environment.

Concretely, we computed the correlation coefficient between direct SIs and upward monotonicity judgments for each participant (across the different environments judged by this participant).¹⁵ We used these numbers as individual scores for the strength of the connection between the two phenomena. We were then in a position to compare the strength of different 'connections' in the population (we will most frequently report the mean of these percentage scores). With this method, we could assess which monotonicity judgments or combination of monotonicity judgments best predict the subjects' judgments about scalar implicatures. We will discuss our results in two parts: first we will discuss direct SIs and then we will discuss indirect SIs.

Before we start, a further statistical note is in order. The p -values we report below come from non-parametric bootstrap tests based on 5000 random samples, parametric t -tests yielded entirely similar results.

7.1.1 Direct SIs

As suggested above, the standard Gricean picture would predict that scalar implicatures are derived only in upward entailing environments. We found, however, that with respect to subjective monotonicity, the correlation between direct SIs and UEness is weaker than the correlation between SIs and DEness (M=16.1% vs. 38.7%, $p < 1/5000$). The directions of the correlations, of course, are different: environments are more likely to support scalar implicatures when they are UE, and more likely to support scalar implicatures when they are *not* DE.

The asymmetry between UEness and DEness is visible in Fig. 8 where one can see that the values for direct SIs are much better predicted when downward monotonicity is taken into account (sub-figure b) than when upward monotonicity is used as a predictor (sub-figure a).

7.1.2 Indirect SIs

The results for indirect SIs are parallel to that of direct SIs. On the pure Gricean view one would expect indirect SIs to be derived in downward-entailing environments. However,

¹⁵We used adjusted r^2 values. The same results were obtained using (i) uncorrected r^2 values, and (ii) non-parametric Kendall correlation coefficients whenever that was possible, i.e. when we investigated correlations with a single predictor.

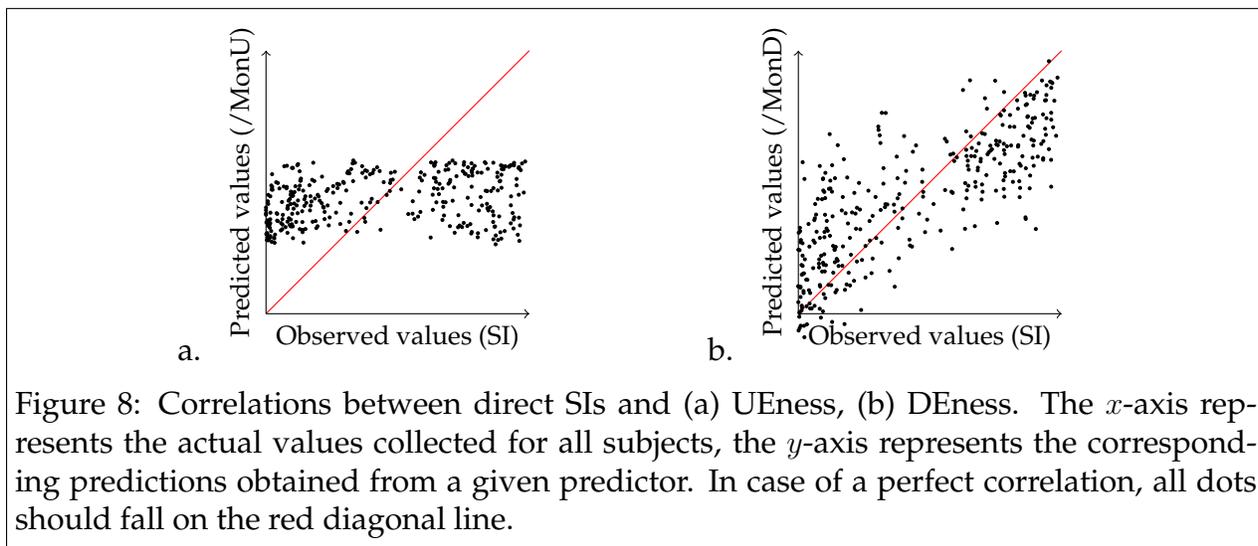


Figure 8: Correlations between direct SIs and (a) UEness, (b) DEness. The x -axis represents the actual values collected for all subjects, the y -axis represents the corresponding predictions obtained from a given predictor. In case of a perfect correlation, all dots should fall on the red diagonal line.

we found that the correlation between indirect SIs and DEness is weaker than the correlation between indirect SIs and UEness ($M=4.4\%$ vs. 15.9% , $p < 1/5000$). Fig. 9 offers a graphical representation of these results.

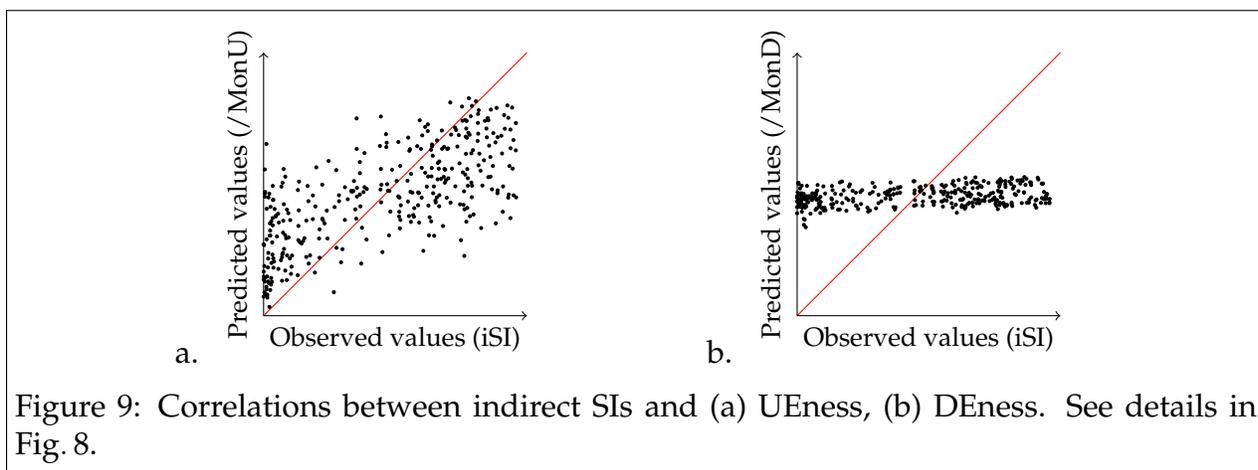


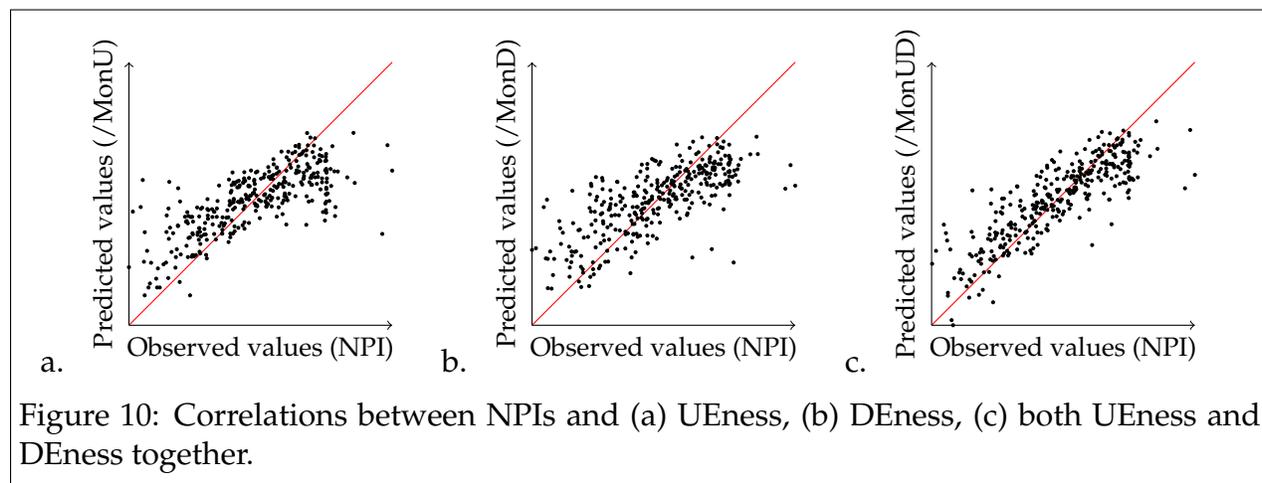
Figure 9: Correlations between indirect SIs and (a) UEness, (b) DEness. See details in Fig. 8.

7.2 NPIs and subjective monotonicity

The standard semantic generalization is that NPIs are licensed in DE environments. So we should expect NPI licensing to be correlated with each participant's own judgments of downward monotonicity.

We used the same methods as with scalar implicature to assess the extent to which NPI grammaticality was correlated with subjective monotonicity. Here we found that the correlation between NPIs and UEness does not differ significantly from the correlation

between NPIs and DEness ($M=23.2\%$ vs. $M=28.1\%$, $p = .24$). However, the combination of UEness and DEness provides yet a better prediction of NPI judgments than any of the two predictors alone ($M=45.8\%$ for the combination — combination vs. DEness alone: $p = .0012$, combination vs. UEness alone: $p = .0004$). These correlations are represented graphically in Fig. 10.

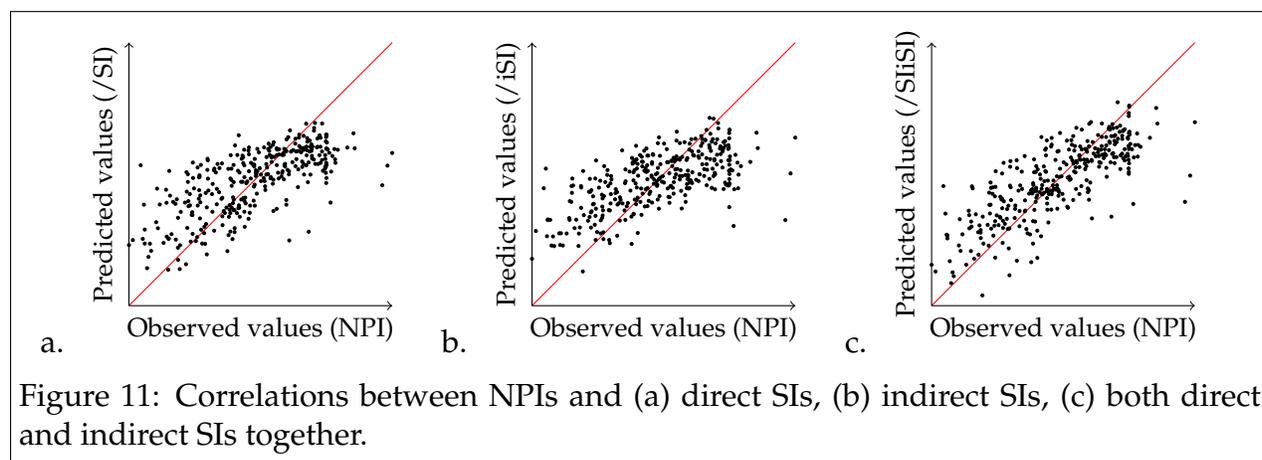


7.3 NPI: monotonicity or implicatures?

In addition to evaluating the traditional semantic generalizations about the degree to which monotonicity predicts SIs and NPIs, our data also allowed us to look at the relation between NPIs and SIs directly. Given that both are correlated with monotonicity we should also expect them to be correlated with each other. We found that the grammaticality of NPIs are better predicted by direct SIs than by indirect SIs ($M=22.4\%$ vs. $M=10.1\%$, $p=.014$). However, indirect SIs do add to the prediction already obtained with direct SIs: combination $M=35.7\%$ vs. direct SIs alone $M=22.4\%$, $p = .0006$ (less surprisingly given the difference above, direct SIs also add to the prediction obtained with indirect SIs: $p < 1/5000$).

As we have seen, the relations between NPIs and monotonicity, and between SIs and monotonicity are qualitatively different. This may provide preliminary qualitative reasons to doubt that the relation between NPIs and monotonicity is mediated by the same operations which play a role to derive SIs. There is some evidence, although it is rather thin (statistically), in favor of a stronger relation between NPIs and monotonicity than between NPIs and SIs. The facts are as follows.

If we consider the best predictor of each domain, we find that NPIs are better predicted by DEness than by direct SIs, although this difference is not quite significant ($p = .058$, it is significant at the traditional .05 level however if we use Kendall correlation coefficient:



28.1% vs. 20.3%, $p = .032$). If we take into account both predictors each time, we obtain that the prediction obtained from the combination of DEness and UEness is better than the one obtained from the combination of direct and indirect SIs, although not significantly so ($p = .078$).

Notice, however, that if the link between NPIs and SIs is implemented via a single operator at the basis of both types of judgments, there may be reason to believe that this operator is one that subserves either direct or indirect SIs, and there may thus be little grounds to take into account together direct and indirect SIs to predict NPIs. Hence, it is worth mentioning that the combination of monotonicity properties offer a better prediction of NPI judgments than the best predictor on the scalar implicature side (direct scalar implicatures): $p = .0002$.¹⁶

7.4 Formal monotonicity

In addition to looking at the extent to which our subjective measures of monotonicity were correlated with NPI and SI judgments, we were also interested in the question of how subjective monotonicity related to formal monotonicity.

For each participant x , we computed a new score which corresponds to the frequency of cases where the judgments of each of the *other* participants leads to a better predictor than the predictor obtained from x 's own judgments. We were then able to show that intra-subject subjective monotonicity is a better predictor than inter-subject subjective monotonicity of NPI judgments. (MonU: $M = 59\%$, $p = .019$; MonD: $M = 60\%$, $p = .012$; MonD*MonU: $M = 60\%$, $p = .021$).^{17,18}

¹⁶Note that there is no way to provide the same advantage to the SI story: the combination of SI predictors is not significantly better than the prediction by DEness alone: $p = .47$.

¹⁷For these tests, we use one-tailed p -values because there is little reason why we would find scores below 50%.

¹⁸We also computed similar statistics for scalar implicatures and obtain a tendency for direct SIs ($M =$

This analysis suggests that participants rely on their own judgments rather than on some shared external notion that would be distributed in the population. Imagine that formal monotonicity is the source of the judgments for NPIs. Then there would be no reason why predictors would work better intra-subject. If a participant x is a better monotonicity judge than another participant y , then x 's monotonicity judgments should be a better predictor of the NPI judgments for both x and y . To put it differently, the ranking of the quality of the monotonicity judgment should provide the goodness of the predictor, and so we would expect half of the participants' target NPI judgments to be better predicted by other participants' monotonicity judgments (for the participants with poor monotonicity judgments) and the opposite for the other half of the participants (the participants with the most reliable monotonicity judgments). Overall, the scores for the participants should fall at 50%. On the other hand, if a participant relies on her own judgments, then the fact that her monotonicity judgments are only poorly related to formal monotonicity will actually increase the discrepancy with the other participants' judgments, and it would increase the chances that her own monotonicity judgments constitute a more reliable predictor of her own NPI judgments.

8 Summary, discussion and conclusion

Our results provide a range of new evidence about the ways in which NPIs and SIs are handled in the grammar. Here we highlight the most significant results:

8.1 Scalar Implicatures

As we discussed in section 2.2, the classical Gricean view predicts that (direct) scalar implicatures are only found in UE environments, such as (27). This is because only in such environments does replacement of a weak scalar item (e.g., **some**) with a strong scalar (e.g., **all**) result in a stronger sentence. On the other hand, some *Neo-Gricean* approaches (as well as more radically different grammatical approaches) predict that SIs are also found in non-monotonic environments, such as (28).

(27) Someone/a few people/John did some of the work.

(28) Exactly two/only two students did some of the work.

Parallel issues arise for indirect scalar implicatures.

Our data provides a direct test of the question of whether SIs are derived in non-monotonic environments. An examination of the bare results show that subjects robustly derived direct and indirect SIs in the (formally) non-monotonic environments (see figure

57%, $p = .068$; $M = 58%$, $p = .043$; $M = 59%$, $p = .027$) and very little for indirect SIs ($M = 51%$, $p = .44$; $M = 47%$, $p = .25$; $M = 51%$, $p = .43$).

5). This result dovetails with previous experiments indicating that direct scalar implicatures are derived in some non-monotonic environments, rather than just in UE environments (Chemla and Spector in press).

More interestingly, our data allowed us to take a subjective view of this empirical dispute. Our results in section 7.1 demonstrated that SIs are better correlated with subjective judgments of DEness rather than UENess. This shows that the derivation of scalar implicatures is better predicted by perception of an alternative as being *not weaker* rather than perceiving of it as stronger. This is unexpected on a classical Gricean approach, but can be accounted for by some neo-Gricean approaches.

8.2 The basis of NPI judgments

Our experiments provide a wealth of data on how NPI judgments related to the other two categories of judgments. Given that the most widely put forward generalization in the literature is that NPIs are licensed in DE environments we might have expected that correlation to be the strongest. Instead we found that DEness and UENess were essentially equally good predictors of NPI grammaticality and the two together were better predictors than either individually.

The fact that DEness and UENess conspire to determine NPI grammaticality is surprising from a formal point of view because the two notions offer somewhat redundant pieces of information. Nonetheless, actual speakers may very well capitalize on both UENess and DEness to form their grammaticality judgments, because the two subjective notions may not be naturally *perceived* as exclusive or redundant.

8.3 Modularity of NPI Judgments

Semantic generalizations such as the claim that NPIs are licensed in DE environments are often taken to suggest a modular model of linguistic processing. The thought goes as follows: since DEness is not an easy property to judge intuitively we might expect that it is merely calculated at some cognitively inaccessible (syntactic) level of representation and that calculation forms the basis of NPI judgments.

The results above show that in fact *introspectively accessible* judgments of monotonicity are good predictors of judgments of NPI grammaticality. This remains compatible with the modular story above. Our results in section 7.4, however, cast doubt on the modular story. For if we suppose that inaccessible representations of formal properties determine NPI judgments, we should not expect an individual's own judgments of monotonicity to be particularly good at predicting *that individual's* grammaticality judgments. However, this is exactly what we found, indicating that subjective judgments of monotonicity are immediately related to judgments of NPI acceptability.

This is a novel result and one that raises exciting conceptual and methodological questions for formal semantics. If subjective judgments of inferences are good predictors of grammaticality judgments then the attempt within much current work on syntax and semantics to capture major generalizations by means of logic-like models may be excellent as a first approximation, but limited in the long run. There are really two issues here. The first is grammatical judgments for items like NPIs simply may not depend on formal properties. Second, it may be, as our results suggest that these judgments are *graded* and they may depend on similarly *graded* judgments about the relevant formal property. This is something current formal methods have been able to approximate, but lack the resources to fully capture. As we see it, our results confirm many of the conclusions of research in formal semantics, since the main claim that NPI judgments and monotonicity are correlated is strongly vindicated, but it also shows the potential limitations of that research.

This article is by no means the first to invoke a measure of *subjectivity* and speaker-dependence in the assessment of the acceptability of polarity items. Other researchers in the theoretical literatures have often invoked subjectivity — in the broader sense of context dependence — to either undermine or reject outright the role of downward monotonicity. Borkin (1971), who focuses on minimizers (e.g., *drink a drop*), underlines the variation among speakers and across utterance contexts, and identifies the intentions and so-called presuppositions of speakers in speech acts as the determining factors in the acceptance of polarity items. In a similar vein, both Fauconnier (1975a, 1978) and Israel (1996, 2011) argue that the meaning of a sentence is subjectively judged and construed, depending on the context of utterance, and cannot be reduced to its objective propositional content: the acceptability of polarity items provides a case in point, as it is, according to these authors, a function of the facilitating effects they have on certain inferences (within a scalar model). These authors therefore emphasize a pragmatic dimension to licensing: they argue that background knowledge and contextual information determines the licensing of NPIs, rather than just the (context-invariant) semantic properties of the sentence.

However, our study uses examples that do not naturally allow background assumptions to facilitate contextual entailments. We did this to try to minimize the extent that background knowledge would facilitate monotonicity judgments and NPI acceptability judgments. Rather we aimed to directly assess, for sentences with relatively little background, what inferences the subject find natural based on the particular quantifiers used. This allows us to measure the subjective monotonicity properties of quantifiers, and use these properties to predict NPI licensing. This sense of the subjectivity of NPI-licensing is compatible with other attempts to introduce subjectivity in NPI-licensing. To our knowledge, however, the present study is the first to explore subjectivity explicitly in this way and to offer the controlled empirical means to test it.

8.4 A Link Between NPIs and SIs

As we discussed in the introduction, authors since Chierchia and Krifka have tried to link judgments of NPI acceptability with the derivation of scalar implicatures. We tested some possible reconstruction of this effort,¹⁹ and our study did not vindicate this effort. Instead we have somewhat weak evidence for the conclusion that monotonicity judgments alone offer a direct predictor of NPI acceptability, a connection which does not seem to be mediated by SI judgments.

9 Conclusion

Formal semantics has its foundations in the mathematical tools developed to study artificial languages. The precision of these tools allowed linguists to state fine-grained, testable generalizations linking linguistic intuitions to abstract, logical properties of meaning. Such generalizations seem to presuppose a modular architecture, in which formal notions inaccessible to introspection form the basis of accessible linguistic intuitions. We proposed to test both the generalizations themselves and the implicit modular view that they suggest. To do so, we focussed on a famous phenomenon at the crossroad between syntax, semantics and pragmatics: *polarity sensitivity*. Our results provide a new form of confirmation of the precise predictions made by formal semanticists: for example, we found that negative polarity items are acceptable in downward-entailing environments and scalar implicatures arise from non-weaker alternatives. Our results provide quantitative empirical data that supports predictions made by semanticists on the basis of a combination of formal models and informal reports of intuitions. In this respect, we demonstrate the power and validity of the standard, qualitative methodology of formal semantics. However, we also tried to show what can be achieved by going beyond the standard methodology. We introduced a method of testing that allows a more fine-grained analysis of speaker's linguistic and *non-linguistic* judgments about linguistic stimuli than the dominant qualitative methods. Our method allowed us to pursue hypotheses not even statable within the standard framework of formal semantics, such as our claim that judgments of NPI-acceptability do not rely on *formal* downward-entailingness but rather on the speaker's perceived, subjective sense of downward-entailingness.

¹⁹For instance, we have not tested whether the acceptability of a given NPI (e.g., **any**) could be related to the robustness of implicatures associated with a corresponding scale (e.g., **<some, all>**) rather than numerals). See also footnote 6.

A Experimental material

This appendix provides the details of the experimental sentences we used. For each of the three phenomena (monotonicity inferences, negative polarity items, scalar implicatures), we used 6 different “contents” (i.e. non-essential, non-logical material made of various verbs and adjectives) in all the environments we wanted to test (scopes and restrictors of various quantifiers). In this appendix, we first give the full paradigm for one “content”, and then provide the other “contents” as they appear when tested in the scope of *every* (*chaque*). The full list of items can thus be obtained by filling in the paradigm of the first example with the “contents” given in the successive *every* examples (although a different final adjective was picked for each different content when embedded in the restrictor of a quantifier: **poilu** *hairy*, **déprimé** *depressed*, **sympathique** *pleasant*, **étourdi** *absent-minded*, **gluant** *gluey*, **rouge** *red*, **obèse** *obese*, **fluo** *fluorescent*, **filiforme** *thin, having a thread-like shape*, **souriant** *smiley*, **bicolore** *bicolor*).

A.1 Monotonicity items

Monotonicity items were constructed from pairs of sentences A and B, in which A contained a strong phrase (e.g., **smoked salmon**) and B a weaker one (**salmon**). Upward-monotonicity was tested as the $A \rightarrow B$ inference, downward-monotonicity was tested as the $B \rightarrow A$ inference. Below we report the list of A and B sentences.

- (29) a. Chaque alien a goûté du saumon fumé.
 Chaque alien a goûté du saumon.
Each alien tasted smoked salmon. / Each alien tasted salmon.
- b. Chaque alien qui a goûté du saumon fumé est poilu.
 Chaque alien qui a goûté du saumon est poilu.
Each alien who tasted smoked salmon is hairy. / Each alien who tasted salmon is hairy.
- c. Aucun alien n’a goûté de saumon fumé.
 Aucun alien n’a goûté de saumon.
No alien tasted smoked salmon. / No alien tasted salmon.
- d. Aucun alien qui a goûté du saumon fumé n’est poilu.
 Aucun alien qui a goûté du saumon n’est poilu.
No alien who tasted smoked salmon is hairy. / No alien who tasted salmon is hairy.
- e. Exactement 42 aliens ont goûté du saumon fumé.
 Exactement 42 aliens ont goûté du saumon.
Exactly 42 aliens tasted smoked salmon. / Exactly 42 aliens tasted salmon.
- f. Exactement 42 aliens qui ont goûté du saumon fumé sont poilus.
 Exactement 42 aliens qui ont goûté du saumon sont poilus.

- Exactly 42 aliens who tasted smoked salmon are hairy. / Exactly 42 aliens who tasted salmon are hairy.*
- g. Certains aliens ont goûté du saumon fumé.
Certains aliens ont goûté du saumon.
Some aliens tasted smoked salmon. / Some aliens tasted salmon.
- h. Certains aliens qui ont goûté du saumon fumé sont poilus.
Certains aliens qui ont goûté du saumon sont poilus.
Some aliens who tasted smoked salmon are hairy. / Some aliens who tasted salmon are hairy.
- i. Moins de 12 aliens ont goûté du saumon fumé.
Moins de 12 aliens ont goûté du saumon.
Less than 12 aliens tasted smoked salmon. / Less than 12 aliens tasted salmon.
- j. Au plus 12 aliens ont goûté du saumon fumé.
Au plus 12 aliens ont goûté du saumon.
At most 12 aliens tasted smoked salmon. / At most 12 aliens tasted salmon.
- k. Seulement 12 aliens ont goûté du saumon fumé.
Seulement 12 aliens ont goûté du saumon.
Only 12 aliens tasted smoked salmon. / Only 12 aliens tasted salmon.
- (30) Chaque alien a fait de la sculpture sur bois.
Chaque alien a fait de la sculpture.
Each alien practiced sculpture on wood. / Each alien practiced sculpture.
- (31) Chaque alien a commandé du vin blanc.
Chaque alien a commandé du vin.
Each alien ordered white wine. / Each alien ordered wine.
- (32) Chaque alien a joué de la guitare électrique.
Chaque alien a joué de la guitare.
Each alien played the electrical guitar. / Each alien played the guitar.
- (33) Chaque alien a bu de l'eau de mer.
Chaque alien a bu de l'eau.
Each alien drank seawater. / Each alien drank water.
- (34) Chaque alien a fait de la danse classique.
Chaque alien a fait de la danse.
Each alien practiced classical dance. / Each alien practiced dance.

A.2 Negative polarity items

The items in the NPI test always included a sentence without an NPI and then a minimally different sentence with an NPI (**le moindre**) added. We report these pairs of sentences below.

- (35) a. Chaque alien a de l'intérêt pour la littérature.
 Chaque alien a le moindre intérêt pour la littérature.
Each alien takes an interest in literature. / Each alien takes any interest in literature.
- b. Chaque alien qui a de l'intérêt pour la littérature est rouge.
 Chaque alien qui a le moindre intérêt pour la littérature est rouge.
Each alien who takes an interest in literature is red. / Each alien who takes any interest in literature is red.
- c. Aucun alien n'a de l'intérêt pour la littérature.
 Aucun alien n'a le moindre intérêt pour la littérature.
No alien has interest for literature. / No alien has any interest for literature.
- d. Aucun alien qui a de l'intérêt pour la littérature n'est rouge.
 Aucun alien qui a le moindre intérêt pour la littérature n'est rouge.
No alien who takes an interest in literature is red. / No alien who takes any interest in literature is red.
- e. Exactement 42 aliens ont de l'intérêt pour la littérature.
 Exactement 42 aliens ont le moindre intérêt pour la littérature.
Exactly 42 aliens takes an interest in literature. / Exactly 42 aliens takes any interest in literature.
- f. Exactement 42 aliens qui ont de l'intérêt pour la littérature sont rouges.
 Exactement 42 aliens qui ont le moindre intérêt pour la littérature sont rouges.
Exactly 42 aliens who take an interest in literature are red. / Exactly 42 aliens who take any interest in literature are red.
- g. Certains aliens ont de l'intérêt pour la littérature.
 Certains aliens ont le moindre intérêt pour la littérature.
Some aliens take an interest in literature. / Some aliens take any interest in literature.
- h. Certains aliens qui ont de l'intérêt pour la littérature sont rouges.
 Certains aliens qui ont le moindre intérêt pour la littérature sont rouges.
Some aliens who take an interest in literature are red. / Some aliens who take any interest in literature are red.
- i. Moins de 12 aliens ont de l'intérêt pour la littérature.
 Moins de 12 aliens ont le moindre intérêt pour la littérature.
Less than 12 aliens take an interest in literature. / Less than 12 aliens take any interest in literature.
- j. Au plus 12 aliens ont de l'intérêt pour la littérature.
 Au plus 12 aliens ont le moindre intérêt pour la littérature.
At most 12 aliens take an interest in literature. / At most 12 aliens take any interest in literature.
- k. Seulement 12 aliens ont de l'intérêt pour la littérature.
 Seulement 12 aliens ont le moindre intérêt pour la littérature.
Only 12 aliens take an interest in literature. / Only 12 aliens take any interest in literature.

- (36) Chaque alien a du talent pour la musique.
 Chaque alien a le moindre talent pour la musique.
Each alien has talent for music. / Each alien has talent for music.
- (37) Chaque alien a un penchant pour l'alcool.
 Chaque alien a le moindre penchant pour l'alcool.
Each alien has a taste for alcohol. / Each alien has any taste for alcohol.
- (38) Chaque alien a un sens artistique.
 Chaque alien a le moindre sens artistique.
Each alien has an artistic sensibility. / Each alien shows any artistic sensibility.
- (39) Chaque alien a fait un effort physique.
 Chaque alien a fait le moindre effort physique.
Each alien made a physical effort. / Each alien made any physical effort.
- (40) Chaque alien a pris du retard sur le projet.
 Chaque alien a pris le moindre retard sur le projet.
Each alien is showing delay with the project. / Each alien is showing any delay with the project.

A.3 Scalar implicatures items

Direct scalar implicatures were tested using a sentence with the scalar item **plusieurs** (*several*) and asking whether (a paraphrase equivalent to) the negation of the alternative sentence with **tous** (*all*) instead of **plusieurs** followed. Indirect scalar implicatures were tested in a similar manner using **tous** (*all*) as the scalar item (and, accordingly, the inference was a paraphrase of the negation of an alternative sentence which would contain an existential quantifier instead of the scalar item). We report the relevant sentence/inference below for both (i) direct and (ii) indirect scalar implicatures.

- (41) a. i. *direct*: "Chaque alien a visité plusieurs musées parisiens."
 → Certains n'ont pas visité tous les musées parisiens.
Each alien visited several Parisian museums. → Some did not visit all the Parisian museums.
- ii. *indirect*:
- iii. *indirect*: "Chaque alien a visité tous les musées parisiens."
 → Certains n'ont pas visité de musée parisien.
Each alien visited all the Parisian museums. → Some did not visit Parisian museums.
- b. i. *direct*: "Chaque alien qui a visité plusieurs musées parisiens est rouge."
 → Certains ont visité tous les musées parisiens et ne sont pas rouges.
Each alien who visited several Parisian museums is red. → Some visited all the Parisian museums and aren't red.
- ii. *indirect*:

- iii. *indirect*: “Chaque alien qui a visité tous les musées parisiens est rouge.”
 → Certains ont visité des musées parisiens (un ou plus) et ne sont pas rouges.
Each alien who visited all the Parisian museums are red. → Some visited Parisian museums (one or more) and aren't red.
- c. i. *direct*: “Aucun alien n’a visité plusieurs musées parisiens.”
 → Certains ont visité tous les musées parisiens.
No alien visited several Parisian museums. → Some visited all the Parisian museums.
- ii. *indirect*:
- iii. *indirect*: “Aucun alien n’a visité tous les musées parisiens.”
 → Certains ont visité des musées parisiens (un ou plus).
No alien visited all the Parisian museums. → Some visited Parisian museums (one or more).
- d. i. *direct*: “Aucun alien qui a visité plusieurs musées parisiens n’est rouge.”
 → Certains ont visité tous les musées parisiens et sont rouges.
No alien who visited several Parisian museums is red. → Some visited all the Parisian museums and are red.
- ii. *indirect*:
- iii. *indirect*: “Aucun alien qui a visité tous les musées parisiens n’est rouge.”
 → Certains ont visité des musées parisiens (un ou plus) et sont rouges.
No alien who visited all the Parisian museums is red. → Some visited Parisian museums (one or more) and are red.
- e. i. *direct*: “Exactement 42 aliens ont visité plusieurs musées parisiens.”
 → Parmi ces 42 aliens, certains n’ont pas visité tous les musées parisiens.
Exactly 42 aliens visited several Parisian museums. → Among these 42 aliens, some did not visit all the Parisian museums.
- ii. *indirect*: “Exactement 42 aliens ont visité tous les musées parisiens.”
 → A part ces 42 aliens, il y en a d’autres qui ont visité des musées parisiens (mais pas tous).
Exactly 42 aliens visited all the Parisian museums. → Except for these 42 aliens, other aliens visited Parisian museums (but not all).
- f. i. *direct*: “Exactement 42 aliens qui ont visité plusieurs musées parisiens sont rouges.”
 → Certains n’ont pas visité tous les musées parisiens et sont rouges.
Exactly 42 aliens who visited several Parisian museums are red. → Some did some did not visit all the Parisian museums and are red.
- ii. *indirect*: “Exactement 42 aliens qui ont visité tous les musées parisiens sont rouges.”

→ Il y a des aliens (d'autres) qui ont visité des musées parisiens et qui sont rouges.

Exactly 42 aliens who visited all the Parisian museums are red. → Some aliens (others) visited Parisian museums and are red.

- g. i. *direct*: "Certains aliens ont visité plusieurs musées parisiens."
→ Aucun n'a visité tous les musées parisiens.
Some aliens visited several Parisian museums. → None visited all the Parisian museums.
- ii. *indirect*:
- iii. *indirect*: "Certains aliens ont visité tous les musées parisiens."
→ Aucun n'a visité de musées parisiens.
Some aliens visited all the Parisian museums. → None visited Parisian museums.
- h. i. *direct*: "Certains aliens qui ont visité plusieurs musées parisiens sont rouges."
→ Aucun n'a visité tous les musées parisiens et est rouge.
Some aliens who visited several Parisian museums are red. → None visited all the Parisian museums and is red.
- ii. *indirect*:
- iii. *indirect*: "Certains aliens qui ont visité tous les musées parisiens sont rouges."
→ Aucun n'a visité des musées parisiens (un ou plus) et est rouge.
Some aliens who visited all the Parisian museums are red. → None visited Parisian museums (one or more) and is red.
- i. i. *direct*: "Moins de 12 aliens ont visité plusieurs musées parisiens."
→ Au moins 12 ont visité tous les musées parisiens.
Less than 12 aliens visited several Parisian museums. → At least 12 visited all the Parisian museums.
- ii. *indirect*: "Moins de 12 aliens ont visité tous les musées parisiens."
→ Au moins 12 ont visité des musées parisiens (au moins une partie des musées parisiens).
Less than 12 aliens visited all the Parisian museums. → At least 12 visited Parisian museums (at least some subset of Parisian museums).
- j. i. *direct*: "Au plus 12 aliens ont visité plusieurs musées parisiens."
→ Plus de 12 ont visité tous les musées parisiens.
At most 12 aliens visited several Parisian museums. → More than 12 visited all the Parisian museums.
- ii. *indirect*:
- iii. *indirect*: "Au plus 12 aliens ont visité tous les musées parisiens."
→ Plus de 12 ont visité des musées parisiens (au moins une partie des musées parisiens).

At most 12 aliens visited all the Parisian museums. → More than 12 visited Parisian museums (at least some subset of Parisian museums).

- k. i. *direct: “Seulement 12 aliens ont visité plusieurs musées parisiens.”*
 → *Parmi ces 12, certains n’ont pas visité tous les musées parisiens.*
Only 12 aliens visited several Parisian museums. → Among these 12, some did not visit all the Parisian museums.
- ii. *indirect:*
- iii. *indirect: “Seulement 12 aliens ont visité tous les musées parisiens.”*
 → *A part ces 12, certains ont visité (une partie) des musées parisiens.*
Only 12 aliens visited all the Parisian museums. → Except for these 12, some visited (a subset of) Parisian museums.

- (42) *“Chaque alien a rencontré plusieurs ministres français.”*
 → *Certains n’ont pas rencontré tous les ministres français.*
Each alien met with several French ministers. → Some did not meet all the French ministers.
- (43) *“Chaque alien a lu plusieurs romans de Balzac.”*
 → *Certains n’ont pas lu tous les romans de Balzac.*
Each alien read several novels by Balzac. → Some did not read all the novels by Balzac.
- (44) *“Chaque alien a étudié plusieurs langues africaines.”*
 → *Certains n’ont pas étudié toutes les langues africaines.*
Each alien studied several African languages. → Some did not study all the African languages.
- (45) *“Chaque alien a pratiqué plusieurs arts martiaux.”*
 → *Certains n’ont pas pratiqué tous les arts martiaux.*
Each alien practiced several Martial Arts. → Some did not practice all the Martial Arts.
- (46) *“Chaque alien a escaladé plusieurs volcans.”*
 → *Certains n’ont pas escaladé tous les volcans.*
Each alien climbed several volcanos. → Some did not climb all the volcanos.

References

- E.G. Bard, D. Robertson, and A. Sorace. Magnitude estimation of linguistic acceptability. *Language*, 72(1):32–68, 1996.
- Ann Borkin. Polarity Items in Questions. *Chicago Linguistics Society*, 7:53–62, 1971.
- Emmanuel Chemla. *Présuppositions et implicatures scalaires: études formelles et expérimentales*. PhD thesis, EHESS, 2008.
- Emmanuel Chemla. Presuppositions of quantified sentences: experimental data. *Natural Language Semantics*, 17(4):299–340, 2009a.

- Emmanuel Chemla. Universal Implicatures and Free Choice Effects: Experimental Data. *Semantics and Pragmatics*, 2(2):1–33, 2009b.
- Emmanuel Chemla and Benjamin Spector. Experimental evidence for embedded implicatures. *Journal of Semantics*, in press.
- G. Chierchia. Broaden your views: Implicatures of domain widening and the “logicality” of language. *Linguistic inquiry*, 37(4):535–590, 2006.
- Gennaro Chierchia. Scalar Implicatures, Polarity phenomenon, and the syntax/pragmatic interface. In Andrea Belletti, editor, *Structures and Beyond*. Oxford University Press, 2004.
- Gennaro Chierchia, Danny Fox, and Benjamin Spector. The Grammatical View of Scalar Implicatures and the Relationship between Semantics and Pragmatics. In Paul Portner, Claudia Maienborn, and Klaus von Heusinger, editors, *Handbook of Semantics*. Mouton, forthcoming.
- N. Chomsky. Syntactic Structures (The Hague: Mouton, 1957). *Review of verbal behavior by BF Skinner, Language*, 35:26–58, 1959.
- W. Cowart. *Experimental syntax: applying objective methods to sentence judgments*. Thousand Oaks, CA, 1997.
- J. Evans, J.L. Barston, and P. Pollard. On the conflict between logic and belief in syllogistic reasoning. *Memory & cognition*, 11(3):295–306, 1983.
- G. Fauconnier. Pragmatic scales and logical structure. *Linguistic Inquiry*, 6(3):353–375, 1975a.
- Giles Fauconnier. Polarity and the Scale Principle. *Chicago Linguistics Society*, 11:188–199, 1975b.
- Gilles Fauconnier. Implication Reversal in a Natural Language. In F. Guenther and S.J. Schmidt, editors, *Formal Semantics and Pragmatics for Natural Languages*, pages 289–301. Dordrecht: D. Reidel Publishing Company, 1978.
- Danny Fox. Free Choice and the theory of Scalar Implicatures. unpublished manuscript, 2006.
- Jon R. Gajewski. *Neg-Raising: Presupposition and Polarity*. PhD thesis, MIT, 2005.
- Bart Geurts and Frans van der Slik. Monotonicity and processing load. *Journal of semantics*, 22(1):97–117, 2005.
- Anastasia Giannakidou. Affective Dependencies. *Linguistics and Philosophy*, 22:367–421, 1999.
- Paul Grice. Logic and Conversation. In *Studies in the Ways of Words*. Harvard University Press, 1967/1989.
- Elena Guerzoni. *Why Even Ask?: On the Pragmatics of Questions and the Semantics of Answers*. PhD thesis, MIT, 2003.
- Irene Heim. A Note on Negative Polarity and Downward Entailingness. *Proceedings of NELS*, 14:98–107, 1984.
- Vincent Homer. Domains of Polarity Items. Ms., UCLA, 2010.

- Laurence Horn. *The Semantics of the Logical Operators in English*. PhD thesis, UCLA, 1972.
- Michael Israel. Polarity Sensitivity as Lexical Semantics. *Linguistics and Philosophy*, 19: 619–666, 1996.
- Michael Israel. *Grammar of Polarity: Pragmatics, Sensitivity and the Logic of Scales*. Cambridge University Press, 2011.
- Nirit Kadmon and Fred Landman. Any. *Linguistics and Philosophy*, 4.2:279–298, 1993.
- M. Kanazawa. Weak vs. strong readings of donkey sentences and monotonicity inference in a dynamic setting. *Linguistics and philosophy*, 17(2):109–158, 1994.
- Manfred Krifka. The semantics and pragmatics of polarity items. *Linguistic Analysis*, 25: 209–257, 1995.
- William Ladusaw. *Polarity Sensitivity as Inherent Scope Relations*. PhD thesis, University of Texas Austin, 1979.
- Utpal Lahiri. Focus and Negative Polarity in Hindi. *Natural Language Semantics*, 6:57–123, 1998.
- Stephen Levinson. *Presumptive Meanings: The Theory of Generalized Conversational Implicatures*. MIT, 2000.
- Marcia C Linebarger. *The Grammar of Negative Polarity*. PhD thesis, MIT, 1980.
- Marcia C Linebarger. Negative Polarity and Grammatical Representation. *Linguistics and Philosophy*, 10:325–87, 1987.
- Richard Montague. *Formal philosophy: selected papers of Richard Montague*. Yale University Press, 1974.
- Robert van Rooij and Katrin Schulz. Exhaustive Interpretation of Complex Sentences. *Journal of Logic, Language and Information*, 13(4):491–519, 2004.
- Daniel Rothschild. Non-Monotonic NPI-Licensing, Definite Descriptions and Grammaticalized Implicatures. In *Proceedings of SALT*, volume 16. CLC Publications, 2006.
- Susan F. Schmerling. A Note on Negative Polarity. In *Papers in Linguistics*, volume 4.1, pages 200–206. Linguistic Research Inc., Champaign, Ill., 1971.
- Carson T. Schütze. *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. University of Chicago Press, 1996.
- Benjamin Spector. Scalar implicatures: Exhaustivity and Gricean reasoning. In Balder ten Cate, editor, *Proceedings of the Eighth ESSLLI Student Session*, Vienna, Austria, 2003. Revised version in Spector (2007b).
- Benjamin Spector. *Aspects de la pragmatique des opérateurs logiques*. PhD thesis, Université Paris 7, Denis Diderot, 2006.
- Benjamin Spector. Aspects of the Pragmatics of Plural Morphology: On Higher-Order Implicatures. In Uli Sauerland and Penka Stateva, editors, *Presuppositions and Implicatures in Compositional Semantics*. Palgrave-Macmillan, 2007a.
- Benjamin Spector. Scalar implicatures: Exhaustivity and Gricean reasoning. In Maria Aloni, Paul Dekker, and Alastair Butler, editors, *Questions in Dynamic Semantics*, vol-

ume 17 of *Current Research in the Semantics/Pragmatics Interface*, pages 225–249. Elsevier, 2007b.

Jon Sprouse. A Test of the cognitive assumptions of magnitude estimation: The cognitive assumptions of Magnitude Estimation do not hold for acceptability judgments. *Language*, 87:274–288, 2011.

Alfred Tarski. The semantic concept of truth: and the foundations of semantics. *Philosophy and Phenomenological Research*, 4:13–47, 1944.

Frans Zwarts. Facets of Negation. In J. van der Does and J. van Eijck, editors, *Quantifiers, Logic, and Language*, pages 385–421. CSLI Publications, Stanford, 1996.

Frans Zwarts. Three Types of Polarity. *Plurality and Quantification*, 69:177–238, 1998.