Experimental investigations of licensing environments for NPIs in English

Jennifer Spenader,[1] Frank Richter[2] and Janina Radó[2]

Abstract: In an experimental investigation of NPIs and their licensing conditions in German, Richter & Radó (2013) showed that Weak Licensors can partially license Strong NPIs. In two experiments we investigate these results for English. In Experiment 1 we empirically identify 16 Strong and 16 Weak NPIs by examining their acceptability when licensed by sentential negation and weak adverb licensors and find results similar to German. In Experiment 2 we compare the licensing abilities of two different types of conditionals, normal, causal conditionals and rhetorical conditionals, such as `If you win, I'll eat my hat'. We did not find statistically significant evidence for our prediction that rhetorical conditionals might be able to license strong NPIs better than normal conditionals, but we discuss possible changes in the experimental design that might lead to different results. Experiment 2 thus serves as a pilot study for future investigations.

1. Introduction

The defining feature of Negative Polarity Items (NPIs) is their restriction to certain environments. Sentential negation has the ability to license all NPIs, hence the name, but the actual licensing properties are much more complex entailment properties. The discovery that the entailment properties of semantic operators have linguistic effects is one of the major triumphs of formal semantics, and explains why research on NPIs is a fundamental part of both formal semantics and linguistics proper. This research area has a rich history and an active present, with new questions and extended theories continually joining the discussion (see e.g. Eckardt, 2012 and Csipak et al. 2013). The seminal work of Frans Zwarts (1981) with his proposal for embedded licensing contexts has contributed

1 Corresponding author. j.spenader@gmail.com, Institute for Artificial Intelligence, University of Groningen.
[2] Institut für England- & Amerikastudien. Goethe Universität Frankfurt am Main frankmrichter@gmail.com, rado@em.uni-frankfurt.de. The second and third author's main contributions were extensive advice on the experimental design and general discussion.

greatly to our understanding of these expressions and the logical properties that they interact with.

Of great interest is the fact that even though licensing properties are extremely abstract, speakers seem to have internalized their properties and can use them implicitly to judge the acceptability of different types of NPIs in different licensing contexts. In this way research in NPIs actually is research into the question which logical properties are relevant for natural language interpretation.

Even though work on NPIs and their licensors is vast, one line of research is only just beginning. We have little experimental confirmation of many of the intuitions frequently discussed in the theoretical literature. A recent exception to this is Richter & Radó (2013) with four experiments testing the acceptability of German NPIs. We lack similar work for English, and this current paper hopes to make a modest start at filling this gap.

The current paper follows the method of Richter & Radó (2014) by first empirically verifying a set of NPIs as (relatively) weak or strong by testing their acceptability under licensors with different entailment properties. We compared weak adverbial licensors like *hardly* and *rarely* to sentential negation, a strong licensor. Based on these results we obtained a set of 16 weak and 16 strong NPIs. Experiment 2 then tests the ability of conditionals to license strong NPIs. Two types of conditionals were used: Normal Conditionals (or causal conditionals) of the types "*If you win the lottery, you can buy a new car"* compared to Rhetorical Conditionals of the type "*If you win the lottery, I'll eat my hat".* Rhetorical conditionals suggest that the antecedent of the conditional will not come true, coming close to an actual negation of future possibility. This similarity to negation explains our intuition that they may be able to license strong NPIs better than normal conditionals can. Unfortunately the results for Experiment 2 did not find a statistically significant difference between the two types of conditionals as licensors. This could be due to the great deal of variation in the ratings and the small number of target items per condition tested with each participant. Specific changes to Experiment 2 to deal with this and other potential problems are discussed at the end of the paper.

2. Background

NPIs vary by the type of semantic context that can license them, and the work of Zwarts (1981) was one of the first to begin to recognize this diversity. Restricting our current discussion to a simplified two-way distinction, weak NPIs (WNPIs) can be licensed by *weak licensors*, which create downward entailing contexts while *strong NPIs* (SNPIs) need to be licensed by *strong licensors,* which create anti-additive contexts. Anti-additive contexts are

a proper subset of downward entailing contexts. Sentential negation is a Strong Licensor that can license all SNPIs as well as all WNPIs, e.g. (1). Sentential adverbs like *hardly* or *seldom* and certain quantifiers, e.g. *Few of the X,* cannot license SNPIs like *give a hoot* in (2), e.g.

(1) The local politicians don't give a hoot about the environment.
(2) *The local politicians hardly give a hoot about the environment.

This brief presentation is a simplification of the amazing complexity of NPIs and their licensors, and subsequent work has argued for even more distinctions, such as *anti-multiplicative* contexts (Zwarts 1998), anti-morphic contexts (Van der Wouden 1994), and nonveridical contexts (Giannakidou 1997), each associated with their own subgroup of NPIs. An overview of the rich history of this research area can be found in Hoeksema (2012). Hoeksema (2012) also adds to the mysterious nature of NPI licensing by presenting carefully collected corpus data that show that the variation in licensing contexts is much richer than the theoretical work suggests, with certain semantic subsets of NPIs only appearing in a rather heterogeneous set of contexts.

In this complex situation experimental work can help extend our understanding. But experimental work with NPIs has been surprisingly absent. A recent exception is the work of Richter & Radó (2013) who did four experiments testing SNPIs and WNPIs in German. Experiments 1 and 2 used typical Strong and Weak Licensors. While researchers seem to generally agree on which NPIs are strong and which are weak, experimentally confirming these categories with naïve speakers using Likert-scale judgments is challenging. Richter & Radó needed multiple experiments varying the types tested and the optimal type of fillers needed to obtain experimental results that align with the intuitive classifications of linguists. This difficulty can be understood as a consequence of one of the main results of the Richter & Radó study: SNPIs used with a weak licensor are not completely unacceptable, contrary to many predictions. In their experiments they had a rating of about 4 on a 7-point Likert-scale (See Richter & Radó 2013: 9).

In Experiment 3 Richter & Radó used the items confirmed in Experiments 1 and 2 to investigate the licensing properties of Negative Raising constructions. Briefly, in Negative Raising constructions, sentential negation in the matrix clause is interpreted as taking scope over the embedded clause, e.g. *John didn't think that Ed was a thief* means John thought that Ed was not a thief. Richter & Radó found that matrix negation in Negative Raising constructions could license SNPIs. However, these sentences were not rated as acceptable as WNPIs in the same context. This result shows two things: overt negation is not entirely necessary to license SNPIs, and there is still a difference in the acceptability of

Negative Raising contexts as licensors because WNPIs were considered more acceptable in these contexts than SNPIs.

Richter & Radó (2013) was done on German, but as far as we know there is no similar published study in English.[3] This work makes a first attempt at filling that gap, by attempting to experimentally confirm the SNPI or WNPI status of a set of NPI items in English. Further, we avoid the classic NPIs, e.g. *any, even* etc. which have all been discussed extensively in the literature, and instead have as a secondary goal to extend our knowledge of the less studied NPIs.[4]

3. Selection of NPIs and Corpus data

The NPIs chosen for the experiments are listed in Table 1. Initially we selected a large set of NPIs, and organized them according to part of speech. Then all candidate SNPIs were examined in the COCA Corpus (Davies 2008) to check that the NPIs did not occur with weak licensors. Finally, for each potential Weak-Strong pair, the relative frequency in the Hoeksema-NPI corpus was considered to make sure that very rare NPIs were not paired with very frequent ones, to avoid unwanted effects.[5]

Candidate WNPIs	Candidate SNPIs
bother to, have a clue about, have the foggiest notion about, be born yesterday, have heard the last of, give a hoot, see eye to eye, can fathom, lift a finger, say no to, worth the paper X is written on, give a rat's ass, will be a moment, would bank on, help matters	*dream of, know the first thing about, budge an inch, be half bad, have a prayer, bat an eyelid, mince words, can shake the idea of, have heard the last of, touch with a ten foot pole, have a leg to stand on, be trifled with, think twice about, would hear of, would harm a fly*

Table 1. List of Strong and weak NPIs used in Experiment 1

[3] Eva Csipak (p.c.) has done a similar investigation as Experiment 1 as a pilot study.

4 Certainly others have had the same desire, e.g. the recent volume *Beyond 'Any' and 'Ever'*, Csipak et al. Eds. (2013).

5 Richter & Radó (2014) also controlled for frequency, though this isn't explicitly mentioned in the paper.

4. Experiment 1

Experiment 1 was a 2 x 2 study with two factors, candidate NPI Type (weak vs. strong) and Licensor Type (weak vs. strong). We selected 16 candidate SNPIs and 16 WNPIs based on the corpus investigations described above (See Appendix A for Target Items). As weak licensors we only used the adverbs *hardly, rarely* or *seldom.* As strong licensors we used sentential negation. Subjects saw only one version of each item, e.g. see (3):

(3) a. Mike would seldom say no to a case involving terrorist accusations.
b. Mike would seldom touch a case involving terrorist accusations with a ten foot pole.
c. Mike wouldn't say no to a case involving terrorist accusations.
d. Mike wouldn't touch a case involving terrorist accusations with a ten foot pole.

Richter & Radó (2013) found that the number and type of fillers has a major effect on the evaluation of NPIs. Based on this previous experience, the current experiment used three types of fillers:

Creative idioms: These were regular idioms but with a creative variation. These items are certainly understandable but not idiomatic., e.g. *Stop beating around the bushes and tell me what's wrong.* (→ beating around the bush)

Idiom Blends: These are two idioms blended together. Examples were modified from those in Coppock (2009), e.g. Lines are open and there are people waiting by to take your call. These are odd, but in a more subtle way than the creative idioms.

Grammatical Fillers: Taken from newspapers and some from the COCA corpus (Davies 2008), e.g. Right now only a handful of American cities have super-fast internet.

We used 8 Creative Idioms, 8 Idiom Blends, and 16 Grammatical Fillers in addition to the 16 Target Items. 32 participants were tested via Amazon's Mechanical Turk service (AMT) and compensated $1.20. The results from one non-native English speaker were discarded leaving 31 participants for analysis (17 female, Mean Age: 37.7, SD 9.8, Range 23-67). Subjects were asked to judge each sentence on a 7 point Likert-scale where 1 was 'Unacceptable' and 7 was 'Acceptable'.

Licensor Type	NPI Type	Mean rating	SD
Weak (adverb)	Weak	5.91	1.33
	Strong	5.12	1.67
Strong (negation)	Weak	6.15	1.14
	Strong	6.18	1.05

Table 2. Results Experiment 1

The mean rating for Grammatical Fillers was 6.08, for Creative Idioms 4.92 and for Blends, 4.03. This means that all NPIs had a higher average rating than Creative Idioms or Idiom Blends, similar to the results in Richter & Radó (2013) that even poorly licensed NPIs are not so bad. In fact they are about as bad as a Creative Idiom, and much better than an Idiom Blend.

Two-factor (Candidate NPI Type x Licensor Type) repeated-measures analyses of variance (ANOVAs) were performed on the aggregated mean ratings, with participants as random factors. There was a significant main effect for NPI type ($F(1,30)=17.35$, $p=0.000$) and licensor ($F1(1,30) = 28.02$, $p = 0.000$) and a significant interaction effect ($F1(1,30) =10.17$, $p=0.003$). Paired t-tests showed that SNPIs licensed by strong licensors were rated significantly higher than SNPIs licensed by weak licensors ($p = 0.000$).

5. Experiment 2

Now that we've empirically verified a set of strong and weak NPIs, we can use them to test other predictions. The licensing properties of different types of conditionals might benefit from experimental investigation. In general conditionals are considered to be weak licensors, and should not allow SNPIs. But there seem to be conditions under which certain SNPIs, such as minimizers can be licensed by certain conditionals, e.g. see Csipak (2014) for how the speech-act type of the consequent might influence SNPI licensing. Rhetorical Conditionals seem to be another exceptional type. Consider the following with the SNPI *bat an eyelid* which means “care” :

(4) a. If local politicians bat an eyelid at what the governor says, I'm the Queen of Sheba.
b. If local politicians bat an eyelid at what the governor says, they will not try to push through new taxes this fall.

By using the phrase “I'm the Queen of Sheba” the speaker signals that they do not believe the antecedent of the conditional, because the speaker is in fact, *not* the Queen of Sheba. In logical terms, the speaker is saying “If A then falsum”. This is logically equivalent to “Not A”. “Not A” is sentential negation. But does this environment then actually have the ability to license SNPIs, as straightforward sentential negation would ? It does require additional inferential steps. On the other hand, the results from Experiment 3 in Richter & Radó (2013) did find that Neg-raising contexts could partially license SNPIs.

If we compare the Rhetorical Conditional in (4a) with the Causal Conditional in (4b), the first author’s intuition is that (4a) is much more acceptable, perhaps even completely fine. But before we can run such an experiment, we have to determine what can be considered a Rhetorical Conditional. Certain expressions clearly belong to this category, and perhaps more than I know myself, but there seems to be a very limited number of well-known examples, e.g.:

(5) If you pass the exam, I’m a Dutchman/ I’m the Empress of China/I’ll eat my hat,…

There are other expressions that can occur in the consequent of a conditional that also strongly suggest that the speaker does not believe that the antecedent of the conditional is true, or will come true, e.g.

(6) If you pass the exam, I’ll be astounded/surprised/amazed.

(5) and (6) seem semantically similar, though (5) is much stronger. In Experiment 2, in order to avoid presenting the more colorful rhetorical consequences more than once to participants *be surprised/astounded/amazed* were also used, keeping the possibility in mind that they might not work quite as well.

5.1 Method

Experiment 2 was again a 2 x 2 design with *NPI Type* (strong vs. weak as established in Experiment 1) and *Conditional Type* (Rhetorical vs. Normal) as factors. Items from Experiment 1 were rewritten to create four new versions of each item. Not all NPIs were able to be naturally rewritten into the antecedent of a conditional, e.g. *be born yesterday* is very odd. For this reason only 12 of the items from Experiment 1 were usable in Experiment 2. An example of an item is given in (7). We predict that (7b) will be more acceptable than (7d):

(7) a. If Mr. Light has a clue about the bills he votes on, I'll eat my hat.
b. If Mr. Light knows the first thing about the bills he votes on, I'll eat my hat.
c. If Mr. Light has a clue about the bills he votes on, he will certainly vote against Proposition 12.
d. If Mr. Light knows the first thing about the bills he votes on, he will certainly vote against Proposition 12.

We tested 55 participants via AMT. **(**Mean age 36.46 (*SD*= 13.05, range 20-76, 27 females; Compensation $1.20). All were self-identifying native speakers of English. None of the participants had taken part in Experiment 1. The fillers were identical to those used in Experiment 1, except 4 of the grammatical sentences were removed and 8 grammatical conditionals were added, for a total of 48 items.

5.2 Results

An ANOVA was performed on the aggregated means with participants as random factors. There was a main effect of *NPI type*. WNPIs were rated more acceptable than SNPIs overall (F(1,55)=6.0, p=0.02). For *Conditional Type*, there was no significant effect. There was a significant interaction effect (F(1,55)=5.409, p=0.02). Planned comparisons showed that WNPIs in normal conditionals were marginally significantly better than SNPIs in normal conditionals (*t-value*= 1.547,p=0.06). There was no significant difference between SNPIs in Normal and Rhetorical Conditionals, even though the means do suggest such a difference. This is probably because of the large amount of variance in the results as indicated by the standard deviation. (see Table 3).

Licensor Type	NPI Type	Mean rating	SD
Normal Conditional	Weak	5.33	1.64
	Strong	4.60	1.99
Rhetorical Conditional	Weak	5.09	1.74
	Strong	4.90	1.86

Table 3. Results Experiment2 Comparing licensing ability of Normal vs. Rhetorical Conditional with Weak and Strong NPIs

6. Discussion

If we consider Experiment 2 as a pilot study, there are number of immediate extensions that could be done to more thoroughly investigate the potentially different licensing features of Normal and Rhetorical Conditionals. First, we could increase the number of observations of each NPI –licensor combination to get more data from each individual participant. The three observations used in Experiment 2 were probably too few to see any clear factor effects. Second, we should probably only use strong Rhetorical Conditionals, and avoid consequences like “I'd be amazed”.

The results from both Experiment 1 and Experiment 2 do suggest that the acceptability of NPIs is more gradual than a direct cut off point given the clear distinction between different entailment properties. This is similar to the results in Richter & Radó (2013) who found that while slightly modified idioms were rated on par with SNPIs and a weak licensor, while twice modified idioms were rated worse than SNPIs with a weak licensor. Zwarts’ initial intuition that there is more to licensing than downward-entailment is certainly confirmed.

7. Conclusion and directions for future research

We can conclude that naïve subjects can experimentally distinguish between weak and strong NPIs in English, further confirming that there are different degrees of licensing ability. Additionally, the numerical means from Experiment 2 showed that the type of conditional may matter for licensing, though the results were not statistically significant. However, the small number items per condition tested was a weakness of the experimental design. Certainly this is an obvious result to build future work on.

Acknowledgements: The first author would like to thank Jack Hoeksema for sharing his data and expertise, and Thomas Weskott for help with the statistical analysis as well as useful discussion.

8. References

Csipak, Eva, Regine Eckardt, Mingya Liu, Manfred Sailer, Eds. (2013). *Beyond 'Any' and 'Ever'.New Explorations in Negative Polarity Sensitivity*. Berlin, Boston: De Gruyter Mouton.

Csipak, Eva (to appear). 'Minimizers in conditional threats and promises.' In E. Chemla, V. Homer and G. Winterstein, Eds. Proceedings of *Sinn und Bedeutung* 18.

Coppock, Elizabeth(2009). 'Parallel grammatical encoding in sentence production: Evidence from syntactic blends', *Language and Cognitive Processes,* 25(1): 38-49.

Davies, Mark. (2008) *The Corpus of Contemporary American English: 450 million words, 1990-present*. Available online at http://corpus.byu.edu/coca/.

Eckart, Regine (2012). 'The many careers of negative polarity items' In C. Bert, H. Cuyckens and T. Leuschner. *Grammaticalization and Language Change: New Reflections.* Amsterdam: John Benjamins.

Giannakidou, A. (1997). *The Landscape of Polarity Items.* Doctoral Dissertation. University of Groningen

Hoeksema, J. (1997) `Corpus study of negative polarity items' *IV-V Jornades de corpus Linguistics*.

Hoeksema, J. (2012). "On the natural history of negative polarity items." *Linguistic Analysis* 38.1: 3.

Richter, Frank, and Janina Radó. (2013) "Negative Polarity in German: Some Experimental Results." *Journal of Semantics* 31.1, 43-65.

Wouden, Ton van der (1994). *Negative Contexts.* Doctoral Dissertation, University of Groningen.

Zwarts, F. (1978). Extractie uit prepositionele woordgroepen in het Nederlands. In: A. van Berkel et al. (eds.), *Proeven van Neerlandistiek, Aangeboden aan prof. dr. Albert Sassen*. Groningen: Nederlands Instituut Groningen, 303-399.

Zwarts, Frans. (1981). Negatief polaire uitdrukkingen I. GLOT 4(1), 35-133.

Zwarts, F. (1995). Nonveridical Contexts. *Linguistic Analysis* 25, 3-4,286-312.

Zwarts, Frans. (1998) "Three types of polarity." *Plurality and quantification*. Springer Netherlands. 177-238.