An experimental investigation of NPI licensing under DE flip-flop

Erika Mayer,[1] Susanne Wurmbrand,[2,3] & Jon Sprouse[2]

[1]University of Massachusetts, [2]University of Connecticut, [3]University of Vienna

1. Introduction

This paper presents the results of an acceptability judgment study investigating NPI licensing in contexts with multiple downward entailing operators. The results support an environment-based approach to NPI licensing that includes licensing domains, such as the one proposed in Homer (2012).

Negative polarity items (NPIs) are a class of words and expressions that are licensed in the scope of negation and unlicensed in positive or affirmative environments. Examples of NPIs in English include *any*, *ever*, *at all*, *lift a finger*, and many more. An NPI's ability to be licensed in a particular location is closely tied to monotonicity, which refers to whether a given environment is upward entailing (UE), downward entailing (DE), or neither. DE environments license inferences from sets to subsets, and are defined formally in the following way (Ladusaw 1979, Homer 2012):

(1) A function f of type $< \sigma$, t$>$ is DE if and only if for all x, y of type σ such that x $\Rightarrow$ y: f(y) $\Rightarrow$ f(x).

UE environments, which include affirmative sentences, license inferences from sets to supersets. In general, NPIs are acceptable, or licensed, in the scope of DE operators and unlicensed in affirmative, UE environments. The NPI *any* is degraded in (2) below, a UE sentence with no DE operators, but licensed in (3), where it is located in the scope of the DE verb *doubt*:

(2) #I think that she felt any relief after the test.

(3) I doubt that she felt any relief after the test.

Approaches to NPI licensing that rely on the notion of downward entailingness can be grouped into three broad categories: operator-based approaches, environment-based ap-

proaches, and sub-environment-based approaches. Operator-based theories state that an NPI must be c-commanded by or within the semantic scope of at least one DE operator to be licensed (Ladusaw 1979, von Fintel 1999). Environment-based theories, as their name suggests, state that an NPI must be located in an environment that is DE with respect to its position. Finally, sub-environment approaches begin with an environment-based framework to NPI licensing and add additional constraints, such as syntactic licensing domains.

For most sentences, operator-based approaches and environment-based approaches to NPI licensing make the same predictions. When an NPI is in the scope of a DE operator, it is generally in a DE environment, and vice versa. One way to tease the two approaches apart is to find a position or environment that is both UE and within the semantic scope of at least one DE operator. Such environments are created when two DE operators both take scope over one position. The addition of a second DE operator can "flip" the monotonicity of an environment from DE to UE, resulting in a UE environment that is nevertheless within the semantic scope of multiple DE operators. (4) shows the entailment pattern for a sentence with one DE operator, *doubt*; such a sentence licenses an inference from the set *cat* to the subset *orange cat*. (5) demonstrates that the addition of a second DE operator, *don't*, reverses monotonicity from DE to UE: the sentence now licenses an inference from the set *cat* to the superset *animal*.

(4) I doubt that Emily bought a cat. ⇒ I doubt that Emily bought an orange cat.

(5) I don't doubt that Emily bought a cat. ⇒ I don't doubt that Emily bought an animal.

Operator and environment-based approaches to NPI licensing make very different predictions for sentences in which an even number of DE operators create a UE environment. Operator-based approaches predict that NPIs should be perfectly acceptable in such environments: the NPI simply needs to be in the scope of one or more DE operators to be licensed. Environment-based approaches, on the other hand, predict that such sentences should not license NPIs, since the environment itself is UE.

The environment-based notion that NPIs can be "anti-licensed" by an even number of DE operators is called flip-flop. The phenomenon of flip-flop has been widely challenged; Chierchia (2004), for example, noted that flip-flop occurs for scalar implicatures but not for polarity items. Not all environment-based accounts to NPI licensing, however, state that flip-flop occurs no matter where the NPI and DE operators are located in a sentence. Some "sub-environment" approaches, such as Homer (2012), posit that while NPIs do need a DE environment, they need not compute monotonicity with regard to the entire sentence in which they are located. Instead, their licensing requirements may be satisfied by smaller constituents that are locally DE.

In Homer's sub-environment-based account to NPI licensing, NPIs are licensed by DE domains. These domains are syntactic constituents in which monotonicity is computed. In order to be licensed, an NPI must find at least one domain that is DE with respect to its position. Licensing domains are lexically-oriented and may be different for different NPIs. For some NPIs, the licensing domain is the polarity phrase (or PolP) of the sentence. The

examples below show how Homer's account makes different predictions for sentences that contain two DE operators in different positions. Brackets represent PolP domains.

(6) I [$_{\text{DOMAIN 1}}$ doubt that Emily [$_{\text{DOMAIN 2}}$ didn't feel any relief after the test.]]

(7) ?I [$_{\text{DOMAIN 1}}$ don't doubt that Emily [$_{\text{DOMAIN 2}}$ felt any relief after the test.]]

In both (6) and (7), the NPI *any* must be in a DE environment with respect to at least one of its two licensing domains. In (6), the larger licensing domain, Domain 1, contains two DE operators, *doubt* and *didn't* (*didn't* is included in both Domain 2 and Domain 1). Together, *doubt* and *didn't* create a UE environment at the position of the NPI *any*. This domain, therefore, is not eligible. The smaller domain, Domain 2, contains only one DE operator, *didn't*. In isolation, Domain 2 is DE with respect to *any*'s position, so *any* is licensed by this smaller domain and is predicted to be acceptable. In (7), however, neither licensing domain is DE with respect to *any*'s position. Domain 1 contains two DE operators, *don't* and *doubt*, that again create a UE environment. Because *don't* and *doubt* are both located in the matrix clause and not the embedded clause, Domain 2 has no DE operators at all and is locally UE. Neither domain is DE with respect to *any*, so (7) is predicted to be unacceptable.

Although Homer's account states that evidence for flip-flop is fairly clear for strict NPIs and French NPIs, he notes that evidence for flip-flop for weak English NPIs is less clear. The acceptability of a weak NPI like *any* in UE environments with multiple DE operators may also vary between dialects or among speakers. Given the inconclusive evidence for flip-flop in English, the experiments presented in this paper were designed to test the predictions of the operator, environment, and sub-environment approaches to NPI licensing by investigating the acceptability of sentences with the NPI *any* and multiple DE operators in various positions.

2. The Experiments

2.1 Design

The goal of our study was to evaluate three types of theories (operator, environment, and sub-environment) by testing the acceptability of the NPI *any* in sentences with multiple DE operators. We constructed eight sentence schemas that contain different combinations of DE operators. For each combination of DE operators, we constructed a control condition with no NPI and a target condition with an NPI. Superficially, these 16 sentences form a 2x2x2x2 design (matrix negation x matrix verb x embedded negation x NPI). However, for the purpose of teasing apart these three theories, we can treat them as eight pairwise-phenomena, where we define an unlicensed NPI as being significantly less acceptable than the same sentence with no NPI.

Table (8) below provides an example of each combination of DE operators, organized such that the first four establish the basic NPI facts, and the second four tease apart the operator (Op), environment (Env), and sub-environment (Sub-env) theories. The predictions of each theory are given in the rightmost three columns. Only target conditions with *any*

are given below; control conditions were identical to target conditions save for the absence of the NPI, and were all predicted to be acceptable. Brackets indicate PolP domains.

(8) *Sample stimuli with examples of each target condition with an NPI*

Pair	Condition	Sentence	Op	Env	Sub-env
1	none	Jeff [thinks that the artist [had **any** pride in the painting.]]	×	×	×
2	neg(embed)	Jeff [thinks that the artist didn't have **any** pride in the painting.]]	✓	✓	✓
3	neg(matrix)	Jeff [doesn't think that the artist [had **any** pride in the painting.]]	✓	✓	✓
4	doubt	Jeff [doubts that the artist [had **any** pride in the painting.]]	✓	✓	✓
5	doubt-neg	Jeff [doubts that the artist didn't have **any** pride in the painting.]]	✓	×	✓
6	neg-neg	Jeff [doesn't think that the artist didn't have **any** pride in the painting.]]	✓	×	✓
7	neg-doubt	Jeff [doesn't doubt that the artist [had **any** pride in the painting.]]	✓	×	×
8	neg-doubt-neg	Jeff [doesn't doubt that the artist didn't have **any** pride in the painting.]]	✓	✓	✓

2.2 Method

Eight unique tokens were created for each of the 16 conditions in lexically matched sets. The matrix verb was always *think* or *doubt*. The 16 conditions were divided into two experiments, each with 8 target conditions. Each list of 8 experimental items was combined with 9 initial practice items and 14 filler items. The resulting lists were pseudorandomized with the constraint that no experimental items occurred consecutively without a filler item in between. Finally, lists were counterbalanced such that each list had four variations. For each experiment, then, there were 32 counterbalanced lists, one for each participant. Each participant rated a total of 31 sentences: 9 practice items, 14 filler items, and 8 experimental items, one from each condition. The task was a seven-point (Likert-esque) rating scale.

One potential problem with investigating the acceptability of the NPI *any* in UE environments is that *any* has a free-choice counterpart. If participants were able to interpret *any* as a free-choice item rather than an NPI, it might appear licensed in a wider range of environments. One way to mitigate this concern is to pair *any* with abstract mass nouns. In an untimed acceptability judgment experiment, Parker & Phillips (2016) found that free-choice *any* is unacceptable with abstract mass nouns, as in (9) below:

(9) #She experienced any relief after the ordeal.

In our study, *any* always modified an abstract mass noun such as *disdain*, *relief*, *amazement*, *resentment*, *pride*, *apprehension*, *poverty*, and *rest*.

2.3 Participants and procedure

For each experiment, 32 participants were recruited via Amazon Mechanical Turk. They were paid $1.50 for the completion of the survey. All participants were located in the United States. In both experiments, 30 out of 32 participants indicated that they were born in the United States and that both of their parents spoke English to them at home. Only these native speakers of English were included in the analyses. Three Amazon Mechanical Turk participants who completed the first experiment completed the second as well. Therefore, there were 57 unique participants in total.

3. Results

The responses of each participant were z-score transformed before analysis to eliminate scale bias. Means and standard deviations for z-scores and raw scores are reported below.

(10) *Means and standard deviations for raw scores and z-scores*

Pair	Condition	Raw mean	Raw SD	Z-score mean	Z-score SD
1	none no-NPI	6.10	1.09	0.74	0.45
1	none NPI	3.07	1.62	-0.68	0.74
2	neg(embed) no-NPI	5.57	1.33	0.36	0.61
2	neg(embed) NPI	6.10	0.99	0.68	0.43
3	neg(matrix) no-NPI	5.70	1.42	0.44	0.70
3	neg(matrix) NPI	6.57	0.73	0.96	0.39
4	doubt no-NPI	5.87	1.36	0.62	0.49
4	doubt NPI	5.87	1.36	0.61	0.51
5	doubt-neg no-NPI	4.83	1.39	0.12	0.57
5	doubt-neg NPI	4.63	1.54	0.06	0.60
6	neg-neg no-NPI	4.00	1.23	-0.33	0.58
6	neg-neg NPI	4.27	1.34	-0.23	0.60
7	neg-doubt no-NPI	5.73	1.46	0.56	0.61
7	neg-doubt NPI	4.37	1.79	-0.08	0.73
8	neg-doubt-neg no-NPI	4.97	1.27	0.08	0.56
8	neg-doubt-neg NPI	5.03	1.30	0.11	0.53

We ran eight paired *t*-tests with a one-tailed Dunn-Bonferroni-corrected *p*-value of .00625 as a criterion for significance. With eight comparisons, this value leads to a maximum familywise Type I error rate of .05. Table (11) presents the results of these eight *t*-tests, with

bold text indicating statistical significance. *P*-values lower than .0001 have been rounded up to .0001.

(11) *Statistical results of paired one-tailed t-test*

Pair	Condition	*t*-statistic	*p*-value
1	none	-11.38	**.0001**
2	neg(embed)	3.03	.9975
3	neg(matrix)	3.55	.9993
4	doubt	-0.14	.4454
5	doubt-neg	-0.46	.3225
6	neg-neg	0.77	.7776
7	neg-doubt	-4.05	**.0002**
8	neg-doubt-neg	0.29	.6127

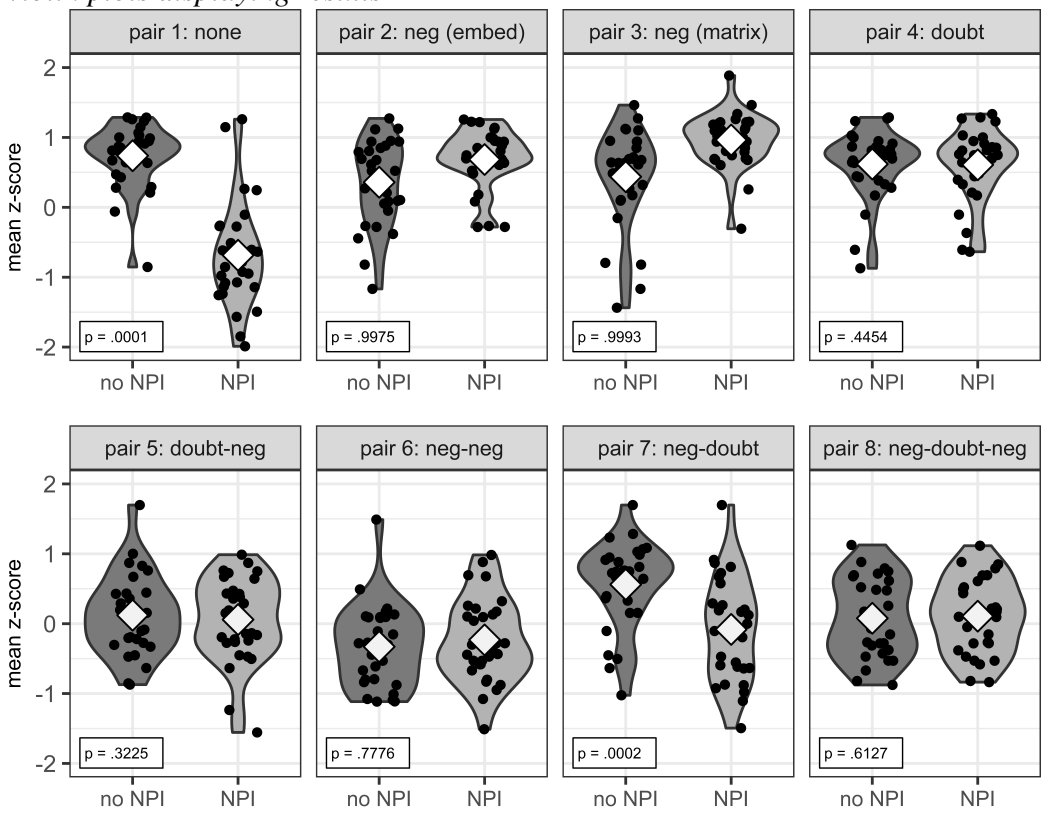
Figure (12) displays the distribution of z-scores for each condition. Pair 1 demonstrates canonical failure of NPI licensing when there is no DE operator in the sentence. Sentences with no DE operators were rated significantly lower when they contained *any* than when they did not. These results suggest that *any* was indeed interpreted as an unlicensed NPI in these contexts, not a free-choice item. Pairs 2, 3, and 4 demonstrate the licensing of NPIs with the DE verb *doubt* and under sentential negation in embedded and matrix positions. For these pairs, there was no significant difference between the NPI and no-NPI conditions.

The critical pairs were 5-8. Pairs 5 and 6 have two DE operators, one in the matrix clause and one in the embedded clause. These two DE operators create a UE environment at the position of the NPI. The environment approach predicts that *any* should be unlicensed due to the global UE environment. The operator and sub-environment approaches both predict that *any* should be licensed, but for different reasons. Under the operator-based approach, *any* should be licensed due to the presence of at least one DE operator taking scope over its position. Under the sub-environment-based approach, NPIs should be licensed due to the local DE domain created by the single DE operator in the embedded polarity phrase. Our results reveal no significant difference between the no-NPI and NPI sentences for these two pairs. This result is inconsistent with an environment-based theory, but consistent with the predictions of both the operator and sub-environment accounts.

Pair 7 is the critical pair for distinguishing operator and sub-environment-based approaches, as it has two DE operators in the matrix clause. The sub-environment-based approach predicts that the NPI should be unlicensed because there is no PolP domain that is DE, whereas the operator approach predicts that the NPI should be licensed because there is at least one DE operator taking scope over the NPI. Our results suggest a significant decrease in acceptability for the NPI sentences, providing evidence that sub-environment theories, such as Homer (2012), are more likely to be correct.

Finally, pair 8, with three DE operators, was predicted to license the NPI under all three approaches: there are multiple DE operators scoping over the NPI, and the NPI's environment is DE. As predicted, there was no significant difference between NPI conditions and no-NPI conditions for pair 8.

(12) *Violin plots displaying results*



Violin plots display the distribution of z-scores for each condition, with white diamonds displaying the mean z-score. Points display individual acceptability ratings. Labels in the lower left corner indicate p-values for each condition. Values lower than .0001 are rounded up to .0001.

4. Discussion

4.1 Evidence for flip-flop

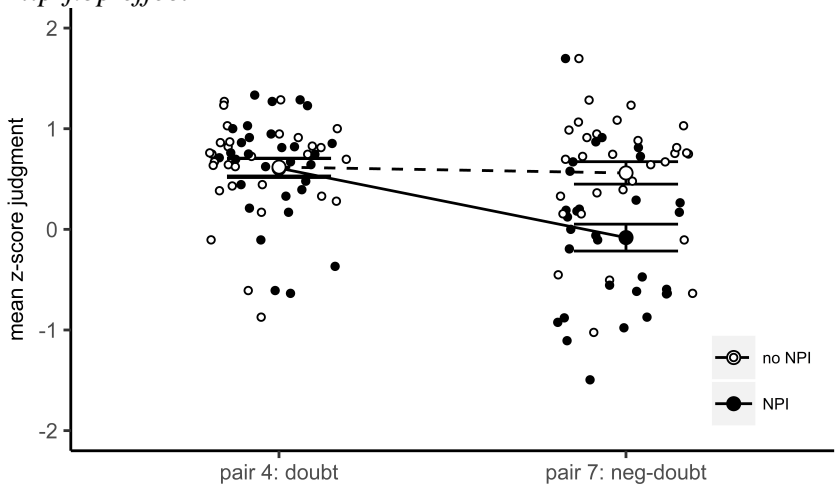
This paper presented the results of acceptability judgment experiments designed to test the predictions of three approaches to NPI licensing for the English weak NPI *any*. These approaches included an operator-based approach stating that NPIs must be within the scope of at least one DE operator; an environment-based approach stating that NPIs must occur in a global DE environment; and a sub-environment-based approach like that of Homer (2012) stating that an NPI must be contained within at least one DE syntactic domain. Our results provide support for Homer's sub-environment approach and suggest that both the position of DE operators and the monotonicity of environments are critical for NPI licensing.

One of the key predictions of an environment or sub-environment-based approach to NPI licensing is that NPIs will be susceptible to flip-flop, a phenomenon by which multiple DE operators degrade the acceptability of an NPI by creating a UE environment. In our study, sentences with two consecutive DE operators (*doesn't* and *doubt*) received lower ratings when they contained *any* than when they did not, suggesting that *any* was unlicensed in these environments. This significant difference provides evidence for flip-flop: despite

multiple DE operators taking scope over its position, the NPI *any* was unlicensed due to its location in a UE environment. Here, an operator-based account makes the incorrect prediction that *any* should be licensed, as it is in the scope of not one but two DE operators. An environment-based approach to NPI licensing, however, makes the correct prediction that the NPI should be unlicensed due to the UE environment created by the two DE operators.

The interaction plot in Figure (13) shows the "flip-flop" effect for sentences with two DE operators in a matrix polarity phrase. For pair 4, which contained only the DE operator *doubt*, sentences with *any* were not significantly different from sentences without an NPI. For pair 7, with the two DE operators *doesn't doubt*, sentences with *any* were rated as significantly less acceptable than those without *any*.

(13) *Flip-flop effect*



Interaction plot showing the difference in acceptability between NPI and no-NPI conditions in environments with one DE operator (pair 4) and two DE operators (pair 7). Error bars represent standard error. Smaller points display individual acceptability judgments.

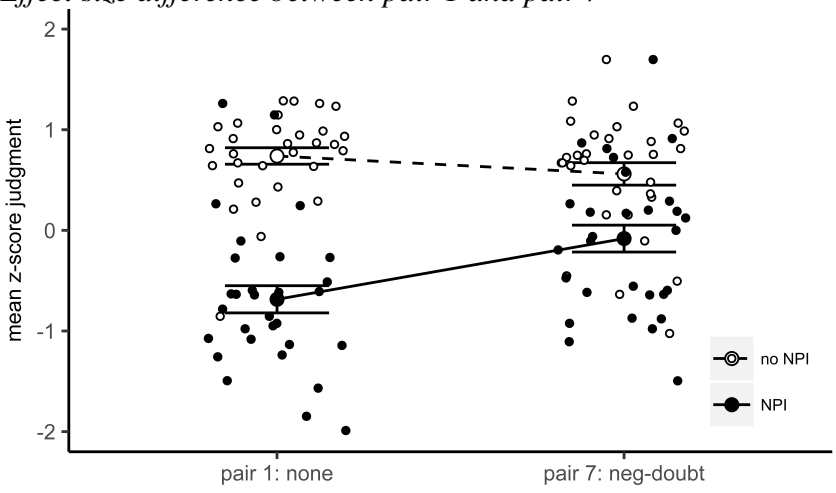
Despite evidence for flip-flop in sentences with *doesn't doubt*, not all pairs with two DE operators led to a flip-flop effect. We found no significant differences between NPI and no-NPI conditions when one DE operator was located in the matrix PolP and the other was located in an embedded PolP (pairs 5 and 6). For these conditions, an environment-based account without any syntactic licensing domains predicts, incorrectly, that there should be significant differences between NPI and no-NPI conditions. The two DE operators together create a UE environment, and this is all that matters for an environment-based approach without syntactic domains: the position or locality of the DE operators is irrelevant.

Unlike the environment-based approach without domains, Homer's sub-environment-based approach to NPI licensing correctly predicts no significant differences between NPI and no-NPI conditions for pairs 5 and 6. In this account, the embedded PolP is a domain in which monotonicity can be evaluated at a more local level. When this domain is considered in isolation, only one DE operator takes scope over the NPI, creating a local DE environment. This DE domain licenses *any* despite the fact that the sentence overall is UE with respect to *any*'s position.

4.2 Effect size and NPI Illusions

Though the pattern of effects that we see in this study confirms the sub-environment predictions, the finer details of the results also help to explain why the phenomenon of flip-flop has been controversial. The issue is that the size of the "flip-flop" effect (i.e., the difference between the no-NPI and NPI conditions) in the unlicensed NPI pair 7 with two DE operators is much smaller than the effect in the unlicensed NPI in pair 1, the typical case. Table (14) shows this difference in effect size.

(14) *Effect size difference between pair 1 and pair 7*



Interaction plot showing the difference in effect size between pair 1 (with no DE operators) and pair 7 (the flip-flop case). Error bars represent standard error for each condition. Smaller points display individual acceptability judgments.

This smaller effect means that the absolute rating of the unlicensed NPI in pair 7 is much higher than the unlicensed NPI in pair 1. For example, although the NPI condition for pair 7 is rated significantly worse than the no-NPI condition, its absolute rating is similar to that of the licensed NPIs in pairs 6 and 8. It is the comparison to a baseline without *any* that suggests the NPI in pair 7 sentences is not licensed rather than the sentences' absolute ratings in comparison to similar sentences.

Mismatches between absolute acceptability and difference effects have become more and more common in the experimental syntax literature (e.g., Featherston 2005), underscoring the importance of well-chosen control conditions and experimental designs. However, in this case, there appears to be a bit more going on. One hypothesis is that this pattern may be due to the well-known grammatical illusions that arise for NPI licensing (Xiang et al. 2009, Parker & Phillips 2016). The typical NPI illusion occurs when negation linearly precedes, but does not c-command, an NPI, as in the following example from Parker & Phillips (2016):

(15) The authors that no critics recommended have ever received acknowledgment for a best-selling novel.

In (15) above, the NPI licensor *no* is embedded in a relative clause, where it does not c-command the NPI *ever*. Even though the NPI licensor is structurally inaccessible in this sentence, its presence often creates the illusion of grammaticality.

In our study, for sentences containing the sequence *doesn't doubt*, two NPI licensors c-command and take scope over the NPI. Here, the "illusion" is a weakening of the UE effect of *doubt*. This hypothesis predicts that participants who do not experience an NPI illusion should find the sequence *doesn't doubt* with an NPI as unacceptable as the unlicensed NPI in pair 1, with no DE operators. Participants who do experience an illusion, however, may find the sequence more acceptable due to the presence of DE operators. Future processing studies can test this account by investigating the extent to which sentences with NPIs and multiple DE operators display the same patterns as other NPI illusions.

4.3 Concluding remarks

The study presented in this paper provides experimental evidence that the weak NPI *any* is sensitive to the monotonicity of the environment in which it is located, not simply the presence of DE operators. It also sheds light on how the position of multiple DE operators in different phrases affects NPI licensing, supporting an account of NPI licensing that takes both monotonicity and syntactic constituency into account (Homer 2012). Although our study provides evidence of a "flip-flop" effect for the NPI *any*, the effect size was not as large as that of conditions with no DE operators at all, and future work may investigate the reasons for this effect size difference.

References

Chierchia, Gennaro. 2004. Scalar implicatures, polarity phenomena and the syntax/pragmatics interface. In *Structures and beyond*, ed. Adriana Belletti. Oxford: Oxford University Press.

Featherston, Sam. 2005. Universals and grammaticality: wh-constraints in German and English. *Lingua* 115:1525–1550.

von Fintel, Kai. 1999. NPI licensing, Strawson Entailment, and context dependency. *Journal of Semantics* 16:97–148.

Homer, Vincent. 2012. Domains of polarity items. Unpublished manuscript.

Ladusaw, William A. 1979. Polarity sensitivity as inherent scope relations. Doctoral dissertation, University of Texas, Austin, TX.

Parker, Dan, & Colin Phillips. 2016. Negative polarity illusions and the format of hierarchical encodings in memory. *Cognition* 157:321–339.

Xiang, Ming, Brian Dillon, & Colin Phillips. 2009. Illusory licensing effects across dependency types: ERP evidence. *Brain and Language* 108:40–55.

Erika Mayer, Susanne Wurmbrand, Jon Sprouse
elmayer@umass.edu, susi@alum.mit.edu, jon.sprouse@uconn.edu