The domains of monotonicity processing

I-An Tan[1] , **Nir Segal**[2] , **Yosef Grodzinsky**[1,3]

[1] Neurolinguistics Lab, Edmond & Lily Safra Center for Brain Sciences, Hebrew University, Jerusalem, Israel

[2] Department of Linguistics, LLCC, Hebrew University, Jerusalem, Israel

[3] Institute of Neuroscience and Medicine (INM-1), Research Center Jülich, Jülich, Germany

Abstract

This paper reports an investigation into the nature of Negative Polarity Item (NPI) licensing conditions from a processing perspective. We approached empirical aspects of this issue not via the standard judgment paradigm, but rather, by measuring continuous variables (error rates, Reaction Times - RTs) in a verification task, in which the truth value of a sentence is determined against a scenario. Our sentences contained one or two Downward-Entailing (DE) operators, that featured in different syntactic configurations. By manipulating both the number of DE operators and the syntactic environment in which they resided, we measured how RT is affected by these factors. We ran these experiments in English and in Hebrew, with different participant populations and different testing methods. Despite the linguistic subtlety of the theoretical issues, our results were remarkably sharp, leading to two firm conclusions: (i) that processing time is determined not by the number of DE operators, but rather, by the monotonicity of the minimal constituent in which they reside; (ii) that DE-ness is not a property of operators, but of environments. We show how our results bear directly on the current debate about the nature of monotonicity, which we describe below. Finally, we provide quantitative tests of alternative, non-semantic explanations, and show how our results do not support them.

1. Introduction

This paper presents the results of two experiments that provide new insight into the debate on the licensing conditions of Negative Polarity Items (NPIs). Specifically, we approached this issue from the perspective of processing by manipulating the number of DE operators as well as the syntactic configurations, and measuring Reaction Time (RT) instead of the standard judgment test. NPIs such as *any* and *ever* in English have interested linguists since the 1960s (Klima, 1964; Baker, 1970; Fauconnier, 1975; Ladusaw, 1980; Xiang et al., 2009; 2013; 2016; *passim*). These items must be licensed by some sort of negativity - a negative marker, such as *no* or *not*, or a downward-entailing (DE) operator, such as *less* and *few*.[1] Hence, the unlicensed NPIs in the sentences in (1a,c,e) result in ungrammaticality, but sentences (1b,d,f) are grammatical, as their NPIs are licensed by a negation (1b) or a DE quantifier (1d,f).

(1) a. *The kids ate *any* cookies at the birthday party.
b. The kids did not eat *any* cookies at the birthday party.
c. *More than 3 kids ate *any* cookies at the birthday party.
d. Less than 3 kids ate *any* cookies at the birthday party.
e. *Many kids ate *any* cookies at the birthday party.
f. Few kids ate *any* cookies at the birthday party.

More difficult to formulate are the precise licensing conditions. Everyone agrees that a DE operator is needed to license an NPI, yet the question is how the licensor and licensee are related. Homer (2021) divides theories of NPI licensing into two groups: an *Operator-Based Approach* (OpBA), by which all an NPI needs is to be in the scope of a DE operator, as stated in (2), and a stricter *Environment-Based Approach* (EnvBA), by which the environment needs to be DE, as in (3).

(2) <u>Operator-Based Approach (OpBA)</u>: An NPI is licensed only if it is in the scope of a downward-entailing (DE) expression. (Fauconnier, 1975; Ladusaw, 1980).

---

[1] Entailment, downward-entailing function and downward-entailing environment are defined as follows (for concreteness, we adopt Crnič's (2014) definition here):

(i) Cross-Categorical Entailment ($\Rightarrow$)
a. For p, q which are truth values: $p \Rightarrow q$ iff $p = 0$ or $q = 1$;
b. For f, g of type $\langle\sigma,\tau\rangle$: $f \Rightarrow g$ iff for all x of type σ, $f(x) \Rightarrow g(x)$.
(ii) Downward-entailing function
A function f is DE iff for any x and y in the domain of f such that $x \Rightarrow y$, $f(y) \Rightarrow f(x)$.
(iii) Downward-entailing environment
A constituent X is DE with respect to a sub-constituent Y of type α iff replacing Y with a variable of type α and binding it by a λ-abstractor adjoined as a sister of X yields a DE function.

Upward-entailing(UE) function and UE environment are defined symmetrically:
(iv) UE function
A function f is UE iff for any x and y in the domain of f such that $x \Rightarrow y$, $f(x) \Rightarrow f(y)$.
(v) UE environment
A constituent X is UE with respect to a sub-constituent Y of type α iff replacing Y with a variable of type α and binding it by a λ-abstractor adjoined as a sister of X yields a UE function.

(3) Environment-Based Approach (EnvBA): An NPI α is licensed in sentence *S* only if there is a constituent A of *S* containing α such that A is DE w.r.t the position of α. (Gajewski, 2005).

The OpBA claims that the licensing conditions of an NPI depend solely on its structural relation to the licensing DE operator. Nothing else is relevant. Homer contrasts this approach with the EnvBA, which argues that the DE-ness of the *syntactic environment* in which the NPI resides is the source of licensing. This DE-ness, moreover, may not be necessarily computed in whole sentences, but rather, in constituents of sentences, which would constitute the syntactic environment necessary for licensing.

To adjudicate between the OpBA and the EnvBA empirically, evidence is required. In this part of the paper, we present arguments that rely on plausibility and acceptability judgments. Here, a sentence is needed, in which an NPI is in the scope of a DE operator, but at the same time, the relevant environment is non-DE. If the sentence is acceptable, the DE-ness of that environment may not matter, as the OpBA would have it; yet if the sentence is unacceptable, then the DE-ness of the environment does matter, as the EnvBA suggests.

Homer (2021) provides evidence from French in favor of the EnvBA. He first presents evidence that *quoi que ce soit* (= anything) is a (weak) NPI, licensed by a negative adjective (4) or a negation (5) :

(4) Il est **im**possible que Jean ait fait *quoi que ce soit* pour aider la Mafia.
it is impossible that Jean have.SUBJ done what that this be.SUBJ to help the Mafia
"It is impossible that Jean did anything to help the Mafia."

(5) Il **n'** est **pas** possible que Jean ait fait *quoi que ce soit* pour aider la Mafia.
It NE is NE possible that Jean have.SUBJ done what that this be.SUBJ to help the Mafia
"It is not possible that Jean did anything to help the Mafia."

The acceptability of (4)-(5) is expected by the OpBA, because both contain a DE operator that can serve as licensor. If so, then the presence of *both* licensors must certainly license the NPI. Yet Homer points out that in "flip-flop" sentences, an NPI is anti-licensed:

(6) *Il **n'** est **pas im**possible que Jean ait fait *quoi que ce soit* pour
It NE is NE impossible that Jean have.SUBJ done what that this be.SUBJ to
aider la Mafia.
help the Mafia (Homer, 2021)
Intended: "It is not impossible that Jean did anything to help the Mafia."

Homer shows how this unacceptability follows from the EnvBA. The intuition is that one cannot find in this sentence a DE environment that contains *quoi que ce soit* and only one of **pas** and **im**possible (which are both in the same TP). These two operators (each individually licensing the NPI) are coupled in this configuration, and thus NPI licensing fails. Homer proposes that this failure is because licensing here may only be done by the node that dominates both of them, which is Upward-Entailing, as the combination of the two DE operators **n'...pas** and **im**possible leads to UE-ness. Therefore, *quoi que ce soit* resides in an Upward-Entailing (UE) environment, and unacceptability follows.

The EnvBA succeeds because it requires that licensing is dependent on a DE environment, not a DE operator. Homer calls this environment "the domain of an NPI":

(7) Domain of an NPI: A constituent $\boldsymbol{\gamma}$ which contains the NPI π is a domain of π if and only if the acceptability of π can be evaluated in $\boldsymbol{\gamma}$.[2]

What is the domain of an NPI? Homer proposes that PolP (Polarity Phrase) is the smallest but not the only domain of *quoi que ce soit*. Since (6) is ungrammatical, *n'...pas* and *impossible* must be in the same domain there; on the other hand, in the acceptable (8), the configuration contains a domain with the NPI and a single licensor. He further points out that in (9), both the conditional-*if* and *impossible* are DE, and *quoi que ce soit* is nonetheless licensed. He concludes that these

[2] Homer (2021) defines *acceptability* of an NPI as: an NPI π is acceptable in a constituent $\boldsymbol{\gamma}$ if and only if $\boldsymbol{\gamma}$ has the appropriate monotonicity w.r.t. the position of π in $\boldsymbol{\gamma}$.

two DE operators cannot be in the same minimal domain, and therefore, there must be a domain below CP, but above VP, where *quoi que ce soit* is licensed. Homer calls this phrase PolP.[3]

(8) Il est **im**possible que Jean **n'** ait **pas** fait *quoi que ce soit* pour
It is impossible that Jean NE have.SUBJ NEG done what that this be.SUBJ to
aider la Mafia.
help the Mafia
"It is impossible that Jean didn't do anything to help the Mafia."

(9) **S**'il est **im**possible que Jean ait fait *quoi que ce soit* pour aider la mafia, je lui présenterai mes excuses.
'If it is impossible that Jean did anything to help the Mafia, I will apologize to him.'

Homer's LFs of (4) - (9) are shown below. They demonstrate the relevant domains (i.e. PolP and TP) and their monotonicity with respect to the NPI *quoi que ce soit*. Subscripted arrows indicate the monotonicity of the domains with respect (⇧ denotes an UE domain and ⬇ denotes a DE domain):

(4)' …[$_{\text{TP.}\Downarrow}$ T [$_{\text{PolP.}\Downarrow}$ **im**possible [$_{\text{CP}}$ que Jean T [$_{\text{PolP.}\Uparrow}$ [quoi que ce soit]$_1$ faire t_1]]]]

(5)' …[$_{\text{TP.}\Downarrow}$ T [$_{\text{PolP.}\Downarrow}$ **pas** possible [$_{\text{CP}}$ que Jean T [$_{\text{PolP.}\Uparrow}$ [quoi que ce soit]$_1$ faire t_1]]]]

(6)' * …[$_{\text{CP}}$ [$_{\text{TP.}\Uparrow}$ T [$_{\text{PolP.}\Uparrow}$ **pas** **im**possible [$_{\text{CP}}$ que Jean T [$_{\text{PolP.}\Uparrow}$ [quoi que ce soit]$_1$ faire t_1]]]]

(8)' …[$_{\text{TP.}\Uparrow}$ T [$_{\text{PolP.}\Uparrow}$ **im**possible [$_{\text{CP}}$ que Jean T [$_{\text{PolP.}\Downarrow}$ **pas** [quoi que ce soit]$_1$ faire t_1]]]]

(9)' …[$_{\text{TP.}\Uparrow}$ [$_{\text{CP.}\Uparrow}$ **Si** T [$_{\text{PolP.}\Downarrow}$ **im**possible [$_{\text{CP}}$ que Jean T [$_{\text{PolP.}\Uparrow}$ [quoi que ce soit]$_1$ faire t_1]]]] T]

By incorporating the concept of domain, Homer derives a more refined version of the Environment-based licensing condition for NPI:

(10) Environment-based Licensing Condition for NPIs (refined): An NPI α is licensed in sentence *S* only if α has a DE domain in *S*.

---

[3] Homer (2021) proposes that TP is another domain of NPI in French, given that the following sentence, which contains two DE operators – conditional-*if* and 'at most five', is grammatical. Assuming that 'at most five' sits in [Spec, TP] and conditional-*if* sits in [Spec, CP], there must be a domain containing only 'at most five' but not conditional-*if*, namely, TP.
(i) Si au plus cinq personnes ont fait quoi que ce soit pour aider la Mafia, nous sommes sauvés.
'If at most five people did anything to help the Mafia, we are good.'

The data that Homer (2021) provides strong support to the EnvBA. An operator-based approach cannot explain why (6) is unacceptable, as the NPI is in the scope of a DE operator.

We approach the issue of DE domains from a processing perspective. That is, we present experiments that measure continuous variables - mainly Reaction Time (RT), and explore their relation to the number of DE operators, as well as the domain in which they are contained. We ask whether the processing of DE domains is more costly than UE, i.e., whether the processing of monotonicity is domain-dependent, as would be the case if licensing and processing cost are related:

(11) *Domain-based Processing Hypothesis* (DPH):

- Parsing is bottom up
- A minimal domain is UE by default
- The monotonicity reversal of a domain of an NPI (MRD) incurs a processing cost

DPH states that, assuming parsing is in a bottom-up fashion (i.e. lower domains are integrated into the higher domains) and the default monotonicity of the minimal domain is UE, every monotonicity reversal of an NPI (MRD) incurs a processing cost. To illustrate, in (5)', MRD occurs once when the embedded PolP integrates into the matrix PolP; in (8)', MRD occurs twice: at first, in the embedded PolP, the monotonicity reverses to DE, and then, when the embedded PolP integrates into the matrix PolP, the monotonicity reverses again, resulting in a UE domain. According to DPH, the processing cost (hence time) of (8)' is expected to be greater than that of (5)'.[4] Note that, according to DPH, if monotonicity reversal occurs twice between two domain boundaries, then they amount to none, as the case in (6)'.

In the current study, we provide quantitative processing evidence supporting the DPH, from two cross-linguistic experiments which substantiate EnvBA: one is a lab experiment with native Hebrew speakers, and the other is an online experiment with native English speakers. We leverage

[4] Note (5)' and (8)' are not a minimal pair because (8)' contains one more word than (5)', which is very likely to be a cause of longer processing time.

the well-studied *monotonicity effect* – DE operators incur added processing cost in verification tasks, relative to their UE counterparts (Just and Carpenter, 1971; Clark and Chase, 1972; Deschamps, Agmon, Loewenstein & Grodzinsky, 2015; Agmon, Loewenstein & Grodzinsky, 2019; Schlotterbeck et al. 2020). We use two DE operators, *not* and *less*, to construct two different structures of double negation: the *intra-domain structure* and the *cross-domain structure*. In the *intra-domain structure*, there is no domain containing only one DE operator. Namely, both of the eligible domains contain either zero or two DE operators (e.g. *Not less than half of the circles are blue*). In the *cross-domain structure*, there is one domain which contains only one DE operator (e.g. *Less than half of the circles are not blue*). We measure reaction times (RTs) for these, as well as for their *more* counterparts in the paradigm of verification tasks. To forecast, in both experiments, participants responded faster to the sentence of the intra-domain structure than those of the cross-domain *stru*cture, as predicted by DPH. Moreover, many participants struggled to comprehend the sentence with the cross-domain structure, as evinced by their elevated error rates.

2. Experimental materials in Hebrew

When the Hebrew NPI *'ey-pa'am* (ever) replaces *quoi que ce soit*, Hebrew mimics French (cf. Appendix 1 for Hebrew replications of Homer's effects). In the current study, we test the flip-flop effect in Hebrew with monoclausal sentences, instead of the embedded clausal structures that Homer provided. We use two DE operators, *less* and *not*, to build two double negative structures, (12d) and (12f), to contrast with sentences including no or one negation, (12a, b, c, e). Note that no NPIs are present, because the experiment is about the processing of multiple negations and domains. Therefore, all sentences are acceptable.:

(12) a. [$_{\text{TP.}\Uparrow}$ [Yoter mi-xezi me-ha-igulim] [T [$_{\text{PolP.}\Uparrow}$ hem kxulim]]] .
more than-half of-the-circles are blue
'More than half of the circles are blue' (0 MRD)

b. [$_{\text{TP.}\downarrow}$ [**Paxot** mi-xezi me-ha-igulim] [T [$_{\text{PolP.}\Uparrow}$ hem kxulim]]] .
less than-half of-the-circles are blue
'**Less** than half of the circles are blue' (1 MRD)

c. [$_{\text{TP.}\downarrow}$ [**Lo** yoter mi-xezi me-ha-igulim] [T [$_{\text{PolP.}\Uparrow}$ hem kxulim]]].
Not more than-half of-the-circles are blue

‘**Not** more than half of the circles are blue’ (1 MRD)

d. [$_{\text{TP.}\Uparrow}$ [**Lo paxot** mi-xezi me-ha-igulim] [T [$_{\text{PolP.}\Uparrow}$ hem kxulim]]].
Not less than-half of-the-circles are blue
‘**Not less** than half of the circles are blue’ (intra-domain structure) (0 MRD)

e. [$_{\text{TP.}\downarrow}$ [Yoter mi-xezi me-ha-igulim] [T [$_{\text{PolP.}\downarrow}$ hem **lo** kxulim]]] .
more than-half of-the-circles are not blue
‘More than half of the circles are **not** blue’ (1 DMR)

f. [$_{\text{TP.}\Uparrow}$ [**Paxot** mi-xezi me-ha-igulim] [T [$_{\text{PolP.}\downarrow}$ hem **lo** kxulim]]].
less than-half of-the-circles are not blue
‘**Less** than half of the circles are **not** blue’ (cross-domain structure) (2 MRD)

To derive the predictions from DPH, we compare how many times MRD occurs in each sentence in the relevant minimal pairs. We have three sets of predictions: (i) $RT_{\textit{not less}} < RT_{\textit{not more}}$ because *not less* includes zero MRD (both PolP and TP are UE), and *not more* includes one (PolP is UE but TP is DE); (ii) $RT_{\textit{less…not}} > RT_{\textit{more…not}}$ because MRD occurs twice in *less…not* (the minimal domain PolP reverses from UE to DE, and then TP becomes UE again) but only once in *more…not* (only MRD only occurs in PolP); (iii) $RT_{\textit{less…not}} > RT_{\textit{not less}}$, as *less…not* has more MRDs. Note that the prediction (i) has been testified previously in a study which adopted a similar paradigm as the current study ($RT_{\textit{not less}}$ (926.9ms) < $RT_{\textit{not more}}$ (987.4ms) and $RT_{\textit{more}}$ (804ms) < $RT_{\textit{less}}$ (888.2ms) in Tan et al. (2023)).

One past study roughly depicts a similar picture (Sherman, 1973, 1976), who found that sentences with the double negative *no one doubted* were easier than those with the single negative *doubted* and were equally difficult to sentences with just *no one* ($RT_{\textit{no one doubted}} < RT_{\textit{doubted}} = RT_{\textit{no one}}$; cf. Schlotterbeck, 2017; Bott et al., 2019, for related works). Sherman suggested an account that was very much in the spirit of DPH: he proposed that subjects might mentally combine *no one* and *doubted* to form an affirmative, which are in the same domain. Interestingly, his design contained at least one other condition with these two negatives, namely, sentences containing *doubted...not, which are across two different domains*. And as Table 2 (p. 148) in the paper shows, $RT_{\textit{no one doubted}}$ < $RT_{\textit{doubted..not}}$, despite the fact that both had the same monotonicity value in the matrix clause. Note that *no one doubted* is an intra-domain double negation structure and *doubted…not* is a cross-domain structure. Sherman's results seem to match our prediction regarding the structural difference.

Before we move on to the experiment, we explore the question of whether the effects of the EnvBA and the DPH may be due to linear adjacency. That is, whether two DE operators are perceived as a single UE operator by virtue of their immediate proximity. The *flip-flop* effect in (6), (8) and (9) is consistent with this perspective: the NPI *quoi que ce soit* is anti-licensed when the two DE operators are adjacent (6), but licensed when they are separated (8), (9). To distinguish between domain and linearity, we need an example in which two DE operators are adjacent, but no *flip-flop* effect occurs (or vice versa). Homer shows this for French (13a), and (13b) demonstrates the same effect for Hebrew: (13b) contains two DE operators, *'im* 'if' and *le-xol-ha-yoter xamiša* 'at most five people', and an NPI *ey pa'am* 'ever'.[5] (13a) and (13b) are both grammatical while the two DE operators are adjacent to each other, just like the case in (6). Thus, linear adjacency cannot account for the *flip-flop* effect of (6).

(13) a. **Si au plus cinq personnes** ont fait quoi que ce soit pour aider la Mafia,
If at most five people have done what that this be.SUBJ to help the mafia,
nous sommes sauvés.
we are saved
'If at most five people did anything to help the Mafia, we are good.'

b. **'im le-xol-ha-yoter** xamiša 'anašim siy'u ' ***ey pa'am*** la-mafia, 'anaxnu beseder
if at-most five people assisted ever the Mafia, we good
"If at most five people ever assisted the Mafia, we are good."

3. Experiment I

[5] The distance between *'im* and *le-xol-ha-yoter xamiša* does not interfere with NPI licensing: even when the two DE operators are separated from each other, the sentence is still grammatical, as shown below:
(i) 'im tagid li še- le-xol-ha-yoter xamiša 'anašim siy'u 'ey pa'am la-mafia, 'ani 'eda še- 'ata
If you-tell me that at-most five people assisted ever the mafia I will-know that you
mešuga
crazy
"If you tell me that at most five people ever assisted the mafia, I will know that you are crazy."

Experiment I was a lab study, implemented in Hebrew. We used two DE operators in Hebrew, *lo* ’not’ and *paxot* ‘less’, along with *yoter* ‘more’, to juxtapose between OpBA and EnvBA. We asked the participants to verify the aforementioned sentences against pictures in the paradigm of speeded sentence picture verification task. According to EnvBA and DPH, we would expect to observe $RT_{\textit{not less}} < RT_{\textit{not more}}$ but $RT_{\textit{less not}} > RT_{\textit{more not}}$ based on the number of MRD. In contrast, if OpBA is correct, we would observe $RT_{\textit{not less}} > RT_{\textit{not more}}$ and $RT_{\textit{less not}} > RT_{\textit{more not}}$ instead since both double negation structures contain one more DE operator than their *more* counterparts.

- Materials

We used four Sentence Types along with the pair of quantifiers, *more* and *less*, to build up our sentential stimuli. As shown in Table 1, there are eight conditions in total. The Sentence Type *plain* is used as a baseline for comparison with other conditions and as a sanity check to make sure the participants are doing the task dutifully. The Sentence Type *intra-domain not* and *cross-domain not*, as discussed in section 2, are the ones which include two DE operators in different configurations. Finally, the Sentence Type *that*-clause was added to counterbalance the number of stimuli with *not* so we would have an equal number of stimuli with and without *not.* The sentences were recorded in Hebrew by a male native speaker, and processed in Audacity to minimize their variations in terms of pitch, amplitude and length. However, since the sentences in *that*-clause contain much more words than the other groups, we did not match their length with others (4490 msec vs. 3350 msec), while within group uniformity was controlled.

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">Factor 2: Quantifier Type</td></tr>
<tr><td>more</td><td>less</td></tr>
<tr><td rowspan="2">Factor 1: Sentence Type</td><td>plain</td><td>More than half of the circles are blue</td><td>Less than half of the circles are blue</td></tr>
<tr><td>intra-domain not</td><td>Not more than half of the circles are blue</td><td>Not less than half of the circles are blue</td></tr>
</table>

	cross-domain *not*	More than half of the circles are not blue	Less than half of the circles are not blue
	***that*-clause**	It is true that more than half of the circles are blue	It is true that less than half of the circles are blue

Table 1 Experimental design

As for the testing images, we used a set of images in which the blue and yellow circles are arranged in a 5 x 5 array. The ratios between the blue and yellow circles are 5:20, 10:15, 15:10 and 20:5. To add variety, for each ratio, there were two types of arrangement of location of the circles. The circles of the same color clustered together to keep the verification task simple.

In order to counterbalance all the conditions, we needed 128 trials (4 Sentence Type x 2 Quantifier Type x 2 Referred Color x 4 Ratio x 2 Arrangement = 128 trials). The Truth Value (true / false) was thus counterbalanced accordingly. Each combination of the trial appeared twice in the experiment. To eliminate the potential learning effect over time, we randomized the order of the trials for each subject.

- Procedure

Before the start of the experiment, the experimenter explained the task to the participants. The participants were told that they would hear a sentence stating the relationship between circles in two colors, and then an image would appear which comprised circles in the two colors. The participants were asked to determine whether the sentence they heard matched the image or not. They were also asked to respond as fast and as accurate as possible. Each experiment included two blocks, and each block contained 16 practice trials and two runs, each of which contained 64 trials. Each subject responded to 32 practice trials and 256 experimental trials in total. To reduce the difficulty of the task, each block contains only two groups of Sentence Types. Considering the occurrence of *not*, two types of grouping are possible: one is [*plain* + *intra-domain not*] v.s. [*that-clause* + *cross-domain not*]; the other is [*plain* + *cross-domain not*] v.s. [*that*-clause + *intra-domain not*]. Because in the first type of grouping, the combination *plain* and *intra-domain*, was the same combination as the stimuli in the previous study (Tan et al., 2023), we decided to assign most of the participants to the second type of grouping and only five participants to the first type of grouping. All the relevant sentences and a few examples of images were shown to the participants before the beginning of each block. The participants were given a break between every two runs.

In each trial, the participants first saw a fixation cross on the screen. After 800 ms, they heard a sentence from the headset while the fixation cross remained on the screen. Each sentence was either 3350 ms long or 4490 ms long. Right after the end of the sentence, the participants saw an image in the middle of the screen. They were allowed to respond between 300 ms - 5000 ms after the onset of the image.[6] Once the participants responded, the trial terminated and the next trial would start. If a participant did not respond within 5000 ms, the response would be recorded as "miss" and the next trial would start. The participants responded by pressing the right arrow key (TRUE) or the left arrow key (FALSE) on the keyboard. The RTs were measured from the onset of the image until the press of a key.

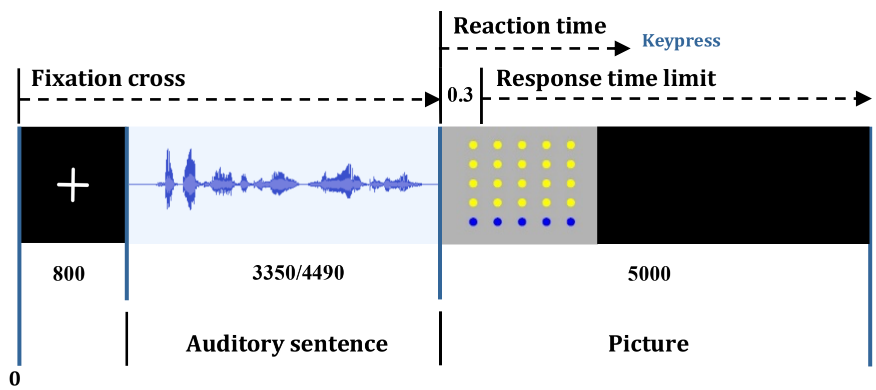


Figure 1 Timeline of Experiment I

- Participants

23 students (12 male and 11 female), aged 26 ± 2.9 (mean ± standard deviation), native Hebrew speakers. All the participants signed an informed consent form approved by the Hebrew University Research Ethics Committee. They received either payment or credits for participating in the experiment.

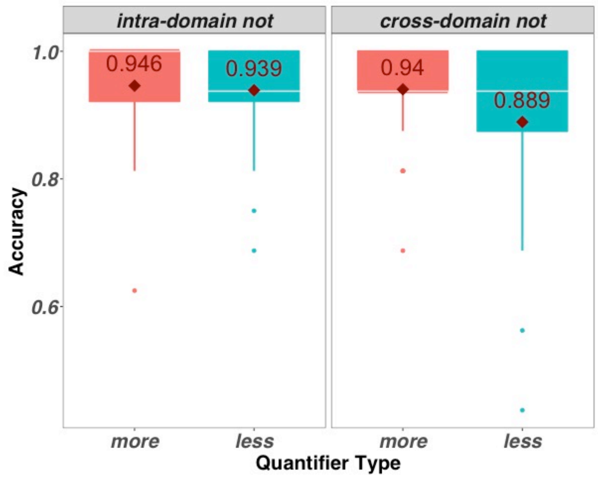
[6] In the pilot runs, we started with 1900 ms of response time limit. However, the participants had very poor accuracy in *less...not* condition. Even after we inserted a 500 ms-break between the audio stimulus and the picture, and prolonged the response time limit to 2200 ms, the accuracy of *less...not* condition still remained lower than 70%. In order to curtail the difference in accuracy between different conditions, we eventually decided to prolong the response time limit to 5000 ms.

- Results and analyses

The data of 22 participants were included (only one participant's data, of which the average accuracy in at least one run was below 75%, was excluded; the average accuracy was 95.12% after screening). In the error domain, within each Sentence Type, the *less* conditions showed lower average accuracy than their *more* counterparts. We show the data of *intra-domain not* and *cross-domain not* in Fig. 2, where the mean accuracy of each condition is marked in dark red in each boxplot (the rest of the data: $accuracy_{more}$ = 99.23%; $accuracy_{less}$ = 96.92)[7]. Among all the conditions, the participants had the lowest average accuracy in *less...not* (88.9%), and the data points also spread wider than in the other conditions.

When analyzing the RT data, only correct responses were taken into account (since it is unclear what cognitive processes are involved in erroneous responses). The data of *plain* and *intra-domain not* parallel with the findings in the previous study (Deschamps et al., 2015; Tan et al., 2023): $\overline{RT}_{more}$ (1106.8ms) < $\overline{RT}_{less}$ (1389.0ms) and $\overline{RT}_{not\ more}$ (1594.8ms) < $\overline{RT}_{not\ less}$ (1510.7ms). As for the pair of *cross-domain not*, it was observed that $\overline{RT}_{more..not}$ (1639.9ms) < $\overline{RT}_{less...not}$ (1885.4ms). The details of the data of *intra-domain not* and *cross-domain not* are exhibited in Fig. 3 with the means marked in dark red in each boxplot. By scrutinizing the RT data, we notice that: (i) the *less…not* was the most costly condition in terms of RT; (ii) within every pair of Sentence Type, except *intra-domain not*, *less* condition took longer to verify than their *more* counterpart. It is noteworthy that even though both *less… not* and *not less* contain two DE operators, our experimental result suggests that the former is much more difficult to process than the latter. The evidence includes longer RT, lower accuracy, and bigger variance of accuracy in the former than the latter.

[7] The Sentence Types that-clause are omitted in the figures and the following analyses because they were for sanity check and for counterbalancing the number of stimuli without *not*, respectively. For those who are curious, the data is close to sentence type plain: $accuracy_{it\ is\ true\ that\ more}$ = 97.44%; $accuracy_{it\ is\ true\ that\ less}$ = 95.88%; $RT_{\text{it is true that more}}$ (1109.5ms) < $RT_{\text{it is true that less}}$ (1364.8ms).



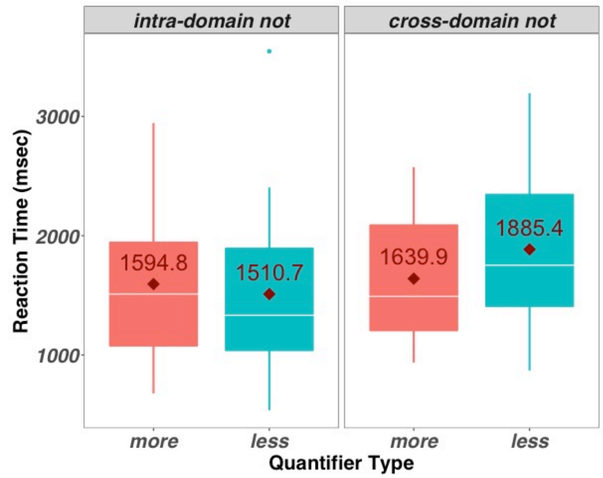


Figure 2 Boxplots of the accuracy data broken down by Sentence Type and Quantifier Type (coral = *more*; and turquoise = *less*). Group distribution statistics are provided as boxes, horizontal midlines and whiskers depicting interquartile range (IQR, defined as first quartile (Q1) to third quartile (Q3)), medians, minimums (≥ Q1 - 1.5*IQR), and maximums (≤ Q3 + 1.5*IQR). In each boxplot, the dark red diamond shows the group mean along with the number. Generally the accuracies are quite high. Only the accuracy of *less...not* condition is slightly lower than 90%.

Figure 3 Boxplots of the original RT data broken down by Sentence Type and Quantifier Type (coral = *more*; and turquoise = *less*). Group distribution statistics are provided as in Fig. 2. Pairwise, *not more* took longer time to process than *not less*; but *more...not* required shorter time to process than *less...not*. Generally, the subjects spent longer time to process *cross-domain not* than to process *intra-domain not.*

To further test the difference between *not less* and *less…not*, we fitted a linear mixed effects model to the data of the *intra-domain not* and *cross-domain not* with the logarithmic transformation of RT as the dependent variable (using R and lme4, Bates et al. 2015). Quantifier Type, Sentence Type and Truth Value and their interactions were used as the fixed effect factors. As random effects, we had an intercept for participants as well as by-participant slope for the effect of Sentence Type. P-values were derived via the lmerTest package in R (Kuznetsova et al. 2017). Visual inspection of residual plots revealed no obvious deviations from homoscedasticity or normality. The statistics are shown in Table 2. Among the statistical results, there are two effects relevant to our research questions: first, there was a main effect of Sentence Type ($t = 4.136$, $p = 0.0004$), as manifested by the fact that the mean RTs of *cross-domain not* were on average higher than those of *intra-domain not.* Secondly, there was a strong significant interaction effect between Quantifier Type and Sentence Type ($t = -6.033$, $p < 0.0001$), indicating that the two double negative structures had very different patterns when compared to their *more* counterpart, respectively.

From the averaged data, we observe that $\overline{RT}_{less...not} > \overline{RT}_{more...not}$, and $\overline{RT}_{not\ less} < \overline{RT}_{not\ more}$. From the results of the mixed effect model, we learned that processing of the pair of *intra-domain not* differs from the processing of the pair of *cross-domain not*, in terms of RT.

Table 2 Summary of the mixed-effect regression model for the data of Experiment I.[8]

	Estimate	**Std. Error**	**t value**	**p-value**
(Intercept)	7.234	0.07529	96.086	< 0.0001***
QType1	-0.0211	0.00913	-2.312	0.02086*
SentenceType1	0.09102	0.02201	4.136	0.00043***
TruthValue1	0.05639	0.00912	6.186	< 0.0001***
QType1*SentenceType1	-0.05506	0.00913	-6.033	< 0.0001***
QType1*TruthValue1	-0.02155	0.00912	-2.363	0.01818*
SentenceType1*TruthValue1	0.00941	0.00912	1.032	0.30200
QType1*SentenceType1*TruthValue1	0.0346	0.00912	3.796	0.00015***

○ Discussion

In Experiment I, we tested four Sentence Types with two Comparative Quantifiers in Hebrew, resulting in eight conditions in total, including *more* / *less* (*plain*), *not more* / *less* (*intra-domain not*), *more* / *less ...not* (*cross-domain not*) and *it is true that more* / *less* (*that-clause*). We are particularly interested in the contrast between participants' RTs in the *cross-domain not* condition and those in *intra-domain not*. Even though both *less...not* and *not less* contain two DE operators,

[8] "contr.sum" was adopted as the contrast scheme in the models for this experiment.

the processing difficulty of the two conditions is very different. Our results show that *less...not* had significantly longer RTs and lower accuracy rate than *more...not*, indicating that it was highly demanding.[9] By contrast, the *not less* condition yielded slightly shorter RTs and slightly lower accuracy rates than *not more*, and was hence not more difficult to process (perhaps even easier).

The difference between *less.. not* and *not less* is further corroborated by the results of the linear mixed-effects model when comparing *cross-domain not* and *intra-domain not*, which is also in line with the results of Sherman's study (1976). A related study (Tan et al., 2023) also showed two DE operators in the same domain (i.e. *not less*) facilitate the processing of a sentence, suggesting that the two DE operators are counted as a single UE one.

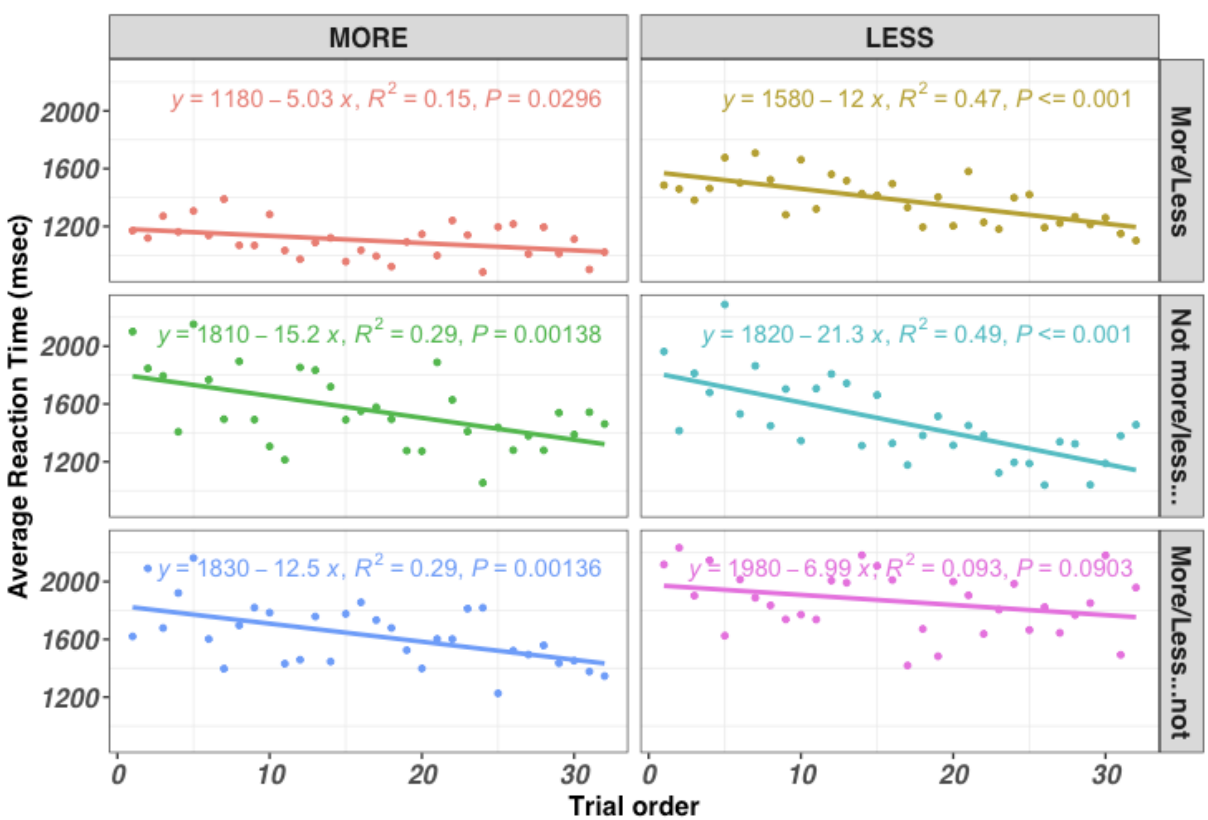
Thus, the processing difference between the *intra-domain* (*less… not*) and *cross-domain* (*not less*) conditions indicates that the processing of two DE operators is not determined by the number of DE operators, suggesting that it is domain-based. As *domain* is a syntactic concept, we conclude that syntax is an important determinant of the processing difficulty of monotonicity. That is, when two DE operators are in the same domain, their processing cost is UE, but when they are not in the same domain, they cannot be integrated into one during processing and cause MRD, which manifests in elevated cost.

- Learning effect

The account above works, but alternatives must be considered and ruled out. One possibility we explored was that the effects obtained were due to selective learning. That is, the differences we observed were because participants' performance improved over the course of the testing session, in a manner that differentiated between the conditions. RTs typically go down in the course of a testing session, and one possibility is that the effect we obtained is due to differential *learning* (some participants reported that they realized that responding is faster if *not less* is converted into *more* during a trial). To see whether this response strategy was indeed used and had a substantial effect on our results, we fitted four regression models for four relevant comparisons, including

[9] A contrast analysis between $RT_{more...not}$ and $RT_{less...not}$ shows that they are significantly different from each other ($t = 4.828$, $p < 0.0001$). On the other hand, there is no significant difference between $RT_{not\ more}$ and $RT_{not\ less}$ ($t = -2.246$, $p = 0.2169$).

<*more*, *less*>, < *not more, not less*>, <*more…not*, *less…not*> and <*not less*, *less…not*>.[10] We first calculated the cross-participant mean RT_n in trial n ($1 \leq n \leq 32$) in each condition (Fig. 4). Then, we fitted a regression model to each comparison, which included RT as the dependent variable, Trial Order as one predictor, either Sentence Type (for <*not less*, *less…not*>) or Quantifier Type (for the other three comparisons) as a second predictor, and their interaction term. The interaction term only came out significant in the model of <*not less*, *less…not*> (p = 0.0157), indicating that the learning rates were only distinguishable between *not less and less…not*, but not in the other three comparisons. The result suggests that the configurational difference between *not less* and *less...not* influences our cognitive functioning not only in terms of processing but also in terms of learning. However, learning cannot explain why the pair of *intra-domain not* exhibited a different pattern from the pair of *cross-domain not*, namely, $RT_{not\ more} > RT_{not\ less}$ but $RT_{more…not} < RT_{less…not}$.



[10]The other comparisons are not to our interest since they are not minimal pairs, e.g. <*more*, *less…not*>, <*not less*, *more … not*>, etc.

Figure 4 The time course of mean RTs in each condition. The x-axis indicates the order of a trial among all the trials of the same condition. ($1 \leq x \leq 32$, $x \in N$). Only correct responses were taken into account for the mean (86.96% of data). Regression lines were added to show the trend.

4. Experiment II

Experiment II was a replication attempt of Experiment I, using the exact same conditions and structure, with three differences: (a) participants were recruited on the Internet and the experiment was conducted online; (b) they were native speakers of English, not Hebrew; (c) they constituted a larger group. As will become clear, this experiment further solidifies our results with a large number of participants.

- Participant Recruitment

We recruited native English speakers online from Prolific's participant pool. Participants were redirected from Prolific.ac to an online experiment hosted on PCIbex Farm. They were promised a monetary reward of £8.00 per hour (which was almost twice the time the test took on average, resulting in an average reward of £4.7). They were also told that completing the study at an overall accuracy of $\geq 95\%$ would award them a bonus - completion of their reward to £8.

- Accuracy-based Screening
  1) Upon registration to the study, as well as while signing the consent form, participants were told: "you are required to meet a 90% accuracy threshold. Keep in mind that you may be rejected in the middle of the experiment, due to unsatisfactory accuracy levels. You have 50 minutes to complete the experiment".
  2) At the beginning of each run, participants were reminded of the bonus and its requirements.
  3) At the experiment's half point, participants got feedback containing their response accuracy rates, along with a comment on how much closer it brings them to the bonus. Participants whose accuracy rate was < 85% were excluded at this point. Eighty four participants moved past the half point at the required pace and accuracy, completing the test at an average of 35 minutes. Of these, six exceeded the time limit, or were below the

overall 90% accuracy threshold at the experiment's end; they were excluded, but paid. Thirty-three (~ 40%) performed at a level that won them a bonus.

○ Procedures

After a general explanation about the experiment, each of the two experimental blocks began with a screen containing a graphic representation of a trial's time-course, accompanied by an explanation of the task, and an invitation to perform a single trial. On the next screen, a table with all 8 sentences to be heard in this block were displayed. This was followed by a practice session containing 16 trials (equally representing all conditions) with feedback ("Correct!", "Wrong!", and "Too slow… please try to respond faster"). Each trial in both the practice and the experimental sessions was accompanied by a display of the keys representing match and nonmatch. This long preparation phase helped reduce errors, with no bias.

The trial structure was very similar to the one in Experiment I, only with slight time differences in order to adapt to the language differences. Details are depicted in Fig. 5.

After receiving feedback at the experiment's middle point, there was a forced 2 minute break (which the participants could choose to extend to 5 minutes). A forced 1-minute break was also given between the two experimental sessions in each block.

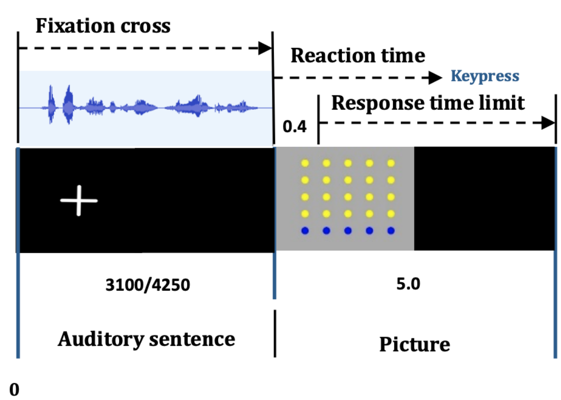


Figure 5 Timeline of Experiment II

○ Results and Analysis

The data of the 78 participants who completed the experiment were now subjected to more stringent screening criteria: admitted to analysis were only those participants who not only performed at 90% and below 50 minutes, but also, at a level of 75% correct on *each* condition (overall mean accuracy: 95.21%). There were no misses. This screening left us with 70 participants. Next, we moved from the error domain to the time domain. We analyzed the RTs of the surviving 70 participants, omitting incorrect responses (4.49% of the data). We found that the *less* condition in each Sentence Type took longer to process than its corresponding *more* condition, except in *intra-domain not*, where the RTs of the two Quantifier Types approximate to each other. The mean RTs in *intra-domain not* and *cross-domain not* are exhibited in Fig. 6 (the mean RTs are marked in dark red in each boxplot.).[11] Despite the difference in experimental method and language, the pattern between the Quantifier Types in each Sentence Type resembles what we found in Experiment II, as shown in Fig. 6. Regarding the two double-DE operator conditions, it seems that *less...not* is much more taxing to process than *not less* because both of the mean and the median of the former are much higher than those of the latter.

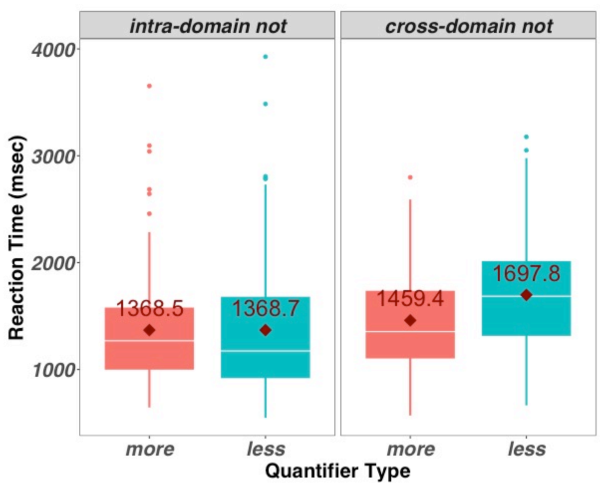


Figure 6 Boxplots of the RT data from the subjects whose accuracies in each condition was over 75%, broken down by Sentence Type and Quantifier Type (the data of 70 subjects are included). Group distribution statistics are provided in the same way as in Fig. 2. The mean $\mathrm{RT}_{\textit{not more}}$ is on a par with the mean $\mathrm{RT}_{\textit{not less}}$ while the mean $\mathrm{RT}_{\textit{more...not}}$ is smaller than the mean $\mathrm{RT}_{\textit{less...not}}$.

To test the significance of the difference between *not less* and *less…not*, we fitted a linear mixed effects model to the data of *intra-domain not* and *cross-domain not* with the logarithmic transformation of RT as the dependent variable (using R and lme4, Bates et al. 2015). Quantifier Type, Sentence Type and Truth Value and their interaction were used as fixed effects. The two

[11] The other mean RTs are : $\mathrm{RT}_{\textit{It is true that less}}$ (1144.8ms) > $\mathrm{RT}_{\textit{It is true that more}}$ (960.5ms); $\mathrm{RT}_{\textit{less}}$ (1112.8s) > $\mathrm{RT}_{\textit{more}}$(947.7ms);

random-effect terms include an intercept of subjects and a random slope of Sentence Type. P-values were derived via the lmerTest package in R (Kuznetsova et al. 2017). Similar to the result in Experiment I, there was a significant main effect of Sentence Type ($t = -4.719$, $p < 0.0001$) as well as an interaction effect between Quantifier Type and Sentence Type ($t = -11.486$, $p < 0.0001$). The results here replicate the results from Experiment I, indicating that the pattern of *cross-domain not* exhibits a quality different from that of *intra-domain not.*

Table 3 Summary of the mixed-effect regression model for the data of Experiment II

	Estimate	**Std. Error**	**t value**	**p-value**
(Intercept)	7.140	0.03322	214.949	< 0.0001***
QType1	-0.0269	0.00476	-5.648	< 0.0001***
SentenceType1	-0.08468	0.01794	-4.719	< 0.0001***
TruthValue1	0.04046	0.00476	8.498	< 0.0001***
QType1*SentenceType1	0.0547	0.00476	11.486	< 0.0001***
QType1*TruthValue1	-0.00854	0.00476	-1.793	0.07298
SentenceType1*TruthValue1	-0.01516	0.00476	-3.184	0.00146
QType1*SentenceType1*TruthValue1	-0.02213	0.00476	-4.648	< 0.0001***

The online experiment along with the meticulous reward mechanism enabled us to collect a huge amount of high-quality data, providing us cross-language evidence which supports the *Domain-based Processing Hypothesis*. We observed the same contrast between *not less* and *less…not* among the English speakers as among the Hebrew speakers that *less...not* is much more taxing than *not less*. The fact that we found the same effect in English reinforces our argument that the configuration of a sentence which contains two DE operators determines the way how we process it.

○ Learning effect

In the same manner as in Experiment I, we calculated the cross-participant mean RT in each trial in each condition for the data in Experiment II, as shown in Fig. 7. Likewise, we fitted four regression models to check the difference between the learning rates in four comparisons, including <*more*, *less*>, < *not more, not less*>, <*more…not*, *less…not*> and <*not less*, *less…not*>. In each model, RT was the dependent variable, with Trial Order as one predictor, either Sentence Type or Quantifier Type as another predictor, and their interaction term. Among the four comparisons, we found no significant interaction effect, indicating that the learning rates were not distinguishable in each pair. The result here is slightly different from the result in Experiment I, such that the interaction effect was significant in Experiment I when comparing *not less* with *less…not*, but not in Experiment II. The possible explanation could be that the participants in Experiment II had stronger incentive to speed up than those in Experiment I, resulting in a more consistent learning effect. Overall, most importantly, it was manifested in both Hebrew and English data, that there were no learning rate differences between *more* and *less* conditions in each Sentence Type.

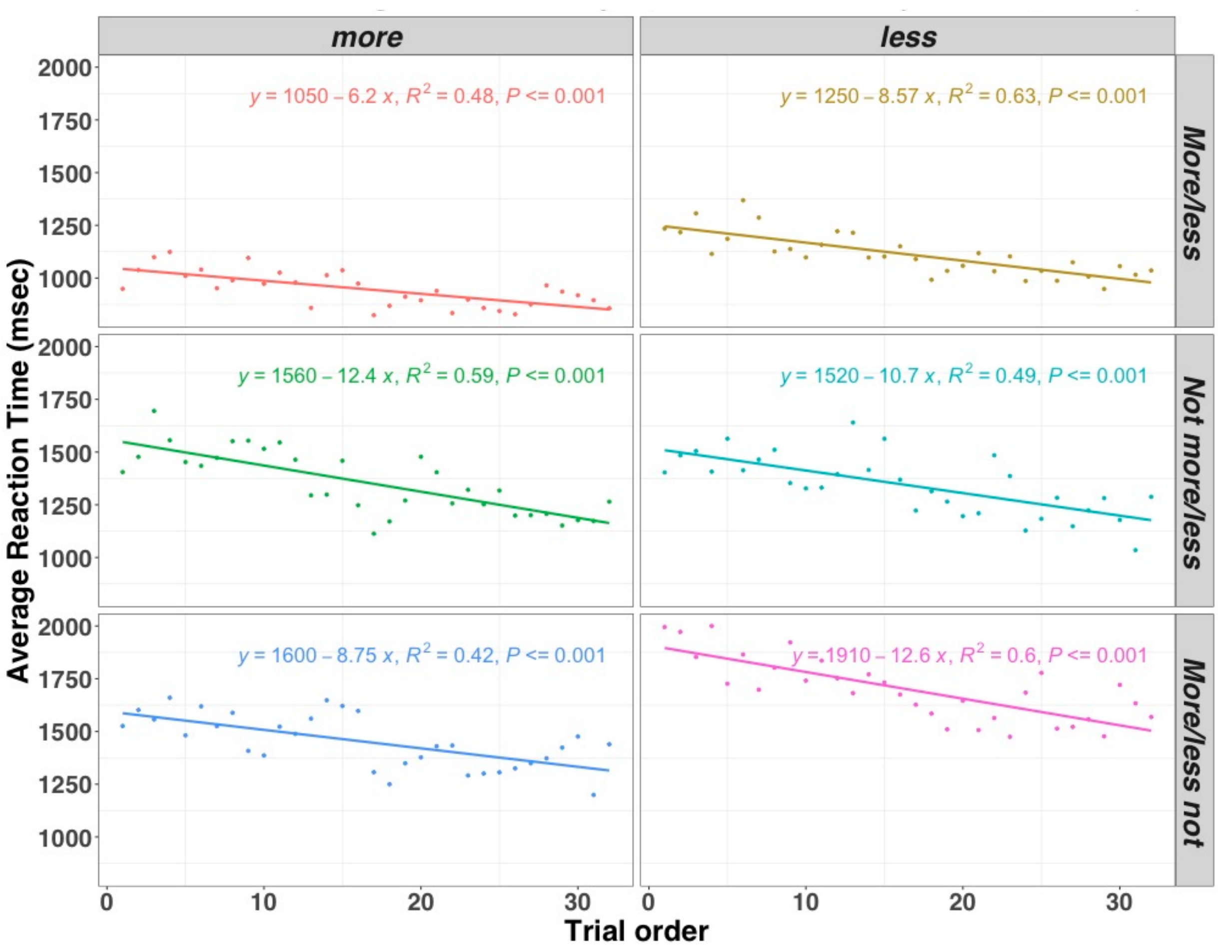


Figure 7 The time course of mean RTs in each condition. The x-axis indicates the order of a trial among all the trials of the same condition. ($1 \leq x \leq 32$, $x \in N$). Only correct responses were taken into account for the mean (95.21% of data). Regression lines were added to show the trend. With the slope of the regression lines, we can compare the learning rates of different conditions.

5. General Discussion

We started by exploring two approaches to NPI licensing, as contrasted by Homer (2011, 2021) in his study of flip-flop phenomena in French. He argued for the EnvBA by showing that only EnvBA could provide a reasonable account of the sentences with two DE operators which do not license NPI. Parallel to Homer's example in French, we showed that there is also a flip-flop of NPI licensing in Hebrew. The Hebrew NPI *ey pa'am* 'ever' is licensed when the two DE operators — *lo* 'not' and *paxot* 'less' – are in different domains, but not when the two sit in the same domain. We proposed the *Domain-based Processing Hypothesis* (DPH), which states that, during processing, if domain-wise monotonicity reverses when a lower domain integrates into a higher domain, extra processing cost is induced. In addition to the judgment test, we further provide experimental evidence substantiating our hypothesis regarding how syntactic structure affects the processing of monotonicity. We demonstrated that, all else being equal, the RT of a sentence containing two DE operators depends on the syntactic relationship between the two. First, compared with their one DE-operator counterparts, respectively, we observed $RT_{\textit{not less}} \leq RT_{\textit{not more}}$ but $RT_{\textit{less...not}} > RT_{\textit{more...not}}$ both in Hebrew and English. Secondly, even though the two double DE-operator structures comprise exactly the same words, it took much longer to process the sentence "*less than half of the circles are not blue*" than "*not less than half of the circles are blue*". We suggested that participants were able to integrate two DE operators as one UE operator only when the two DE operators were situated in the same domain.

To conclude, our findings - that come from the world of speeded behavior in which the RT, the variable of interest, is continuous - show remarkable convergence with the judgment data that Homer (2021) presented. They therefore provide further evidence to the Environment-based Licensing Condition for NPIs he proposed: NPIs are sensitive to monotonicity of their syntactic environment on the basis of domain because, cognitively, the processing of monotonicity is domain-based. In other words, when two DE operators occur in the same domain, we perceive the domain as one UE domain, resulting in the anti-licensing of NPI in such cases.

Author’s correspondence address

The Edmond and Lily Safra Center for Brain Sciences, Edmond J. Safra Campus, The Hebrew University, Jerusalem, 9190401, Israel

E-mail: i-an.tan@mail.huji.ac.il

Acknowledgements

This research has received funding from an internal grant from Edmond & Lily Safra Center of Brain Sciences, and funding from the European Union’s Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 945539 (Human Brain Project SGA3). This research was supported by the Joint Lab “Supercomputing and Modeling for the Human Brain”.

Data availability statement

The data that support the findings of this study are openly available in Open Science Framework at https://osf.io/mj7s8/?view_only=5ee0c5cc6ac64efdb8a47a90bd4ae28d.

Appendix 1: French “flip flop” and our Hebrew experimental materials

This appendix shows the direct connection between the sentence materials of our experiment and the environmental characterization of NPI licensing domains as presented by Homer (2020). Our experiment was conducted in Hebrew main clauses (containing zero, one or two DE operators in subject position). A demonstration that Hebrew is like English and French is thus in order.

We begin with the “flip-flop asymmetry” Homer discusses, inspired by Chierchia (2004), Gajewski (2005) and Guerzoni (2006): a weak NPI is licensed by the DE operator in sentence (A1), but not in (A2). Homer account for this asymmetry by assuming that NPI licensing is syntactically constrained. He posits a Polarity Phrase (PolP) as a domain of NPIs, which DE operators are said to be part of. A weak NPI requires a DE licensor (A1). An NPI is licensed if contained within a PolP with an odd number of DE operators (A2); otherwise, it is anti-licensed (A3).[12]

(A1) *…[TP ⇧ T [PolP ⇧ possible [CP que Jean [TP ⇧T [PolP ⇧ [*quoi que ce soit*]$_1$ faire t_1]]]]]

(A2) …[TP ⬇ T [PolP ⬇ **im**possible [CP que Jean [TP ⇧T [PolP ⇧ [*quoi que ce soit*]$_1$ faire t_1]]]]]

(A3) *…[TP ⇧ T [PolP ⇧ **pas** **im**possible [CP que Jean [TP ⇧T [PolP ⇧ [*quoi que ce soit*]$_1$ faire t_1]]]]]

(A4) [TP ⇧ T [PolP ⇧ **im**possible [CP que Jean [TP ⇧T [PolP ⬇ **pas** [*quoi que ce soit*]$_1$ faire t_1]]]]]

Example (A5), where DE-ness is induced in the antecedent clause of a conditional sentence, confirms Homer's hypothesis. In (A5) a PolP is found, that contains the NPI and one DE operator ***im****possible*. Thus, the second DE operator ***si*** in the antecedent of the conditional is above the roof of the licensing domain, unlike in (A3), where both are under the roof:

(A5) [TP ⇧ [CP ⇧ **Si** T [**PolP ⬇** **im**possible [CP que Jean [T [PolP ⇧ [*quoi que ce soit*]$_1$ faire t_1]]]]] T]…

In addition to PolP, Homer posits TP as an another domain of NPIs, based on (A6), which contains two DE operators, *si* "if" and *au plus cinq personnes* "at most five people". Assuming the subject *au plus cinq personnes* sits at [Spec, TP], for *quoi que ce soit* to be licensed, there must be a domain between CP and PolP, namely, TP, as Homer proposes.

(A6) [TP ⇧ [CP ⇧ **Si** [TP⬇ [au plus cinq personnes] T [PolP ⇧ ont [*quoi que ce soit*]$_1$ faire t_1]]] T]…

Hebrew offers a perfect replication of the examples that get Homer's theory off the ground (with the weak NPI *'ey-pa'am* = ever):

(A1-heb) *…[TP ⇧ T [PolP ⇧ efšari [CP še Dani [TP ⇧ T [PolP ⇧ nirdam *'ey-pa'am* be-šmira]]]]]
possible that Dani fell.asleep ever while.on.guard
'… possible that Dani ever fell asleep while on guard'

(A2-heb) …[TP ⇧ T [PolP ⬇**bilti**-efšari [CP še Dani [TP ⇧ T [PolP ⇧ nirdam *'ey-pa'am* be-šmira]]]]]

[12] Some examples in Appendix 1 have appeared earlier in the main text. They are iterated here for the readers' convenience.

impossible that Dani fell.asleep ever while.on.guard

'… impossible that Dani ever fell asleep while on guard'

(A3-heb) *…[TP ⇧ T [PolP ⇧ **lo** **bilti**-efšari [CP še Dani [TP ⇧ T [PolP ⇧ nirdam *'ey-pa'am*

not impossible that Dani fell.asleep ever

be-šmira]]]]]

while.on.guard

'… not impossible that Dani ever fell asleep while on guard'

(A4-heb) …[TP ⇧ T [PolP⬇ **bilti**-efšari [CP še Dani [TP ⇧T [PolP ⬇ **lo** nirdam *'ey-pa'am*

impossible that Dani not fell.asleep ever

be-šmira]]]]]

while.on.guard

'… impossible that Dani did not ever fall asleep while on guard'

(A5-heb) [TP ⇧ [CP ⇧ **'im** [PolP ⬇ **bilti**-efšari [CP še Dani [TP ⇧T [PolP ⇧ nirdam *'ey-pa'am*

if impossible that Dani fell.asleep ever

be-šmira]]]]] T]…

while.on.guard

'If it is impossible that Dani ever fell asleep while on guard, …'

(A6-heb) [TP ⇧ [CP ⇧ **'im** [TP⬇ [**le-xol-ha-yoter** xamiša 'anašim] T [PolP ⇧ siy'u *'ey-pa'am* la

if at.most five people assisted ever the

mafia]]] T]…

mafia

'If at most five people ever assisted the mafia, …'

Appendix 2: extension of "flip flop" to simple, monoclausal sentences

Yet embedded clauses and conditionals turn out to not be the only flip-flop environments. Homer presents (A7), a monoclausal sentence containing two DE operators, *au plus cinq* and *n'…pas*. Again, assuming the subject *au plus cinq* is at [Spec, TP], there must be a DE domain below TP, which licenses the NPI, namely, PolP.

(A7) *Au plus cinq personnes n' ont pas fait quoi que ce soit pour aider la Mafia.*

at most five people ne have neg done what that this be.subj to help the Mafia

'At most five people didn't do anything to help the Mafia.'

Unfortunately, Homer did not provide a minimal contrast where *at most* combines with a constituent negation (i.e. *not at most*). This is important to our experiment, as our materials were simple double-negative sentences with either constituent negation or sentential negation. Below we show that Homer's hypothesis is valid even among the simple sentences, as we found a "flip-flop asymmetry" between different syntactic structures.

We used mono-clausal sentences whose subjects are generalized quantifiers with modified numerals (*yoter/paxot me-xamiša* = more/less than five), and their negated counterparts (*lo yoter/paxot me-xamiša* = not more/less than five). A weak NPI (*'ey pa'am* = ever) in object position is not licensed by a UE-quantifier (A8); yet it is licensed by the DE-quantifier *paxot* (A9). As in (A7), when a sentential negation *lo* is added to a sentence with a DE operator in the subject position, an NPI is still licensed (A10). The flip-flop happens when a constituent negation is added to (A9), as no DE domains are available in this case (A10).

(A8) *[TP ⇧ [Yoter me- xamiša ratzim][T [PolP ⇧ [higi'u *ey pa'am* la-gmar]]]].
more than-five runners reached ever to-the-finish-line
'More than five runners have ever passed the finish line.'

(A9) [TP ↓ [**paxot** me-xamiša ratzim][T [PolP ⇧ [higi'u *ey pa'am* la-gmar]]]].
less than-five runners reached ever to-the-finish-line
'Less than five runners ever reached the-finish-line.'

(A10) [TP ⇧ [**paxot** me-xamiša ratzim] [T [PolP ↓ **lo** [higi'u *ey pa'am* la-gmar]]]].
less than-five runners not reached ever to-the-finish-line

(A11)*[TP ⇧ [**lo** **paxot** me-xamiša ratzim] [T [PolP ⇧ [higi'u *ey pa'am* la-gmar]]]].
not less than-five runners reached ever to-the-finish-line

These examples indicate that Homer's effects can be extended to monoclausal sentences. The two DE operators in (A11) integrate and together provide a UE environment to *ey pa'am*, whereas in (A10), the *not* itself may provide a DE domain to *ey pa'am*. The key point here is how the integration of two DE operators occurs, since eventually, in the maximal domain, i.e. the sentence, *ey pa'am* is in a UE domain in both (A10) and (A11). The contrast between (A10) and (A11) suggest that the integration happens stage by stage (*cf.* Papeo et al., 2016).

These Hebrew data are align with Homer's hypothesis that PolP and TP are both valid domains for NPIs, showing that our experimental materials are well-suited to distinguish between the hypotheses.

References

Agmon, Galit, Yonatan Loewenstein, & Yosef Grodzinsky, (2019). Measuring the cognitive cost of downward monotonicity by controlling for negative polarity. *Glossa: a Journal of General Linguistics*, 4(1).

Baker, Carl. L. (1970). A note on scope of quantifiers and negation. *Linguistic inquiry*, 1(1), 136-138.

Bates, Douglas, Martin Maechler, Ben Bolker, & Steve Walker. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software 67*(1). 1–48.

Bott, Oliver, Fabian Schlotterbeck, & Udo Klein, (2019). Empty-set effects in quantifier interpretation. *Journal of Semantics*, 36(1), 99-163.

Chierchia, Gennaro. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. *Structures and beyond*, 3, 39-103.

Clark, Herbert. H., & William. G. Chase, (1972). On the process of comparing sentences against pictures. *Cognitive psychology*, 3(3), 472-517.

Crnič, Luka. (2014). Non-monotonicity in NPI licensing. *Natural Language Semantics*, 22, 169–217.

Deschamps, Isabelle, Galit Agmon, Yonatan Loewenstein, & Yosef Grodzinsky, (2015). The processing of polar quantifiers, and numerosity perception. *Cognition*, 143, 115–128.

Fauconnier, Gilles. (1975). Polarity and the scale principle. *Proceedings of Chicago*, 11,188-199.

Gajewski, Jon. R. (2005). *Neg-raising: Presupposition and polarity*. PhD Thesis, MIT. Cambridge, MA.

Gajewski, Jon. R. (2011). Licensing strong NPIs. *Natural Language Semantics*, 19, 109-148.

Guerzoni, Elena. (2006). Intervention effects on NPIs and feature movement: Towards a unified account of intervention. *Natural Language and Semantics*, 14, 359–98.

Homer, Vincent. (2010). *Polarity and modality.* PhD Thesis, University of California, Los Angeles.

Homer, Vincent. (2021). Domains of polarity items. *Journal of Semantics*, 38(1), 1-48.

Just, Marcel. A., & Patricia. A. Carpenter, (1971). Comprehension of negation with quantification. *Journal of verbal learning and verbal behavior*, 10(3), 244-253.

Klima, Edward. (1964). Negation in English. In Jerry A. Fodor and Jerrold J. Katz (eds.), *The structure of language*. New Jersey: Prentice Hall. 246-323.

Kuznetsova, Alexandra, Per Bruun Brockhoff & Rune Haubo Bojesen Christensen. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13). 1–26.

Ladusaw, William A. (1980). Affective Or, Factive Verbs, and Negative-Polarity Items. In J. Kreiman and A. E. Ojeda (eds.), *Papers from the Sixteenth Regional Meeting of the Chicago Linguistics Society*. Chicago. III(16), 170-184.

Papeo, Liuba, Jean-Rémy Hochmann & Lorella Battelli. (2016). The default computation of negated meanings. *Journal of Cognitive Neuroscience*, 28(12), 1980–1986.

Schlotterbeck, Fabian. (2017). *From truth conditions to processes: how to model the processing difficulty of quantified sentences based on semantic theory*. PhD Thesis, University of Tübingen. Germany.

Schlotterbeck, Fabian, Sonia Ramotowska, Leendert van Maanen, & Jakub Szymanik. (2020). Representational complexity and pragmatics cause the monotonicity effect. In *CogSci*.

Sherman, Mark. A. (1973). Bound to be easier? The negative prefix and sentence comprehension. *Journal of Verbal Learning and Verbal Behavior, 12*(1), 76–84.

Sherman, Mark. A. (1976). Adjectival negation and the comprehension of multiply negated sentences. *Journal of Verbal Learning and Verbal Behavior, 15*(2), 143–157.

Tan, I-An, Nitsan Kugler-Etinger & Yosef Grodzinsky (2023). Do two negatives make a positive? Language and logic in language processing. *Language, Cognition and Neuroscience*, 1-16.

Von Fintel, Kai. (1999). NPI licensing, Strawson entailment, and context dependency. *Journal of semantics*, 16(2), 97-148.

Xiang, Ming, Brian Dillon & Collin Phillips. (2009) Illusory licensing effects across dependency types: ERP evidence. *Brain and Language*, 108, 1, 40-55.

Xiang, Ming, Julian Grove & Anastasia Giannakidou. (2013). Dependency dependent interference: NPI interference, agreement attraction, and global pragmatic inferences. *Frontiers Psychology*, 4:708. doi:10.3389/fpsyg.2013.00708

Xiang, Ming, Julian Grove & Anastasia Giannakidou. (2016). Semantic and pragmatic processes in the comprehension of negation: an event related potential study of negative polarity sensitivity. *Journal of Neurolinguistics*, 38, 71-88.