The Processing Cost of Downward Entailingness: The Representation and Verification of Comparative Constructions[1]

Yosef GRODZINSKY – *Edmond and Lily Safra Center for Brain Sciences (ELSC) HUJI, INM-1, Forschungszentrum Jülich*
Galit AGMON – *Edmond and Lily Safra Center for Brain Sciences (ELSC) HUJI*
Kedem SNIR – *Edmond and Lily Safra Center for Brain Sciences (ELSC) HUJI*
Isabelle DESCHAMPS – *Faculté de Médecine, Université Laval*
Yonatan LOEWENSTEIN – *Edmond and Lily Safra Center for Brain Sciences (ELSC) HUJI, Department of Neurobiology, The Federmann Center for the Study of Rationality, HUJI*

Abstract
We bring experimental considerations to bear on the structure of comparatives and on our understanding of how quantifiers are processed. At issue are mismatches between the standard view of quantifier processing cost and results from speeded verification experiments with comparative quantifiers. We build our case in several steps: 1. We show that the standard view, which attributes processing cost to the verification process, accounts for some aspects of the data, but fails to cover the main effect of monotonicity on measured behavior. We derive a prediction of this view for comparatives, and show that it is not borne out. 2. We consider potential reasons – experimental and theoretical – for this theory-data mismatch. 3. We describe a new processing experiment with comparative quantifiers, designed to address the experimental concerns. Its results still point to the inadequacy of the standard view. 4. We review the semantics of comparative constructions and their potential processing implications. 5. We revise the definition of quantifier processing cost and tie it to the number of Downward Entailing (DE) operators at Logical Form (LF). We show how this definition successfully reconciles the theory-data mismatch. 6. The emerging picture calls for a distinction between the complexity of verified representations and the complexity of the verification process itself.

Keywords: quantification, monotonicity, negation, comparative constructions, Logical Form, adjectival antonyms, decomposition, quantifier processing, speeded verification, reaction time.

1. The landscape

1.1. Monotonicity

That monotonicity is a property of many natural language quantifiers has been recognized since Barwise and Cooper (1981). Monotonicity is defined by entailment patterns:

(1)   a. A quantifier Q is Upward Entailing (UE), if $A \subseteq A' \Rightarrow Q(A) \subseteq Q(A')$
         b. A quantifier Q is Downward Entailing (DE), if $A \subseteq A' \Rightarrow Q(A') \subseteq Q(A)$
         c. A quantifier Q that is neither UE nor DE is non-monotone

---

Quantifiers in richer structures such as Q(A)(B) may be UE or DE on their restrictor (A) and nuclear scope (B) arguments. We illustrate this with *more than two* and *fewer than three*. As {tall men}⊆{men} and {have a red beard}⊆{ have a beard}, the following entailments hold:

(2)   a. left UE: [$_Q$ *More than two*] [$_A$ tall men] [$_B$ have a beard]
           ⇒ [$_Q$ *More than two*][$_A$ men] [$_B$ have a beard]
       b. right UE: [$_Q$ *More than two*] [$_A$ tall men] [$_B$ have a red beard]
           ⇒ [$_Q$ *More than two*][$_A$ tall men] [$_B$ have a beard]
       c. left DE: [$_Q$ *Fewer than three*] [$_A$ tall men] [$_B$ have a beard]
           ⇐ [$_Q$ *Fewer than three*][$_A$ men] [$_B$ have a beard]
       d. right DE: [$_Q$ *Fewer than three*] [$_A$ tall men] [$_B$ have a red beard]
           ⇐ [$_Q$ *Fewer than three*][$_A$ men] [$_B$ have a beard]

1.2. Monotonicity and processing: verification and the DE Cost Effect

How are sentences with monotone quantifiers processed and verified? Barwise and Cooper famously proposed a witness set (WS) verification algorithm that is based on iterated sampling. It distinguishes UE from DE quantifiers in terms of verification complexity (cf. Szymanik, 2016 for a comprehensive recent review). The clearest case may come from proportional quantifiers: with UE ones, (e.g., *more-than-half*), a single positive Example found in a scenario suffices for verification by the WS algorithm. Yet with their DE counterparts (*less-than-half*), exhaustive scrutiny of the whole scenario is required. Thus here, the WS algorithm requires more steps to verify a DE sentence than its UE counterpart. This verification method is predicted to bring about behavioral UE/DE differences. As Barwise and Cooper comment: "*we predict that response latencies for verification tasks involving decreasing quantifiers would be somewhat greater than for increasing quantifiers…These predictions are based on the complexity of the checking procedure we have suggested*" (p. 192). Though not explicitly discussed by Barwise and Cooper, falsification with this checking method reverses the complexity of the WS algorithm and, correspondingly, the prediction (at least with proportional and degree quantifiers): the WS strategy above predicts that falsifying DE sentences should be faster than falsifying their UE counterparts (and similar, or even equal, to verifying their UE counterparts). We can use verification time, quantified using reaction time (RT) measurements in a verification experiment, to test this theory. Imagine a verification experiment with a 2×2 design, in which the Polarity contrast ($RT_{DE} - RT_{UE}$) is pitted against Truth-value ($RT_F - RT_T$). Under the WS algorithm, no main effects (of a Polarity factor across Truth-value, or a Truth-value factor across Polarity) are expected. Yet we expect a Polarity×Truth-value interaction, due to the DE/UE ordering reversal in the case of falsification: longer RT in True DE and False UE verifications (Figure 1a).
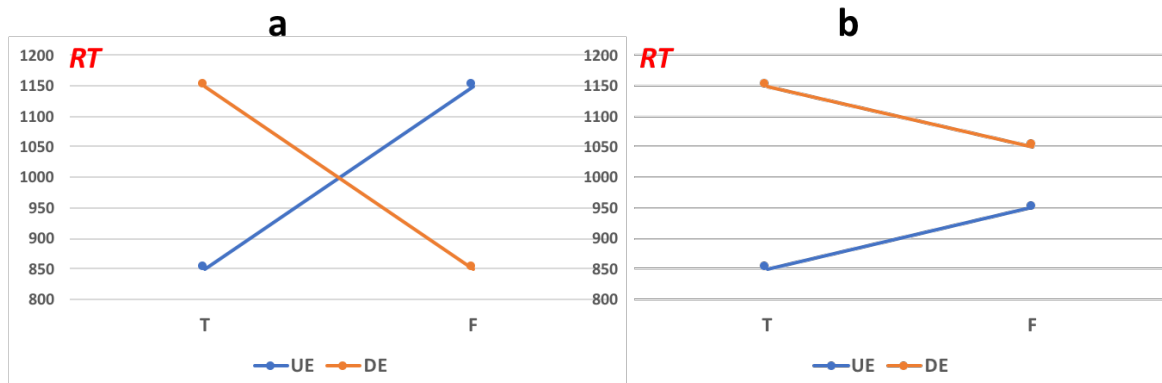
Figure 1: Response patterns with polar degree quantifiers. **a**. Predictions of the WS strategy: no main effect of Polarity or Truth-value; disordinal (non-order preserving) Polarity×Truth-value interaction. **b**. Schematic illustration of Just and Carpenter's (1971) results: main effect of Polarity; all other effects – same as in **a**.

Actual results of such an experiment were in fact already available at the time of Barwise and Cooper's writing: Just and Carpenter (1971) presented data from speeded verification, in which quantifiers Polarity was pitted against Truth-value. Sentences with UE and DE degree quantifiers were verified against scenarios that contained 2 black and 14 red dots (or 14 black and 2 red dots).[2] Half of the scenarios made each sentence true, and the other half made each sentence false:

(3)   a. UE: Many of the dots are black
        b. DE: Few of the dots are red

Just and Carpenter found a Polarity×Truth-value interaction effect, as the WS algorithm would later predict. Yet they also obtained a main effect for Polarity, $RT_{DE} >^{sig} RT_{UE}$ (illustrated in Figure 1b), not predicted by the WS algorithm. We call this main effect the DE Cost (DEC) Effect. It is found across Truth-values, hence it is independent of verification.[3]

The shape of the interaction effect (disordinal: the slopes of the imaginary lines connecting the RTs for each quantifier are opposite to one another) is consistent with the WS, yet the unexpected main effect of Polarity suggests that the processing complexity of quantified sentences has two independent components. Indeed, Just and Carpenter proposed an account that reflects this independence. It attributes the main (DEC) Effect to costly lexical decomposition [*few*=NOT(*many*)], and the Polarity×Truth-value interaction to attentional shifts during verification.[4]

---

[2] *Many* and *few* are arguably ambiguous between adjectival and quantificational denotations. As we show below, the same results are obtained from unambiguously proportional determiners (*more/less-than-half*).
[3] A comment by Dan Goodhue played an important role in clarifying this point.
[4] They proposed an attentional strategy, imposed by the scenario, which forces a participant to attend to the larger set of dots first; an attentional shift to the smaller set (driven by sentence content) is costly. In the experimental context, the UE stimulus sentence in (3a) is true iff followed by an image in which the larger set of 14 dots is in the predicate color. Focusing on this set for Truth-value determination is appropriate, making for a speedy response. However, this UE sentence is false when the *smaller* set of 2 dots is in the predicate color; attention must therefore be shifted to the larger set for the determination of Truth-value, and as the shift is supposedly costly, a longer response time is expected. The same logic applies to the DE case (3b), but in reverse.

1.3. Prospectus

This note focuses on one of the two components, the DEC Effect – the processing signature of the monotonicity of degree quantifiers *across Truth-value*. DEC can be expressed thus:

(4)   DEC Effect (preliminary)
        $\Delta RT = RT_{DE} - RT_{UE} > 0$

RT_{DE} and RT_{UE} are speeded verification times of the sentences containing UE and DE quantifiers, respectively. The DEC Effect is robust and has been repeatedly reported (Geurts and van der Slik, 2005; Deschamps, Agmon, Loewenstein and Grodzinsky, 2015). It is at the heart of our exploration of the representational and verification complexity of comparative quantifiers via empirical tests of their perceptual complexity. We try to identify the source of the DEC Effect and see how such an understanding bears on the semantics of comparatives.

For the remainder of Section 1, we review some recently published experimental tests, which seem to support a decompositional analysis of DE quantifiers along the lines of Just and Carpenter (cast in current lingo). We proceed to a puzzle that arises with respect to a DEC Effect found for comparative constructions and consider possible solutions (Section 2). We rule out one of these through an experiment we report (Section 3) and then delve into the details of the Downward Entailingness in the context of comparative quantifiers (Section 4): we motivate a decompositional analysis of *more-* and *less-*comparatives (Büring, 2007a,b; Heim, 2006; Rullman, 1995) and show that each of these posits a different number of DE operators in these comparatives. This difference may help in revising the DEC Effect to fit our experimental data. In Section 5, we redefine the DEC Effect accordingly and show how this definition not only helps to account for the problematic data from comparatives, thereby lending support to the decompositional analysis, but also serves as a tool for the identification of hidden DE-operators.

In Section 6, we reflect on the view that the processing complexity of quantifiers is determined by two components. First, the Polarity main effect is captured by DEC. We call it the representational component, as it is determined by the structural complexity of the quantifiers at issue. Second, the Polarity×Truth-value interaction is in keeping with the WS algorithm. We therefore call it the verificational component. All in all, we show how results that come from the continuous time domain can be explicitly mapped onto linguistic representations and brought to bear on linguistic theory, even though this theory only features categorical variables.

1.4. Recent experimentation on quantifier monotonicity

Deschamps et al. (2015) report the results of three speeded verification experiments with polar quantifiers, in which matched auditory sentences were coupled with images that contain blue and yellow circles in varying proportions:

(5)   a. More than half of the circles are blue
        b. Less than half of the circles are blue

(6)   a. Many of the circles are blue
        b. Few of the circles are blue

(7)   a. There are more blue circles than yellow circles
        b. There are fewer blue circles than yellow circles

Each trial began with a visual fixation point followed by an auditory sentence, which was then followed by an image which participants were asked to verify (Figure 2). In addition to measuring a DEC Effect, these experiments tried to see whether DE Cost is affected by properties of the truth-making scenario (in this case, by Weber's Law). Therefore, the blue/yellow proportion in the scenarios was varied along a seven-valued parameter. This proportion determined both Truth-value (T/F) and task difficulty. As the proportion approached 1, the task was more difficult. In Figure 2, for Example, we see a more difficult true case and an easier false one:
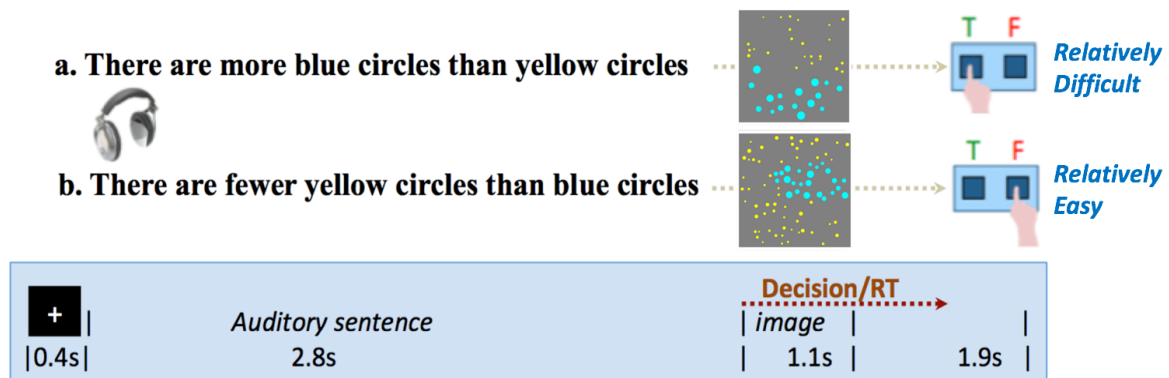

Figure 2: form, content and time-course of stimuli.

Three tests using the sentences shown in (5)-(7) and a host of control conditions were carried out in the same verification paradigm. RT functions, time-locked to image onset as seen in Figure 2, behaved in keeping with the inequality in (4) modulo Weber's Law, across all seven values of the proportion parameter and across True and False instances. Figure 3 collapses participants' scores across Truth-value and proportion and presents the DEC Effect for the polar pairs in (5)-(7) by showing the difference between means (***=p<.001)[5]. The effect was robust, manifesting in almost all individual participants.

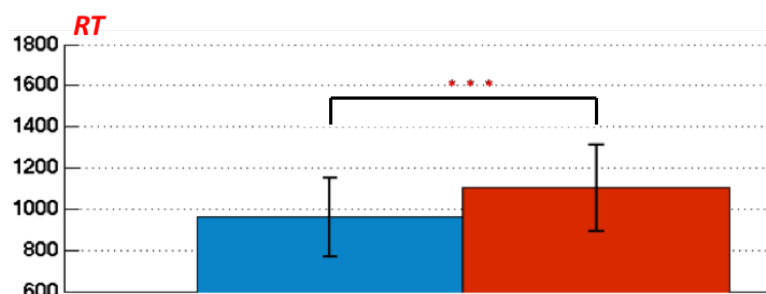**Figure 3**: RTs in msec for UE (blue) vs. DE (red) stimuli. Error bars mark 1 population SD, Deschamps et al. (2015).

---

[5] We collapse across proportions because no DEC×proportion interaction was found, see Deschamps et al. (2015) for details.

Deschamps et al. drew several conclusions from this result. Of these, one is of interest here: a DEC Effect is found in a variety of instances but not in control conditions[6].

2. Problems and possible solutions

2.1. The RT puzzle in comparatives

Among the cases tested, was the pair of comparative sentences (7).[7] It contains polar quantifiers (*more*, *less*), and as such, it seems at first blush that a DEC Effect is also expected. But processing complexity is determined by properties of generalized quantifiers,[8] which is reflected by entailment patterns. These are mixed in comparatives – each sentence seems to have a UE and a DE component, except in reversed order:[9]

(8)   a. <u>More **cats**</u> than snakes died $\Rightarrow$
         <u>More **mammals**</u> than snakes died   ({**cats**}$\subset${**mammals**})
       b. More cats <u>than **reptiles**</u> died $\Rightarrow$
         More cats <u>than **snakes**</u> died   ({**snakes**}$\subset${**reptiles**})

(9)   a. <u>Fewer **mammals**</u> than snakes live in deserts $\Rightarrow$ <u>Fewer **cats**</u> than snakes live in deserts
       b. Fewer cats <u>than **snakes**</u> live in big cities $\Rightarrow$ Fewer cats <u>than **reptiles**</u> live in big cities
A natural construal of the DEC Effect here is to fix it to the cumulative monotonicity of a sentence. Decomposing each comparative left-to-right, we obtain that (7a-b) do not differ in total DE-ness, as that they both contain a DE environment and a UE environment. We can therefore compute the predicted relation between their verification times under DEC:

(10)   a. $RT_{(7a)} = RT_{UE} + RT_{DE}$
        b. $RT_{(7b)} = RT_{DE} + RT_{UE}$
        hence
        c. $RT_{(7a)} = RT_{(7b)}$, or $\Delta RT = RT_{DE} - RT_{UE} = 0$ (at the very least: $\neg(RT_{(7b)} > RT_{(7a)})$)

Deschamps et al., however, found that in comparatives, $RT_{(7b)} >^{sig} RT_{(7a)}$, or $\Delta RT > 0$. By (10), this effect is not expected. We now address possible reasons for this puzzle.

2.2. Possible solutions

We are aware of three logically possible explanations for this theory/data mismatch:

---

[6] Deschamps et al. Compared quantifier polarity to the direction of algebraic inequalities. When the expression to be verified was not a sentence, but rather, a quasi-algebraic expression with "<" or ">", not polarity effect analogue was found.

[7] Note that the WS verification algorithm cannot work for comparatives, as the quantifier has no restrictor which can be sampled to determine Truth-value. See Section 6.2 for further elaboration.

[8] Deschamps et al. also consider a frequency-based account of the DEC Effect, by which it is due to differences in the lexical frequency between DE and UE quantifiers. They reject this account on several grounds, one of which coming from frequency differences between UE quantifiers (i.e., $f_{more} > f_{many}$) that do not manifest in the RT domain (i.e., in sentence verification, $RT_{more} < RT_{many}$).

[9] These entailment patterns are among the reasons for the characterization of these as "A-not-A" (Schwarzschild, 2008).

I.   The experimental results are compatible with alternative interpretations.
II. Assumptions regarding comparative structure are incorrect.
III. The definition of DEC is incorrect.

What follows is a consideration of these possibilities, and an amended account of the data.

3. A possible experimental wrinkle and its fix

3.1. The issue

Consider the experimental paradigm illustrated in Figure 2, where all image stimuli are composed of objects in two colors (blue and yellow). Prior to testing, participants are informed that these two colors would be the only ones to feature in the images. Sentences (5)-(6) contain a single color term, realized in the right argument of the quantifier – the last word in the sentence. Correct Truth-value judgment in the binary-choice task requires a complete parse of the sentence.

Comparatives are different. Consider the phrasal comparatives used in Deschamps et al.'s study, repeated here in (11): they contain two color terms, that is, both *blue* and *yellow*.

(11)   a. There are more blue circles than yellow circles
       b. There are fewer blue circles than yellow circles

In the context of the task, a participant has enough information to perform correctly with only part of the sentence, as parsing the first part of the comparative is sufficient for verification. If the first argument is *blue*, she can safely conclude that the second must be *yellow* and vice versa. Attending to the rest of the sentence would convey no further critical information.

But what would make her stop listening amid sentence? Recall that the task involves *speeded* verification, i.e., participants are motivated to decide and respond as fast as they can after the image appears. In a repetitive experimental session such as ours (>200 trials per run), compliant participants act efficiently: they try to perform as instructed, but at the same time seek to reduce effort. After a few trials, they can quickly learn that a partial parse is sufficient. This may result in the deployment of a time-saving strategy with no accompanying loss of accuracy.

If this strategic response method is possible, we should compute the DEC Effect only on the first part of the comparative. This would make the comparative quantifier mono-argumental, akin to (5a-b). This would make (7a), *more*-comparative, a UE sentence; and (7b), a *less*-comparative, a DE one. With this strategy, the observed result, $\Delta RT > 0$, is expected.

Deschamps et al.'s experiment cannot rule out this interpretation. We must therefore modify the experimental paradigm so as to rule out this strategic performance option and ensure that the experiment tests what it aims to test. This is what we did in a new experiment, run with a group of Hebrew University undergraduate students (n=22), all native Hebrew speakers, who participated for either payment or credit giving their informed consent. The experiment was approved by the Hebrew University Research Ethics Committee.

3.2. Materials and methods

The experiment used Hebrew versions of the sentences in (11), but did so in a context that forced participants to produce a complete parse of the comparative[10]. The experimental design used the above stimuli but added an image type. That is, like before, each image contains a proportion of circles in two colors, but these were picked out of three colors (red, blue, and yellow), producing combinations as in (12), Figure 4:
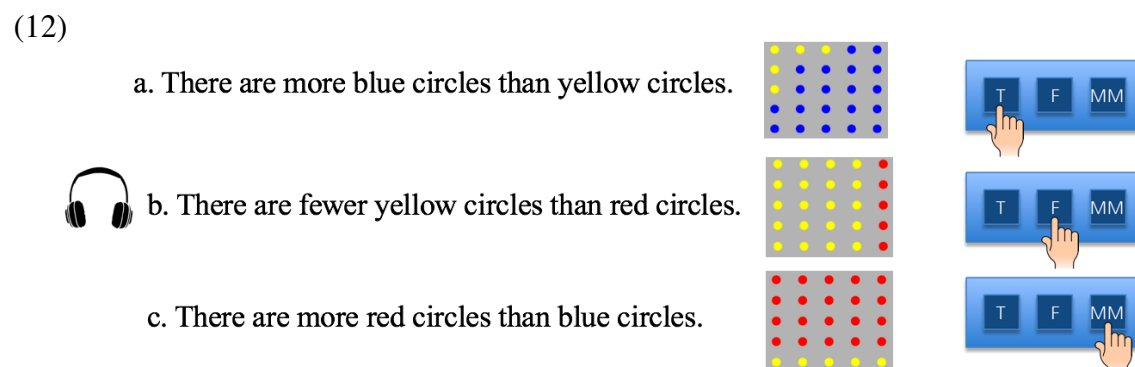
(12)

    a. There are more blue circles than yellow circles.

    b. There are fewer yellow circles than red circles.

    c. There are more red circles than blue circles.

Figure 4: design and conditions of the present comparatives experiment.

Some images contained a color that was not mentioned in the sentence, which made the sentence infelicitous. Participants were instructed about these possibilities and were asked to mark these as MisMatches (MM). They were given a third button in addition to the true and false ones (MM, 12c).[11] In this context, the correct response was now discoverable only at the end of the sentence. Participants were forced to pay attention throughout the sentence and parse the complete comparative with both its UE and DE parts. No DEC Effect was therefore predicted.

This 2×3 design, with Polarity and Image Type as factors and 20 stimuli per type, led to a test that had 120 trials, preceded by a short training session. Sentence-image stimuli were presented in a random order using a Presentation© code.

3.3. Results

The results draw a clear picture (Table 1, Figure 5): a. mean error rates were low across all conditions (in parentheses, Standard Deviations). b. RTs present a main Polarity effect (the difference between the means for *more* and *less*, annotated blue and red in Table 1, across Truth-value); that is, even in the present test paradigm, where participants are forced to parse the sentence in its entirety, there is a significant difference between the processing of *more*- and *less*-comparatives across Truth-value ($F(1,21)=97.236$, $p<.0000001$). We return to other aspects of these data later.

---

[10] The Hebrew sentences: *Yeš yoter/paxot ʻigulim kxulim me-ʻadumim*, etc.

[11] Although it would also be possible to create a MisMatch using the first color term, we only tested instances in which the mismatch was realized at the end of the sentence, namely those in which the offending color was in the comparative.

(13)		Mean proportion correct (SD)			Mean RT in msec (SD)		
Sentence	Truth-value	T	F	MM	T	F	MM
a. There are *more* blue circles than yellow circles		0.954 (0.04)	0.885 (.08)	0.940 (0.05)	**897** (181)	**1047** (175)	798 (100)
b. There are *fewer* blue circles than yellow circles		0.898 (0.09)	0.856 (0.07)	0.950 (0.04)	**1022** (208)	**1115** (195)	831 (112)

Table 1: colored RTs are in keeping with the bar-graph below.
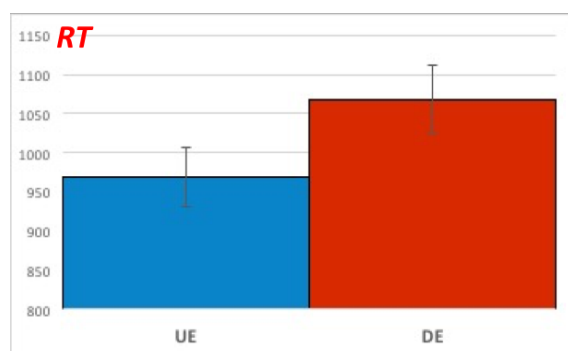

Figure 5: the Polarity main effect with UE (blue) vs. DE (red) sentences (RT in msec). Error bars mark SEM.

3.4. Interim conclusion

We can assert with confidence that the RT puzzle is not due to an experimental artifact, as it persists even when we can guarantee complete processing of the comparative. Next, we reexamine the assumptions that led to the puzzle in the first place. For the RT mismatch to be solved, one needs to look deeper into the structure of comparatives in an attempt to uncover what hidden complexity elements they may contain.

4 Comparatives and monotonicity

4.1 The need for a revised analysis

In this Section, we consider how comparatives should be analyzed. We present two well-known puzzles that the monotonicity-based analysis of comparatives faces. a. NPIs, licensed only in DE environments, are nonetheless licensed in the UE part of *less*-comparatives. b. an ambiguity in *less*-comparatives is mysteriously resolved by the insertion of an NPI. Neither puzzle arises in *more*-comparatives. We review a decompositional solution that derives the UE-ness of the comparative clause in *less*-comparatives from two DE operators. We then proceed to point out problems in this analysis, which have led some investigators to endorse an alternative, non-decompositional, lexical account, which we then sketch. Finally, as we later argue that our experimental results help to fortify the decompositional account, we demonstrate that phrasal comparatives – the cases we tested – are indeed relevant.

4.1.1 An apparent mismatch

There is a well-known mismatch between entailment patterns and the licensing patterns of Negative Polarity Items (NPI) in comparatives (exemplified here with phrasal ones): above, in (8)-(9), *more/fewer*-comparatives were shown to have "mixed", *A-not-A*, monotonicity. One would expect this pattern to correlate with that of NPI-licensing. That is, all DE environments, and only those, should license NPIs. Yet the observed correlation is only partial: as expected, *more* licenses an NPI just in the DE *than*-phrase (14), but *fewer* licenses a NPI not only in its left, DE argument (15a), but also in its right, UE argument (15b) (Seuren, 1973):

(14)   a. #This city has <u>more cats that *ever* meow</u> than snakes   *(NPI in a UE environment)*
       b.   This city has more cats <u>than snakes that *ever* bite</u>   *(NPI in a DE environment)*

(15)   a.   This city has <u>fewer cats that *ever* meow</u> than snakes   *(NPI in a DE environment)*
       b.   This city has fewer cats <u>than snakes that *ever* bite</u>   *(NPI in a UE environment)*

Where does this mere partial correlation between NPI-licensing and entailment patterns come from? The comparative part of (15b) is UE by its entailment pattern, but the NPI within it cannot be licensed unless in the scope of a DE operator. This mismatch not only creates a linguistic puzzle, but also a cognitive one. An explanation proposed to the merely partial entailment/NPI licensing correlation might come to the rescue of the DEC Effect puzzle. Let's dig deeper.

4.1.2. Ambiguities in comparatives and their resolution with NPIs

To get to the bottom of the issue, we look into the comparative part of the *less*-comparative by returning to Rullman's (1995) discussion of a curious ambiguity (again following Seuren). Assuming that a jet plane can fly at heights between 1000-20000ft, we get this ambiguity:

(16)   The helicopter is flying less high than *a jet plane can*
       Context: jets can fly at 1000-20000ft
       Reading 1: The helicopter is flying below 1000ft
       Reading 2: The helicopter is flying below 20000ft

The parallel *more*-comparative evinces no such ambiguity:

(17)   The helicopter is flying higher than *a jet plane can*
       Reading: The helicopter is flying above 20000ft

Returning to *less*-comparatives, Rullman shows that only reading 1 must be analyzed as containing a DE environment. Assume that {x/x is a cargo plane}$\subseteq$ { x/x is a plane}. Then (18a) entails (18b) only on the "less-than-minimum" reading:

(18)   a. The helicopter is flying less high than a plane can $\Rightarrow$
       b. The helicopter is flying less high than a cargo plane can

That is, if the helicopter is flying at a (maximal) height that is below the minimum height a plane can fly (=1000ft), then in particular, it is flying below the minimum height at which cargo plane can fly. By contrast, this entailment does not go through on a "less-than-maximum" reading: if the maximum height that a plane can achieve is 20,000ft (below which the helicopter is flying), then it does not follow that the helicopter is flying below the maximal height a cargo plane can reach (as this plane's maximum height may be less than 20,000ft). It is entirely possible, then, that the helicopter in (18a) is flying higher than the cargo plane but lower than the maximum height a plane can reach (20,000ft), thereby falsifying (18b). The representation of the "less-than-maximum" reading thus does not contain a DE environment. [12] Thus generally, as Rullman (p. 87) puts it, on the "less-than-minimum" interpretation, *less than*-comparatives are DE.

Moreover, NPIs are only licensed on the "less-than-minimum" reading, because it is the only reading whose representation contains a DE environment. Indeed, (19) is not ambiguous:

(19)   The helicopter is flying less high than *any* plane can

How can these complex facts be explained? Below, we show how lexical decomposition (Rullman, 1995; Heim, 2006; Büring, 2007a,b) accounts for these by breaking DE quantifiers into pieces. We will later argue that this analysis, coupled with straightforward assumptions about processing, accounts for the RT puzzle with which we began.

4.2. A sketch of the decompositional account

Below, we conflate the three accounts just mentioned into one, as they all share the relevant property which we recruit in order to account for the processing results. This account is built out of three ideas. I. quantifiers undergo lexical decomposition; II. the comparative's missing part, taken to be a relative-clause-like constituent, is copied from the main clause; III. some word parts cannot compose due to type mismatch, and are forced to QR. More explicitly:

I. <u>Lexical Decomposition</u>: *higher* decomposes into *high+–er*, and *less* into *little+–er*. Lexical entries and types: adjectives contain a degree argument (20a). *–er* is a comparative DE operator over sets of degrees, which checks for an inclusion relation between two degree segments. It is designed to deliver the right meaning when composing with both UE and DE adjectives (20b). *Little* (20c) is a DE operator (essentially, a negation), whose meaning is "not as much as d" (type $<d<<dt>t>>$), where d is a generalized quantifier over degrees (type $<dt,t>$). Both *–er* and *little* are parts of a DegP, but they are blocked from composing due to type mismatch, which forces *–er* to QR at LF. *Can* is the usual existential modal (20d).

(20)   a. $[[high]]^w = \lambda d. \lambda x. Height_w(x)>d$
        b. $[[-er]] = \lambda P_{<d,\ t>}.\lambda Q_{<d,t>}.P \subset Q$
        c. $[[little]] = \lambda d. \lambda P_{<d,t>}.P(d)=0$
        d. $[[can]] = \lambda p.\exists w \in W_{Acc}: p(w)$

---

[12] No such ambiguity is observed with the UE *–er* – (i) only has the "more-than-maximum" reading:
        (i) The plane is flying high-*er* than a helicopter can

II. <u>Copy</u>: at LF, the comparative has a silent complement for *can* – a copy of the complement of the main predicate (excluding tense and *–er*). Very schematically:

(21)   The helicopter is flying high–*er* [than a plane can [fly high]]

III. <u>QR</u>: The main clause and the comparative clause, then, are sets of degrees, which *–er* takes as its arguments. As *-er* cannot compose with *little*, it must raise, which makes it outscope *little* at LF. Nothing blocks either *-er* or *little* from outscoping the modal. As a result, the minimum/maximum ambiguity can be derived, because these scopal orderings are possible:

(22)   a. *er* > *little* > modal
        b. *-er* > modal > *little*

The less-than-minimum/maximum ambiguity is thus derived via two different LFs – *little* works as a negation, and its scope relative to the existential modal determines the meaning:

(23)   a. LF1, the "less-than-minimum" reading of the *less*-comparative:
        [*–er* than]$_1$ wh$_3$[t$_3$ ***little***]$_4$[**can** [a plane fly t$_4$ high]][t$_1$ ***little***]$_2$[The helicopter is flying t$_2$ high]
        b. LF2, the "less-than-maximum" reading of the *less*-comparative:
        [*–er* than]$_1$ wh$_3$[**can** [t$_3$ ***little***]$_4$[a plane fly t$_4$ high]][t$_1$ ***little***]$_2$[The helicopter is flying t$_2$ high]

This scopal account also works to license the NPI in (19), as it posits a DE operator that scopes over the comparative. And while this analysis overgenerates,[13] as Heim herself points out, it nonetheless marks an advance in that it is fully compositional.[14,15]

4.3. Counting DE operators

The setup in (23) opens the door to a new perspective on the processing complexity of comparatives. Assume lexical entries as in (20), by which both *–er* and *little* are DE operators. For each of the polar comparatives (24) we obtain a count of the number of these (25):

(24)   a. X is higher than Y
        b. Y is less high than X

---

[13] Example (19) is unambiguous, as only the "less-than-minimum" reading is available. And yet, Heim's account stops short of blocking the other, "less-than-maximum", reading because it provides no way to block the modal from outscoping *little* when the disambiguating NPI is present.

[14] Another issue pointed out by Heim is the lack of full synonymy between *little* α and its antonym. To get around this problem, Heim proposes to stick to the "non-evaluative, open-scaled, adjectives in the comparative (like *less fast – slower, less old – younger*), which do seem to form truth-conditionally equivalent pairs in simple sentences" (Heim, 2006, p. 21).

[15] Rullman doubts the validity of the restricted decomposition account. He observes that the less-than-minimum/maximum ambiguity is more widespread than expected: in addition to the previous cases (i), it is also attested in sentences where *more* combines with its negative antonym (ii):
        (i) The helicopter is flying *less high* than a jet plane can
        (ii) The helicopter is flying *lower* than a jet plane can
This forces a decomposition of adjectival antonyms, which Rullman stops short of. But Heim and Büring do decompose adjectives, in the spirit of Kennedy (2001), e.g., low=*little*(high).

(25) DE count
     a. *higher* (1): $[$*–er* than$]_1$ $wh_2[Y\ t_2\ high][X\ is\ t_1\ high]$
     b. *less high* (3): $[$*–er* than$]_1$ $wh_3[[t_3$ ***little***$]_4[Y\ t_4\ high]][[t_1$ ***little***$]_2[X\ is\ t_2\ high]]$

Next, we use the measure of processing complexity in order to translate this count into predictions about RT in verification tasks.

5. DEC redefined: adjudicating between the accounts

5.1. A revised DEC Effect

A reminder: the DEC Effect, whose definition is based on the monotonicity properties of environments within sentences, has not been a complete success. It correctly predicted the results of some DE/UE experimental contrasts but failed to account for the results for comparatives: contrary to fact, it predicted that $\Delta RT_{less\text{-}more}=0$, as both *more* and *less* comparatives have a DE and a UE part.

The decompositional analysis we have seen supplies a 3:1 proportion of DE to UE operators in *less*/*more* comparatives. If DEC can be tied to this count, the RT contrast that is repeatedly found would be accounted for. But for that, a change in perspective on the processing complexity of quantifiers is required. DEC will no longer be based on the DE-ness as measured by inferential properties. Rather, it will be expressed in terms of the number of DE operators, n_{DE}, in a given LF.[16] Our proposal builds on $\boldsymbol{n_{DE}}$, the number of DE operators in an LF:[17]

(26)   The DEC Effect (revised)
       $\boldsymbol{n_{DE}(LF_2) > n_{DE}(LF_1) \Rightarrow RT(LF_2) > RT(LF_1)}$.

The decompositional analysis takes *more*-comparatives to contain one DE operator (*–er*) and *less*-comparatives to contain three (*–er, little*). Hence, $\boldsymbol{\Delta n_{DE}=2}$, and the predicted DEC Effect is correct: $\boldsymbol{\Delta RT}=RT_{fewer} - RT_{more}\boldsymbol{> 0}$.

5.2. Fitting our results from polar phrasal comparatives to the new DEC

The above analysis pertains to clausal comparatives, which leaves us one step short of deriving the results of our own experiment, in which we showed that *phrasal* comparatives are fully processed:

(27)   a. There are more blue circles than yellow circles
       b. There are fewer yellow circles than blue circles

Our discussion thus far featured clausal comparatives, and we therefore need to consider whether phrasal ones, whose semantics is slightly different from their clausal counterparts (e.g., Beck, Hohaus and Tiemann, 2012), fit the bill. The experimental sentences in (27) have the

---

[16] The DEC may also at some other semantic representation, if one sticks to an account that does not assume LF in its technical sense.
[17] See also Hackl (2009); Szymanik and Zajenkowski (2010).

same truth conditions. We focus on their logical forms, as these bear on DEC.

The decompositional analysis turns *more* into *many+–er*, and *fewer* into *little+many+er*; the rest follows as in (20)-(23), resulting in LF representations with an unequal number of DE operators, namely one DE operator for *more* (28a), but three for *fewer* (28b):[18]

(28) a. [*–er* than]$_1$ wh$_2$[t$_2$ many yellow circles][there are t$_1$ many blue circles]
    b. [*–er* than]$_1$ wh$_2$[t$_2$ *little*]$_3$[t$_3$ many blue circles][t$_1$ *little*]$_2$[there are t$_2$ many yel. circles]

This analysis, then, once coupled with the new DEC, predicts our results.

The theoretical informativeness of the revised DEC goes beyond the data we discussed: by DEC, differential response times in verification experiments with quantifiers (**ΔRT**) should correlate with the differential count of DE operators (**Δ*n*$_{DE}$**). Experiments that measure the DEC Effect may therefore serve as a tool for the discovery of hidden DE operators through RT patterns. That is, in every case where there is a UE environment due to two DE operators (or more generally, where 2*n*$_{DE}$ =*n*$_{UE}$), we expect response times to be elevated relative to "true" UE environments. The processing signature of such operators, even if covert, should be revealed experimentally. We are currently engaged in further experimentation along these lines.

We note, moreover, that other work in our lab may provide preliminary hints regarding differences between adjectives and quantifiers with respect to hidden negation – differences that RH might welcome, but Büring's approach would not predict, as it gives all negations the same status. At this point, there is little evidence regarding adjectival antonyms and negations that they may conceal, yet Agmon, Deschamps, Loewenstein and Grodzinsky (2016) have begun to explore this issue both behaviorally and though functional MRI. The preliminary evidence they produced suggests a marked difference between the two types: when a two-way ANOVA is performed over data from positive and negative adjectival antonyms and polar quantifiers, an interaction effect emerges. That is, the difference between antonyms is significantly smaller that the difference between polar quantifiers (see also Tucker, Tomaszewicz and Wellwood, 2017 for a recent experimental exploration).

6. Representational and verificational complexity

The new definition of DEC views the representational complexity of a sentence as a central determinant of its processing cost. This complexity is not acknowledged by the standard WS-based view, inspired by Barwise and Cooper, which focuses on the interaction between monotonicity and Truth-value. This view ascribes the longer RTs for DE sentences to the verification procedure and not to the structure of the quantifiers themselves (cf. also Koster-Moeller, Varvoutis and Hackl, 2008). As noted earlier, this algorithm has two predictions for verification experiments with polar quantifiers: <u>a</u>. a Polarity×Truth-value interaction – higher RTs are expected for true DE sentences, compared to their UE counterparts, as is a reversal of this relation for false sentences; <u>b</u>. no main effect – no difference is expected between the mean RTs for DE and UE sentences (Figure 1a). Prediction <u>a</u> is borne out, but prediction <u>b</u> is false (Figure 1b). The present view links complexity to the number of DE operators at LF to explain

---

[18] Additional assumptions regarding the copying process in the context of phrasal ellipsis are suppressed, as they are orthogonal to the issue under consideration, namely, they do not affect the number of DE operators at LF.

the main effect. The interaction, we argue, is explicable by the WS algorithm. It follows that *a complete account of the RT data requires two components: representational and verificational.* This analysis of the processing data receives further support when data from several studies are broken down by Truth-value and compared. Consider Figure 6a, with Deschamps et al.'s results from a UE/DE pair of simple quantified sentences are analyzed by the factor Truth-value, and compare it to a similar breakdown of the present experiment with comparatives (Figure 6b):
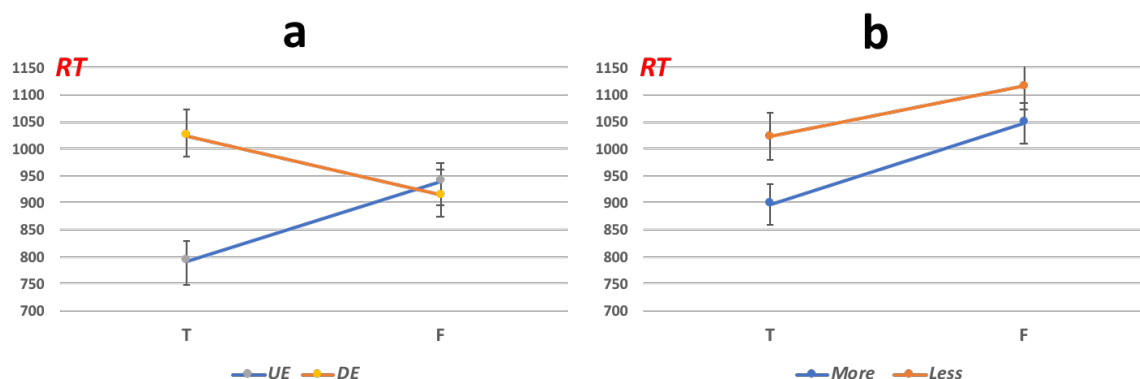

Figure 6: Polarity×Truth-value interaction in two experiments with sentences containing *more* (**blue**) vs. *less* (**orange**). **a**. Deschamps et al.'s results from quantifiers in simple sentences – disordinal interaction effect. **b**. the comparatives experiment – ordinal interaction effect. Error bars mark SEM.

We see a nuanced picture: both studies evince a DEC Effect, attributed to representational complexity. A Polarity×Truth-value interaction is also observed in both, but it takes different shapes. The interaction for the simple sentences [$F(1,16)=14.755$, $p<0.001$] is disordinal, with lines whose slopes go in different directions (Figure 6a). The (smaller) effect for the comparatives [$F(1,21)=7.84$, $p<0.02$]), by contrast, is ordinal – the lines in Figure 6b, though not parallel, nonetheless have slopes in the same direction.

Assuming verification by the WS strategy, this contrast is expected: when applied to simple quantificational sentences, this strategy expects performance inversion when Truth-value switches (see Bott, Klein and Schlotterbeck, 2013, Szymanik, 2016). For comparatives, no such strategy can be employed, and the lines are indeed near parallel. The slopes in Figure 6b (the difference between RTs for true and false sentences) remain unaccounted for.

Thus, despite various objections to the WS approach to verification, the experimental results, once properly handled, appear in keeping with this approach in places where it applies. We make no claim, though, regarding the algorithm that is used to verify comparatives.

7 Final thoughts

The robust DEC Effect found for phrasal comparatives (i) sharpens our view of the way the processing cost of DE-ness is manifested and shows that the complexity of quantifier processing is bi-componential; (ii) supports a decompositional view of *less*-comparatives, and (iii) underscores the value of experimental work as a powerful tool for the discovery of hidden linguistic structure.

References

Agmon, Galit, Isabelle Deschamps, Yonatan Loewenstein and Yosef Grodzinsky (2016). Less is more: differences in the processing of negative polarity in quantifiers and in adjectives. Society for the Neuroscience of Language Conference, London.

Barwise, Jon, and Robin Cooper (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy, 4*(2), 159-219.

Beck, Sigrid, Vera Hohaus and Sonja Tiemann (2012). A note on phrasal comparatives. In *Semantics and Linguistic Theory, 22*, 146-165.

Bott, Oliver, Udo Klein and Fabian Schlotterbeck (2013). Witness Sets, Polarity Reversal and the Processing of Quantified Sentences. In *Proceedings of the Amsterdam Colloquium*, 59-66.

Büring, Daniel (2007a). More or less. *Proceedings from the annual meeting of the Chicago Linguistic Society, 43.2*, 3-17. Chicago Linguistic Society.

Büring, Daniel (2007b). Cross-polar nomalies. *Semantics and Linguistic Theory, 17*, 37-52.

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

Geurts, Bart, Napoleon Katsos, Chris Cummins, Jonas Moons and Leo Noordman (2009). Scalar quantifiers: Logic, acquisition, and processing. *Language and Cognitive Processes, 25.1*, 130-148.

Geurts, Bart and Frans van Der Slik (2005). Monotonicity and processing load. *Journal of semantics*, *22. 1*, 97-117.

Hackl, Martin (2009). On the grammar and processing of proportional quantifiers: most versus more than half. *Natural Language Semantics, 17*(1), 63-98.

Heim, Irene (2006). Little. In M. Gibson and J. Howell (eds.) *Proceedings of SALT XVI*, CLC Publications, Cornell University, Ithaca.

Just, Marcel and Patricia Carpenter (1971). Comprehension of negation with quantification. *Journal of Verbal Learning and Verbal Behavior, 10*(3), 244-253.

Koster-Moeller, Jorie, Jason Varvoutis and Martin Hackl. (2008). Verification procedures for modified numeral quantifiers. In Proceedings of the 27th West Coast Conference on Formal Linguistics, Somerville, MA: Cascadilla Press.

Rullmann, Hotze (1995). *Maximality in the Semantics of Wh-Constructions*. UMass Amherst Ph.D. thesis.

Schwarzschild, Roger (2008). The semantics of comparatives and other degree constructions. *Language and Linguistics Compass*, *2.2*, 308-331.

Seuren, Pieter (1973). The comparative. *Generative grammar in Europe,* 528-564. Springer, Dordrecht.

Seuren, Pieter (1978). "The Structure and Selection of Positive and Negative Gradable Adjectives." In W.M. Jacobsen Farkas, Donka and K.W. Todrys, eds., Papers from the Parasession on the Lexicon, *Chicago Linguistic Society*, 336-346. Chicago: Chicago Linguistic Society.

Schwarzschild, Roger and Karina Wilkinson: 2002, 'Quantifiers in Comparatives: A Semantics of Degree Based on Intervals', *Natural Language Semantics, 10*, 1-41.

Szymanik, Jakub (2016). *Quantifiers and Cognition. Logical and Computational Perspective*s, Studies in Linguistics and Philosophy, Springer.

Szymanik, Jakub and Marcin Zajenkowski (2010). Quantifiers and working memory. *Lecture Notes in Computer Science, 6042*, 456-464.

von Stechow, Arnim (1984). "My Reaction to Cresswell's, Hellan's, Hoeksema's, and Seuren's Comments." *Journal of Semantics 3*, 183-199.

Tucker, Daniel, Barbara Tomaszewicz and and Alexis Wellwood (2017). Decomposition and processing of negative adjectival comparatives. Ms. Northwestern.