Licensing Sentence-internal Readings in English

An Experimental Study

Adrian Brasoveanu & Jakub Dotlačil

Linguistics, UCSC, 1156 High St., Santa Cruz, CA 95064
abrsvn,j.dotlacil@gmail.com

Abstract. Adjectives of comparison (AOCs) like *same*, *different* and *similar* can compare two elements sentence-internally, i.e., without referring to any previously introduced element. This reading can only be licensed if a semantically plural NP is present. In contrast to almost all previous literature, we argue in this paper that it is incorrect to describe a particular NP as either licensing or not licensing the sentence-internal reading of a specific AOC: licensing is more fine-grained. We use experimental methods to establish which NPs license which AOCs and to what extent. We show how the results can be interpreted against the background of a formal semantics analysis of AOCs and we argue that using Bayesian methods to analyze this kind of data has several advantages over the more traditional, frequentist approach.

1 The phenomena

Most, if not all, languages have lexical means to compare two elements and express identity / difference / similarity between them. English uses adjectives of comparison (henceforth AOCs) like *same*, *different* and *similar* for this purpose. Often, the comparison is between an element in the current sentence, e.g., the italicized NP *the same movie* in (1b) below, and a sentence-external element mentioned in the previous discourse, e.g., the underlined NP 'Waltz with Bashir' in (1a). AOCs can also compare sentence-internally, that is, without referring to any previously introduced element, as shown in (2). In this kind of cases, the sentence itself, as it were, provides the context for the comparison, hence the label of **sentence-internal** reading.

(1) a. Arnold saw 'Waltz with Bashir'.
    b. Heloise saw *the same movie* / *a different movie*.

(2) Each of the students saw *the same movie* / *a different movie*.

The sentence-internal reading is available only if the sentence in which the AOC occurs also contains a semantically (but not necessarily morphologically) plural noun. Importantly, not all semantically plural NPs can license sentence-internal readings of AOCs. This has already been observed in previous literature on the topic (see [1], [2], [3], [4], [5], [8], among others). The previous literature also noted that many NPs license sentence-internal readings of only some AOCs (see [3] for a recent detailed discussion and summary of the previous literature).

However, it is much less known that the majority of semantically plural NPs cannot be described as either licensing or not licensing the sentence-internal reading of a specific AOC. Licensing is more fine-grained. The gradient nature of AOC licensing has not been acknowledged with the exception of [5] for *different* and has not been systematically studied. In this paper we report one experiment that begins to address this issue by establishing which NPs license which AOCs and to what extent. Furthermore, we argue that using Bayesian methods to analyze the resulting experimental data has several advantages over the more traditional, frequentist approach. The paper concludes with a brief discussion of the consequences of the experimental results for the semantic analysis of AOCs.

2 Experiment

2.1 Method

We used questionnaires to test people's intuitions about sentence-internal readings of three AOCs—*same*, *different* and *similar*, with four licensors—NPs headed by *each*, *all*, *none* and *the* (for a total of $3 \times 4 = 12$ conditions). Each condition was tested four times, twice in a scenario in which the condition was most likely judged as true and twice in a scenario in which the condition was most likely judged as false. There were 32 fillers.

An example of a scenario and three test items testing the sentence-internal reading of *similar*, *same* and *different* is given below. In the actual setup, each scenario was followed by five items, two of which were fillers. For each scenario, each of its corresponding test items had a different AOC and a different licensor.

(3) Gustav, Ryan and Bill are three bank managers who share a passion for Volvo, Rolls Royce and Porsche automobiles. Last year, each of them bought a new car. Gustav bought a Volvo PY30, Ryan bought a Volvo XRT2000 and Bill bought a Volvo H4.

a. Each of the bank managers chose a similar car.

b. All the bank managers chose the same car brand.

c. None of the bank managers chose a different car brand.

Each item was judged with respect to (*i*) TRUTH: whether it is true, false or unknown given the accompanying scenario and (*ii*) ACCEPT(ABILITY): how acceptable it is on a 5-point scale (5=completely acceptable, 1=completely unacceptable). TRUTH was measured so that it could be distinguished from ACCEPT.

A total of 42 subjects in two undergraduate classes at UCSC completed the questionnaire for extra-credit. For each subject, we randomized both the order of the scenarios in the questionnaire and the order of the items for each scenario. We excluded two subjects because of their incorrect responses to fillers and one because only TRUTH was completed; one of the remaining 39 subjects filled in only three fourths of the questionnaire. Final number of observations: $n = 1856$.

Barplots of ACCEPT for the 12 conditions are shown in Figure 1, from the least acceptable, i.e., sentence-internal *different* when the licensor NP is headed by *none*, to the most acceptable, i.e., sentence-internal *same* when the licensor NP is headed by *all*.

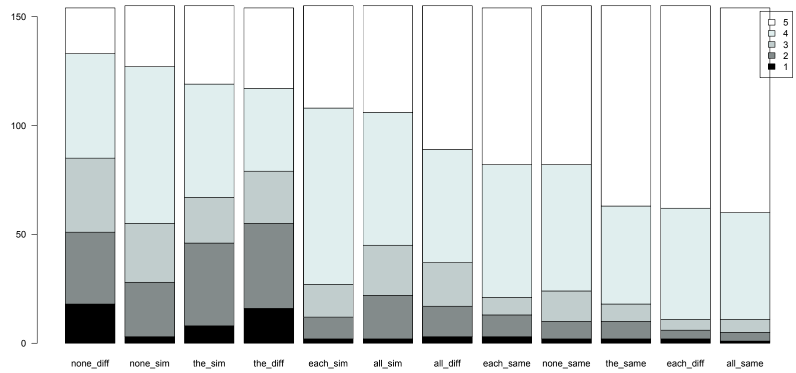

Fig. 1. Barplots of responses by quant-AOC combination

2.2 Statistical modeling and resulting generalizations

The response variable ACCEPT is ordinal, so we use ordered probit regression models to analyze the data. These models are similar to linear regression models in that the predictors are linearly combined and the weights / coefficients for each predictor are estimated from the data. The linear combination of predictors provides the mean for a normal distribution with a fixed variance (set to 1^2 for simplicity). To put it differently, the linear combination of predictors provides an 'offset' for the mean of the standard normal distribution. The area under the probability density function obtained in this way is partitioned into five regions by four thresholds (also estimated from the data) and each region corresponds to one value of the ordinal variable.

We have 2 fixed-effects predictors: (*i*) QUANT-AOC—factor with 12 levels since we have 12 licensor-AOC combinations, reference level: the *each-different* combination; (*ii*) TRUTH—factor with 3 levels T(rue), F(alse), U(nknown), reference level: T. Our main interest is in how QUANT-AOC affects ACCEPT while controlling for / factoring out the influence of TRUTH on ACCEPT.

A frequentist analysis shows that adding either of the fixed effects to the null (intercept-only) model significantly decreases deviance, but the interaction of the fixed effects does not (p=0.31). That is, licensor-AOC combinations and truth-value judgments significantly and additively influence acceptability judgments. Adding intercept-only random effects for items accounts for practically no variance, but adding random effects for subjects does (std.dev=0.56). Thus, the final regression model $\mathcal{M}$ we henceforth focus on has 2 fixed effects, QUANT-AOC and TRUTH (no interaction), and intercept-only random effects for subjects.

Our primary interest is to establish which NPs license sentence-internal readings of which AOCs and to what extent. That is, we are interested in a wide range of pairwise comparisons between various licensor-AOC combinations. But doing this in the null-hypothesis significance testing framework would require an unfeasibly large amount of data to achieve significance given the necessary α-level correction for running all pairwise comparisons between the 12 licensor-AOC combinations (66 comparisons in total).

In contrast, any number of pairwise comparisons can be carried out in a Bayesian framework because we do not use p-values as a criterion for decision making. Instead, we simply study the multivariate posterior distribution of the parameters obtained given our prior beliefs, the data and our mixed-effects order probit regression model $\mathcal{M}$. Pairwise comparisons of various licensor-AOC combinations are just different perspectives on, i.e., different ways of marginalizing over, this posterior distribution (see [6], [7] and references therein for more discussion). To determine whether there is a credible difference between any two conditions, we check whether 0 (=no difference) is in the 95% highest posterior density interval (HDI; basically, a 95% confidence interval) of the difference: if 0 is outside the HDI, the two conditions are credibly different.

The Bayesian model we estimate has the following structure: (i) we assume low-information / vague priors for the non-reference levels of QUANT-AOC and TRUTH—independent normal distributions with mean 0 and variance 10^2; (ii) the subject random effects are assumed to come from a normal distribution with mean 0 and variance σ^2, with σ taken from a uniform distribution $\mathit{Unif}(0, 10)$. As we indicated above, the function linking the linearly combined predictors and the response ordinal value is the standard normal cumulative distribution function Φ. The range of Φ is partitioned into five intervals (since the acceptability scale was 1–5) by 4 cutoff points / thresholds; the low-information priors for the thresholds are also independent normal distributions with mean 0 and variance 10^2. We estimate the posterior distributions of the predictors QUANT-AOC and TRUTH, the standard deviation σ of the subject random effects and the 4 thresholds by sampling from them using Markov Chain Monte Carlo techniques (3 chains, $125,000$ iterations per chain, we discard the first $25,000$ iterations and record only every 50^{th} one).

The posterior histograms for the most relevant comparisons are shown in Figures 2–4 below, grouped by AOC. The resulting generalizations are summarized at the top of each set of plots, where > means the licensor(s) on the left is / are preferred to the licensor(s) on the right.

Figure 2 below shows that *each* is a better licensor of sentence-internal *different* compared to *all*, which in turn is better than definite plurals and negative quantifiers. However, we cannot confidently distinguish between definite plurals and negative quantifiers since the HDI of the difference between them includes 0. Nonetheless, it is likely that slightly more data would enable us to distinguish between the acceptability of these last two licensors—in contrast to the rightmost plot in Figure 3, for example, where the difference between two licensors is virtually non-existent.

Thus, examining the posterior probability distributions reveals not only whether two conditions are different or not, but also how confident we are in postulating a difference and the magnitude of this difference. The corresponding generalizations for the licensors of *same* and *similar* are provided in Figures 3 and 4.

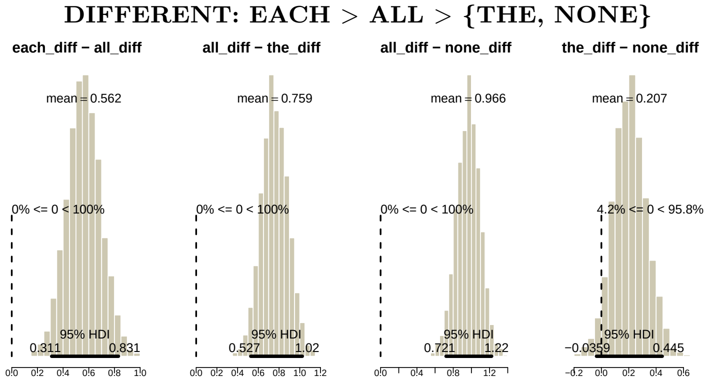

Fig. 2. Differences in acceptability between licensors of *different*

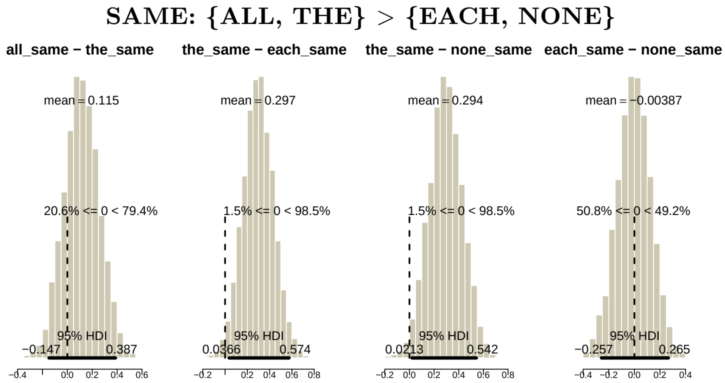

Fig. 3. Differences in acceptability between licensors of *same*

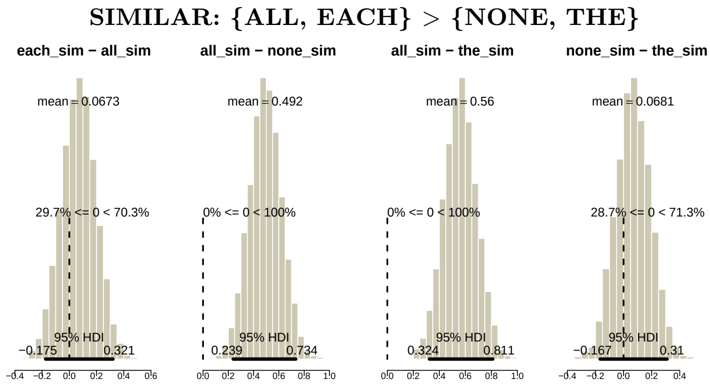

Fig. 4. Differences in acceptability between licensors of *similar*

Finally, Figure 5 shows the posterior distributions of the two non-reference levels of TRUTH and the four thresholds. False sentences (F) and sentences whose truth values are unknown (U) (due to their grammatically unclear status) lead to lower acceptability compared to true sentences. The rightmost plot shows the mean posterior thresholds plotted together with the standard normal probability density function. The fourth (rightmost) threshold, for example, is the cutoff point between values 4 and 5 of the ACCEPT response variable. We see that value 5 has the highest probability (the largest area under the probability density function).

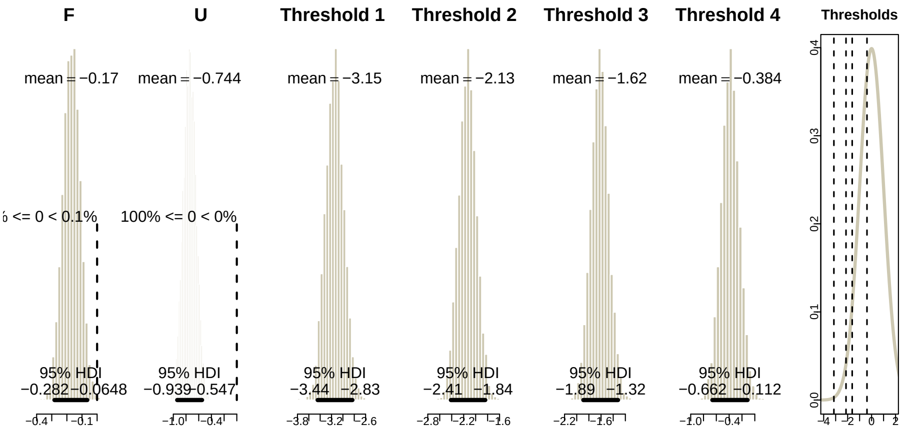

Fig. 5. Posterior distributions of TRUTH and thresholds

3 Consequences for the semantic analysis of AOCs

There is a long tradition of connecting the sentence-internal reading of at least some AOCs with distributivity. Here, we follow [3], who analyzes AOCs in a dynamic semantics that provides semantic values for natural language expressions in terms of sets of sequences of individuals and models these sequences as stacks.

Consider (4) below and the sequence of figures in (5a)–(5c) depicting the sequence of dynamic updates contributed by (4). The update contributed by *each boy* stores all the boys as the value of some variable, u_0 in our example. This is pictorially depicted by the one-column table following the leftmost arrow in (5a). The interpretation of the distributive operator **dist** contributed by *each boy* and of sentence-internal *different* are pictorially depicted in (5b): **dist** provides a temporary context inside of which the interpretation proceeds in three steps, namely (*i*) pick two distinct boys, (*ii*) check that each of the two boys recited a poem and (*iii*) check that the two poems are different. In (5b-i), this sequence of steps is depicted for *boy*$_1$ and *boy*$_2$ and their corresponding poems. But **dist** requires these three steps to be repeated for any pair of boys in the set u_0, as shown in (5b-i)–(5b-v). For more details and the exact logical formulas, see [3].

(4) Each^{u_0} boy recited a^{u_1} different$^{2}_{u_1}$ poem.

(5) a. $\emptyset \xrightarrow{\text{Each}^{u_0}\text{boy}} \begin{array}{c} u_0 \\ \hline \multicolumn{1}{|c|}{boy_1} \\ \hline \multicolumn{1}{|c|}{boy_2} \\ \hline \multicolumn{1}{|c|}{boy_3} \\ \hline \end{array} \xrightarrow{\textbf{dist}_{u_0}(\text{recited a}^{u_1}\,\text{different}^2_{u_1}\,\text{poem})}$

b. $$\left\{ \begin{array}{l} \text{i. } \begin{array}{cc} u_0 & u_1 \\ \hline \multicolumn{1}{|c|}{boy_1} & \multicolumn{1}{c|}{poem_1} \\ \hline \end{array} * \begin{array}{cc} u_0 & u_1 \\ \hline \multicolumn{1}{|c|}{boy_2} & \multicolumn{1}{c|}{poem_2} \\ \hline \end{array} \;\; \& \;\; poem_1 \neq poem_2 \\[2ex] \text{ii. } \begin{array}{cc} u_0 & u_1 \\ \hline \multicolumn{1}{|c|}{boy_1} & \multicolumn{1}{c|}{poem_1} \\ \hline \end{array} * \begin{array}{cc} u_0 & u_1 \\ \hline \multicolumn{1}{|c|}{boy_3} & \multicolumn{1}{c|}{poem_3} \\ \hline \end{array} \;\; \& \;\; poem_1 \neq poem_3 \\[2ex] \text{iii. } \begin{array}{cc} u_0 & u_1 \\ \hline \multicolumn{1}{|c|}{boy_2} & \multicolumn{1}{c|}{poem_2} \\ \hline \end{array} * \begin{array}{cc} u_0 & u_1 \\ \hline \multicolumn{1}{|c|}{boy_1} & \multicolumn{1}{c|}{poem_1} \\ \hline \end{array} \;\; \& \;\; poem_2 \neq poem_1 \\[2ex] \text{iv. } \begin{array}{cc} u_0 & u_1 \\ \hline \multicolumn{1}{|c|}{boy_2} & \multicolumn{1}{c|}{poem_2} \\ \hline \end{array} * \begin{array}{cc} u_0 & u_1 \\ \hline \multicolumn{1}{|c|}{boy_3} & \multicolumn{1}{c|}{poem_3} \\ \hline \end{array} \;\; \& \;\; poem_2 \neq poem_3 \\[2ex] \text{v. etc.} \end{array} \right\}$$

c. $\xrightarrow{\text{sum all updates}} \begin{array}{cc} u_0 & u_1 \\ \hline \multicolumn{1}{|c|}{boy_1} & \multicolumn{1}{c|}{poem_1} \\ \hline \multicolumn{1}{|c|}{boy_2} & \multicolumn{1}{c|}{poem_2} \\ \hline \multicolumn{1}{|c|}{boy_3} & \multicolumn{1}{c|}{poem_3} \\ \hline \end{array} \quad \text{where} \quad \begin{array}{l} boy_1 \text{ recited } poem_1 \\ boy_2 \text{ recited } poem_2 \\ boy_3 \text{ recited } poem_3 \end{array}$

Thus, in this account, the **dist** operator distributes over *pairs* of individuals and is necessary for sentence-internal readings of *different*. Besides the pairwise distributivity, [3] postulates another operator, **dist-Comp**, which creates a temporary context consisting of an individual and stacks that store all entities in the domain of quantification different from that individual. In (4), the **dist-Comp** operator would create a context comparing each boy and all the other boys.

Both the **dist** and the **dist-Comp** operator can capture distributive interpretations (hence their label **dist**), and both of them can account for sentence-internal readings of *different* and *same*. However, sentence-internal readings of *similar* seem to be compatible only with **dist-Comp**: similarity is computed over the entire domain of quantification, which **dist-Comp** provides, and not simply over the individual pairs contributed by **dist**.

Consider the following example:

(6) Each manager bought a similar car.

Suppose there are three managers and two of them bought the same car brand, say, Volvo. The third manager bought a BMW, the color of which is similar to one Volvo and the design of which is similar to the other Volvo. In that case, it is true that for each pair of cars, the paired cars are similar (in some respect)—but (6) is intuitively false. We can capture this if sentence-internal *similar* is licensed by **dist-Comp** as opposed to **dist**.

Finally, *same* (and plural *different*, which we do not discuss in this paper) has another interpretation that gives rise to sentence-internal readings. If there is no distributivity in the clause, then *same* checks that only one entity (possibly plural) was introduced by its NP. The whole system of sentence-internal readings and their licensors is summarized in Table 1.

	dist	**dist-Comp**	no distributivity
different	✓	✓	*
same	✓	✓	✓
similar	*	✓	*

Table 1. Distributivity and sentence-internal readings in [3]

We are now going to discuss how this analysis, along with other accounts of sentence-internal readings, can account for the data from our experiment.

It has been observed in [5] that the distributive interpretation of predicates like *build a snowman* depends on the type of subject. In particular, the following cline in the availability of the interpretation was observed, where > means the determiners / quantifiers on the left are preferred to the determiners / quantifiers on the right.

(7) Distributive interpretation: **EACH** > **ALL** > **THE**

The parallelism between the gradience of distributivity 'strength' associated with these determiners / quantifiers and the gradience of acceptability associated with sentence-internal readings of *different* provides support for accounts in which sentence-internal *different* requires distributivity to be licensed, as is the analysis of [3] discussed above, as well as [2], [4], [5], [8]. This is true regardless of the explanation for the gradient nature of distributivity 'strength' (but see [5] for one account).

At the same time, the results are problematic for accounts like [1], in which sentence-internal readings are incompatible with distributively interpreted licensors. From the perspective of [1], we would expect that **all** and **the** are better licensors than **each**, contrary to the facts.

Finally, none of the current accounts can explain why negative quantifiers are dispreferred licensors for *different*. These points have already been made in [5] with respect to the Dutch data. This paper extends them to English.

Regarding *same*, we have seen the following ordering of licensors:

(8) Same: **{ALL, THE}** > **{EACH, NONE}**

The difference between the two orderings of licensors supports the account of *same* in [1]. Under that analysis, *same* should not give rise to sentence-internal readings with distributive quantifiers, which squares well with the degraded status of **each** and **none**. The remaining question is why **each** and **none** are only slightly degraded, not uninterpretable, as the account in [1] would predict.

One possibility is that *same* is ambiguous, as discussed above and as assumed in [3] and [8]. One of the two meanings for *same* needs to appear in the scope of **dist** to have a sentence-internal reading, while the other meaning is compatible with a non-distributive plural licensor (see Table 1). Given the ordering in (8), the former meaning must be dispreferred / less accessible. Thus, our experiment seems to provide evidence for an ambiguity account of *same*, even though we still need to explain why one meaning of *same* should be preferred over the other.

One possibility is that the more complex meaning is dispreferred. Consider (9a) below: under the account in [3], **dist** creates temporary contexts storing pairs of non-identical boys and *same* needs to check that within each pair, the recited poems are identical. In contrast, the meaning of *same* in (9b) only needs to check that exactly one poem was introduced in discourse by the direct object. This second meaning of *same* is simpler in that we do not need to repeatedly examine pairs of poems (stored in two distinct discourse referents), we simply contribute a cardinality requirement on a set of witnesses. An investigation of the hypothesis that meaning (and / or processing) complexity can explain the licensing gradience in (8) is left for future research.

(9) a. Each boy recited the same poem.
    b. All the boys/The boys recited the same poem.

Finally, sentence-internal *similar* is associated with the following scale of licensors:

(10) Similar: **{ALL, EACH}** > **{NONE, THE}**

The scale in (10) indicates that *similar* is close to *different*. The only difference between the two is that *similar* does not distinguish between **all** and **each**. These fine-grained parallelisms and contrasts between *similar* and *different* (or *same*) have not been previously noticed so far, as far as we know. As indicated above, the account in [3] generalizes to *similar* if we stipulate that NPs have another way of introducing distributivity, **dist-Comp**. But once again, the degraded status of negative quantifiers is surprising. While the account is stipulative, there are currently no better alternatives, as far as we know. A more careful specification and investigation of this account, including the contrast between **dist** and **dist-Comp** on the theoretical side and the contrasts between *similar*, *same* and *different* on the empirical side, is also left for future research.

4 Conclusion

We have discussed experimental evidence showing that licensing sentence-internal readings of AOCs is gradient in nature. We have argued that this gradience supports an analysis of sentence-internal readings that connects them with distributivity. Furthermore, the particular ordering of licensors for *same* vs. *different* vs. *similar* provides evidence for an ambiguity account of *same*, as well as for two different distributivity operators. Some issues, like the particular status of negative quantifiers as licensors of *different* and *similar*, remain unclear and are left for future research.

Bibliography

[1] Barker, C.: Parasitic scope. Linguistics and Philosophy 30, 407–444 (2007)

[2] Beck, S.: The semantics of Different: Comparison operator and relational adjective. Linguistics and Philosphy 23, 101–139 (2000)

[3] Brasoveanu, A.: Sentence-internal *Different* as quantifier-internal anaphora. Linguistics and Philosophy 34, 93–168 (2011)

[4] Carlson, G.: Same and Different: some consequences for syntax and semantics. Linguistics and Philosphy 10, 531–565 (1987)

[5] Dotlačil, J.: Anaphora and Distributivity. A study of *same*, *different*, reciprocals and *others*. Ph.D. thesis, Utrecht University, Utrecht (2010)

[6] Kruschke, J.K.: Bayesian data analysis. WIREs Cognitive Science 1, 658–676

[7] Kruschke, J.K.: Doing Bayesian Data Analysis: A Tutorial with R and BUGS. Academic Press/Elsevier, Oxford

[8] Moltmann, F.: Reciprocals and *same/different*: Towards a semantic analysis. Linguistics and Philosophy 15(4), 411–462 (1992)