Polarity particles revisited

Farkas, D.F.; Roelofsen, F.

DOI
10.3765/sp.12.15
Publication date
2019
Document Version
Final published version
Published in
Semantics and Pragmatics

Semantics & Pragmatics Volume 12, Article 15, 2019
https://doi.org/10.3765/sp.12.15

EARLY ACCESS

Polarity particles revisited*

Donka F. Farkas and Floris Roelofsen

June 18, 2019

1 Introduction

Claus, Meijer, Repp, and Krifka (2017) present novel experimental evidence concerning polarity particle responses in German, and discuss the challenges these findings raise for various approaches to such responses, namely the *salience account* in Krifka (2013), the *feature model* in Roelofsen and Farkas (2015), and the *ellipsis approaches* in Kramer and Rawlins (2012) and Holmberg (2013, 2016). The authors then sketch a way to account for the data within the feature model, as well as a revised version of the salience account.

Our first goal is to work out in detail an account of the new German data in the feature model. In the process, we clarify and better articulate those aspects of the model that are responsible for explaining preference patterns. Our second goal is to deepen the comparison between this model and the salience approach by taking a wider cross-linguistic perspective.

The structure of the paper is as follows. Section 2 summarizes the feature model, focusing both on relevant language specific analyses and general typological predictions. Section 3 turns to the findings of Claus et al. (2017), and develops an account of them within the feature model. Section 4 compares the feature model and the salience account from a wider cross-linguistic perspective, and Section 5 concludes.[1]

2 Synopsis of the feature model

The earliest work on polarity particles—*yes*, *no*, and their counterparts in other languages—concentrated on cross-linguistic differences in the use of such particles in agreeing responses to negative initiatives, exemplified in (1) (Kuno, 1973; Pope, 1976; Sadock and Zwicky, 1985).

(1) A: Sam is not home.
B: Yes/No, he isn't.

Languages were claimed to be sensitive to one of two parameters, which we dub here, for reasons that will become evident soon, the *relative* and the *absolute* parameter. The relative parameter concerns the relationship between the response and its antecedent, i.e., whether the response agrees

*We are grateful to our editor Kristen Syrett and three anonymous reviewers for very detailed and helpful comments, to Berry Claus, Manfred Krifka, Marlijn Meijer, Sophie Repp, Michael Wagner, and the participants in a UCSC seminar on sentence types for very useful discussion, and to Hitomi Hirayama and Wataru Uegaki for Japanese judgments. Finally, we gratefully acknowledge financial support from the Netherlands Organisation for Scientific Research (NWO).

[1]For reasons of space, we do not discuss the ellipsis approaches here. These are compared with the feature model in some detail in Roelofsen and Farkas (2015), and evaluated against the German data in Claus et al. (2017).

with or rejects the antecedent. The absolute parameter concerns the polarity of the response itself. In languages said to be sensitive to the relative parameter, known in the literature as *truth-based* languages (e.g., Chinese and Japanese), agreeing responses to negative antecedents pattern with agreeing responses to positive antecedents: the same 'agreeing' particle is used in both cases. In languages said to be sensitive to the absolute parameter, known in the literature as *polarity based* languages (e.g., French and Swedish), agreeing responses to negative antecedents pattern with reversing responses to positive antecedents: the same 'negative' particle is used in both cases.[2]

Building on Farkas and Bruce (2010), Roelofsen and Farkas (2015) (henceforth R&F) replace this binary typological classification by a finer-grained system that extends naturally to languages in which both parameters appear to be relevant. The feature model they develop implements this idea by proposing that polarity particles in all languages realize two types of polarity features, *relative* polarity features and *absolute* polarity features, which connect to the relative and the absolute parameter, respectively.

(2) a. *Relative polarity features*: [AGREE] and [REVERSE]
b. *Absolute polarity features*: [+] and [−]

It is assumed that these features characterize polarity particles across languages. Polarity responses are taken to involve a polarity head that hosts a relative and an absolute polarity feature, and a prejacent clause, which may be fully or partially elided.

R&F assume that sentences introduce propositional discourse referents, and that these discourse referents are marked for polarity; the polarity of a discourse referent is − if introduced by a negative sentence, and + otherwise. The semantic contribution of the polarity features is as follows (see R&F for formal details).

(3) *Relative features*
a. The relative polarity feature [AGREE] encodes the presupposition that the immediately preceding discourse provides a unique most salient propositional discourse referent that agrees with the prejacent in polarity and interpretation.
b. The relative polarity feature [REVERSE] encodes the presupposition that the immediately preceding discourse provides a unique most salient propositional discourse referent whose interpretation and polarity are the opposite of that of the prejacent.

(4) *Absolute features*
a. The absolute polarity feature [+] encodes the presupposition that the polarity of the prejacent is positive.
b. The absolute polarity feature [−] encodes the presupposition that the polarity of the prejacent is negative.

Given this characterization, polarity features can only occur in responses, and the absolute polarity feature must agree with the sentence polarity of the prejacent.

Polarity features are realized morphologically by polarity particles. Which features are realized by which particles is a language specific matter, captured in the feature model by two factors. First, a language specific *feature-particle mapping* associates each response particle in the language to the feature(s) or feature-combination(s) that can be realized by that particle. Second, the form of

[2]Sadock and Zwicky (1985) recognize a third type of language, one which relies primarily on an 'echo' strategy: polarity responses repeat some part of the verb in the antecedent, enough to overtly mark the polarity of the response. It has been noted early on (see, for instance, Jones 1999 on Welsh) that languages may use both polarity particles and the echo strategy. We return to this point below.

polarity responses may be affected by *realization constraints*, which favor the realization of certain features over others.[3]

This model makes three interconnected predictions. First, it predicts the possibility of 'mixed' languages, i.e., languages whose particle systems are sensitive to both parameters simultaneously. Such a language may involve 'double-duty' particles, that is, its feature-particle mapping may connect a single particle to both absolute and relative features. Second, the model predicts the possibility of languages with more than two polarity particles. For instance, a language may have particles realizing the two absolute features, and in addition it may have particles realizing relative features. Third, it is predicted that languages may have particles that realize certain feature *combinations*, for instance [REVERSE,+] or [AGREE,−], in addition to particles realizing individual features.

Building on Pope (1976), R&F propose that the use of polarity responses in a particular language, as well as the typological space of such systems is delimited by a series of markedness distinctions concerning polarity features. In particular, markedness considerations constrain feature-particle mappings in the case of a multi-functional particle, that is, a particle that may realize either a relative or an absolute feature. Given that [AGREE] and [+] are the unmarked relative and absolute feature, respectively, and [REVERSE] and [−] the marked ones, in cases of multifunctionality [AGREE] and [+] will be connected, by harmonic alignment, to one particle, and [REVERSE] and [−] to another. What would be unexpected is a language with two particles, where one can realize [AGREE] and [−], and the other [REVERSE] and [+].

The English polarity particle system In R&F the feature-particle mapping in (5) is given for the English particles *yes* and *no*:[4]

(5) *Feature-particle mapping for English yes and no*
   a. *yes* can realize [AGREE] and [+]
   b. *no* can realize [REVERSE] and [−]

Under this proposal, English is a mixed language in the sense defined above: its polarity particles are sensitive to both the relative and the absolute parameters.

This account correctly predicts that in [AGREE,+] responses, *yes* can be used but *no* can not, while in [REVERSE,−] responses, *no* can be used but *yes* can not. It also explains why both particles can be used in agreeing responses to negative antecedents, i.e., in [AGREE,−] responses, as well as in reversing responses to negative antecedents, i.e., in [REVERSE,+] responses:[5]

(6) A: Paul did not call.
   B: No, he didn't. / Yes, he didn't. [AGREE,−]

(7) A: Paul did not call.
   B: No, he DID. / Yes, he DID. [REVERSE,+]

R&F further claim that in languages like English, where a choice between *yes* and *no* is allowed in [AGREE,−] responses, each particle realizing one of the two features carried by the response, markedness considerations may be called upon to explain preference patterns. Other things being

[3] We should note that the terminology chosen here differs from that in Roelofsen and Farkas (2015), where a single term was used to refer both to feature-particle mappings and realization constraints.

[4] As noted in R&F, footnote 22, English has an additional particle, *yeah*, which can realize [AGREE]. In [AGREE,−] responses, the combination *yeah no* is also possible, where *yeah* realizes [AGREE] and *no* realizes [−]. More empirical research is needed to determine the precise distribution of *yeah* and *yeah no*.

[5] For the role of intonation in [REVERSE,+] responses in English, see Goodhue and Wagner (2018).

equal, in such cases one expects a preference for realizing a marked feature over an unmarked one. This preference is rooted in a general pressure for the overt realization of marked elements (see, for instance, Horn 1984, on the general connection between marked meaning and marked form). Thus, given that in [AGREE,−] responses [−] is marked and [AGREE] is unmarked, other things being equal, markedness considerations lead one to expect *no* to be preferred over *yes* in such responses. This prediction is confirmed by experimental evidence in Brasoveanu et al. (2013) for English for responses to simple negative sentences containing no quantifiers. The results, however, also show that other things are not always equal: the preference for *no* in [AGREE,−] responses disappears in case the antecedent is more complex. What factors influence this preference in these more complex cases remains an open issue. But these data show that markedness cannot be the only factor in accounting for preference patterns in case the grammar allows a choice of particles for a given response.

Beyond English R&F discuss two types of languages with three polarity particles, namely (i) languages whose third particle realizes [REVERSE], the marked relative feature, and (ii) languages whose third particle realizes [REVERSE,+], the most marked feature combination.

The two languages in the [REVERSE] third particle group for which R&F propose a detailed analysis are Hungarian and Romanian. To account for the complex pattern of polarity responses in these two languages, R&F make two further assumptions. First, the absolute polarity feature of a response may be realized not only by a polarity particle but also by the (possibly truncated) prejacent. This is the strategy used in so-called 'echo' responses (see footnote 2 above).

Second, as already mentioned, markedness considerations are not alone in determining which features are overtly realized in particular languages. Particle choice may also be affected by realization constraints that favor the overt realization of a particular type of feature. Thus, R&F argue that in Romanian, absolute features must always be realized overtly, while in Hungarian only the most marked absolute polarity feature [−] is obligatorily realized.

R&F also briefly discuss two languages in the [REVERSE,+] third particle group, namely French and German. The feature-particle mapping proposed for German is given in (8):

(8) *Feature-particle mapping for German*
   a. *ja* realizes [AGREE] and [+]
   b. *nein* realizes [REVERSE] and [−]
   c. *doch* realizes [REVERSE,+]

Thus, on this account, German is like English in having two multifunctional particles, but, additionally, it has a dedicated [REVERSE,+] particle, *doch*. R&F claim that *doch* blocks both *ja* and *nein* in [REVERSE,+] responses.

The predictions this account makes for German are given in (9):

(9) *Main predictions for German*
   a. In [AGREE,+] responses, only *ja* can be used.
   b. In [REVERSE,−] responses, only *nein* can be used.
   c. In [REVERSE,+] responses, only *doch* can be used.
   d. In [AGREE,−] responses, both *ja* and *nein* can be used
   e. In [AGREE,−] responses, there is a preference for *nein* over *ja* due to markedness.

Summary Whether a polarity feature is realized or not in a particular situation is taken to depend on four factors. The first is whether the language has a means of realizing that feature or

not. The second factor concerns the markedness of the feature relative to the other feature in the response. For instance, in [AGREE,−] responses realizing the marked feature [−] is preferred, other things being equal, over realizing the unmarked feature [AGREE]. The third factor is whether there is a realization constraint favoring the realization of that feature. Finally, blocking is relevant in cases such as [REVERSE,+] responses in German, where the language allows, in principle, a choice between the specific particle *doch* realizing the feature combination as a whole, and two more general particles, *nein* realizing [REVERSE] and *ja* realizing [+].

3 An account of the experimental results of Claus et al. (2017)

Below we review the experimental results of Claus et al. (2017) concerning polarity particles in German, and further refine our earlier proposal for this language so as to capture these results.

Main experimental findings Confirming the prediction in (9d), Claus et al. found that both *ja* and *nein* are possible in confirming responses to negative declaratives (our [AGREE,−] responses). They also found that manipulating the saliency of the negative propositional discourse referent introduced by the preceding declarative in the larger context does not affect the choice of particle in these responses. This result is in line with the feature model but unexpected on the original salience account of Krifka (2013).

When it comes to preference patterns in [AGREE,−] responses, however, Claus et al. found that in contrast to English, *ja* appears to be preferred over *nein*, a result that contradicts prediction (9e). Looking more closely at the data, Claus et al. (2017, §4.2) argue that German speakers in fact fall into two groups, a large group that prefers *ja* over *nein* in [AGREE,−] responses, and a smaller group that prefers *nein* over *ja* in such responses. They refer to the former as the *ja*-group and to the latter as the *nein*-group. The experimental results show that for some participants, the preference for one particle over the other in [AGREE,−] responses was quite pronounced. In Experiment 2 the *ja*-group consisted of 31 participants, 15 of which rated *ja* ≥ 6 and *nein* ≤ 4 on a scale from 1 to 7. However, for other participants the ratings for *ja* and *nein* were quite close, and for some they were identical.

Finally, Claus et al. found that in [REVERSE,+] responses, while *doch* is the favored response *nein* is relatively acceptable as well. As predicted, *ja* was the least acceptable choice. In their Experiment 4, where participants had to rate the acceptability of all three particles in [REVERSE,+] responses with an explicit prejacent, *doch* received the highest median rating (7 out of 7), while the median rating for *nein* was 3.75, and the median rating for *ja* was 1. The prediction in (9c) captures the high and low ratings for *doch* and *ja* respectively but not the fact that *nein* elicits a median rating.

Account of [AGREE,−] responses We develop an Optimality Theoretic account, building on the feature model proposal sketched in Claus et al. (2017, §4.1.1 and §4.2).[6]

To account for the judgments of the speakers in the *ja*-group, who prefer *ja* over *nein* in [AGREE,−] responses, we propose that, in addition to the feature-particle mapping in (8), the grammar of these speakers contains a realization constraint favoring the realization of relative features over absolute features, which outranks the constraint favoring the realization of marked features over that of unmarked ones. In [AGREE,−] responses, this leads speakers in the *ja*-group to prefer *ja* over *nein*.

[6] As mentioned by Claus et al., this account was suggested by us in personal communication in response to Meijer et al. (2015), which already presented some of the experimental results included in Claus et al. (2017).

Turning to the speakers in the *nein* group, who prefer *nein* over *ja* in [AGREE,−] responses, their judgments are accounted for by the analysis in R&F without further modification. In this case, just like in English, there is no operative preference for realizing either relative or absolute features, and therefore particle choice in [AGREE,−] responses is primarily sensitive to markedness, leading to a preference for *nein* over *ja*.

The grammar of the speakers in the *ja* group is similar to that of Romanian: in both there is a constraint favoring the realization of certain features, independent of markedness. The main difference between them is that the constraint targets absolute features in Romanian and relative features in German.

Thus, the two constraints governing particle choice in [AGREE,−] responses in German are the two possibly conflicting faithfulness constraints in (10) and (11):

(10) MAXIMIZE MARKED
Maximize the realization of marked polarity features or feature combinations.

(11) MAXIMIZE RELATIVE
Maximize the realization of relative polarity features.

MAXIMIZE MARKED is rooted in a general pressure for the overt realization of marked features. In [AGREE,−] responses it militates for the realization of the marked feature [−], resulting in a preference for *nein*. On the other hand, MAXIMIZE RELATIVE is a constraint that militates for the realization of relative features, which in this case results in a preference for *ja*. German speakers in the *ja*-group rank MAXIMIZE RELATIVE over MAXIMIZE MARKED, while speakers in the *nein*-group have the opposite ranking. In English and Romanian MAXIMIZE RELATIVE is inoperative, in that it is ranked so low as not to have an effect.

To provide an explicit account for the preferential rather than categorical nature of the judgments reported in these experiments, we adopt the Linear OT framework proposed in Keller (2000). There are other variants of OT that are designed to capture gradient acceptability data, such as Stochastic OT (Boersma and Hayes, 2001) and Harmonic Grammar (Legendre et al., 1990; Pater, 2009). We adopt Linear OT here for concreteness and simplicity, but this choice is not essential for our account (see Keller, 2006, for a concise comparison between Linear OT, Stochastic OT, Harmonic Grammar, and other alternatives).[7] Linear OT differs from classical OT in two ways. First, it represents constraint rankings by means of numerical weights, rather than as partial orders. Secondly, it assumes that the acceptability of a given expression is proportional to the sum of the weights of the constraints that it violates. More explicitly, if C is the set of constraints that a given expression violates, and if the weight of a constraint $c \in C$ is denoted as $w(c)$, then the acceptability of the expression is assumed to be proportional to $-\sum_{c \in C} w(c)$, which we refer to as its *score*.

We assume that German speakers of the *ja*-group assign a higher weight to MAXIMIZE RELATIVE than to MAXIMIZE MARKED, while speakers of the *nein*-group assign a higher weight to MAXIMIZE MARKED. In (12) and (13) below we provide tableaux for speakers of the *ja*-group and the *nein*-group, respectively. For concreteness, we assume that in the former case, MAXIMIZE MARKED has weight 2 and MAXIMIZE RELATIVE has weight 1, and in the latter case these weights are reversed.

[7] We follow Claus et al. (2017) here in assuming that the gradient pattern they found should be captured within a competence model rather than on the basis of performance factors. An anonymous reviewer raises doubts about this assumption. If one gives it up, a comparison between the feature model and the salience approach on this point becomes moot. We maintain it here to show that a competence-based account within the feature model is possible. We are not, however, strongly committed to the claim that these data necessarily have to be captured entirely within a competence model.

(12) Tableau for [AGREE,−] responses in German, *ja*-group:

Weight	MAXIMIZE RELATIVE 2	MAXIMIZE MARKED 1	*Score*
nein	⋆		−2
ja		⋆	−1

(13) Tableau for [AGREE,−] responses in German, *nein*-group:

Weight	MAXIMIZE MARKED 2	MAXIMIZE RELATIVE 1	*Score*
nein		⋆	−1
ja	⋆		−2

Recall that Claus et al. also found that some participants gave *ja* and *nein* the same acceptability rating. Assuming that this finding is not due to task-related reasons, it can be accounted for by assuming that these participants assign equal weight to MAXIMIZE RELATIVE and MAXIMIZE MARKED. Thus, the experimental results of Claus et al. with regard to [AGREE,−] responses can be fully captured within the feature model.

Account of [REVERSE,+] responses In [REVERSE,+] responses *doch* was rated very high, *ja* very low, and *nein* in between, as schematized in (14):

(14) *doch* > *nein* > *ja*

Claus et al. note that neither the R&F analysis of German nor the salience account of Krifka (2013) predicts this pattern. Recall that under the R&F account, the particle *doch*, which realizes the feature combination [REVERSE,+], is taken to block both *nein*, which in [REVERSE,+] responses realizes [REVERSE], and and *ja*, which in these responses realizes [+]. The difference in acceptability between *nein* and *ja* thus remains unexplained.

Claus et al. propose an analysis of the pattern in (14) in a modified version of the salience account, and sketch a possible account in the feature model as well (see p.37 of Claus et al., 2017). Below we show in more detail how the feature model can capture the pattern in (14). We take blocking to be a mechanism that "adjudicates between those outputs which express either all of the input meaning (feature content) or some subpart of it" (Kiparsky 2005; see also Wunderlich 1995, Rainer 2016, among others). If what needs to be expressed is the feature combination [REVERSE,+], *doch* is the optimal choice because it expresses all feature content, while *ja* and *nein* only express part of it. We implement this by assuming the faithfulness constraint in (15):

(15) EXPRESSIVENESS
Maximize the expression of feature content.

The use of *doch* in [REVERSE,+] responses obeys EXPRESSIVENESS, while that of *nein* and *ja* violates it. In classical OT this would result in *doch* being optimal and therefore grammatical in these responses, and *nein* and *ja* being non-optimal, and therefore both unacceptable. However, if we adopt a version of OT that allows us to capture gradient acceptability patterns, such as Linear OT, the difference in acceptability between *nein* and *ja* can be captured.[8] More specifically, when cast

[8]The revised salience account of [REVERSE,+] responses given in Claus et al. (2017) employs essentially the same

in Linear OT, the feature model account makes the right predictions, independently of the relative weight assigned to the relevant constraints, because *doch* in a [REVERSE,+] response violates no relevant constraints, *nein* in such a response violates EXPRESSIVENESS but obeys MAXIMIZE MARKED and MAXIMIZE RELATIVE, while *ja* violates all three relevant constraints. We summarize this in the tableau below, where we assume for concreteness that all constraints have the same weight—any other assignment of weights would predict the same order of acceptability in [REVERSE,+] responses.

(16) Tableau for [REVERSE,+] responses in German:

	EXPRESSIVENESS	MAXIMIZE MARKED	MAXIMIZE RELATIVE	
Weight	1	1	1	*Score*
doch				0
nein	⋆			−1
ja	⋆	⋆	⋆	−3

Thus, the account predicts the pattern in (14): *doch* is fully acceptable in [REVERSE,+] responses, *ja* is unacceptable, and *nein* is somewhere in between these two extremes.

This completes our account of the experimental data in Claus et al. (2017) concerning polarity particle usage in German. We now turn to a comparison of the feature model with the salience approach.

4 Comparison between the salience approach and the feature model

We first identify some basic differences between the two approaches with reference to English and German (Section 4.1), and then take a wider cross-linguistic perspective (Sections 4.2-4.4).

4.1 Basic differences between the two approaches

On the salience approach, polarity particles are propositional anaphors that affirm or reject their antecedent discourse referent. Just like in the feature model, propositional discourse referents are taken to be marked as being either positive or negative. Positive sentences introduce positive discourse referents. In responses to such sentences, *yes* and *ja* pick up the positive discourse referent and affirm it, while *no* and *nein* pick up the same discourse referent and reject it.

An essential difference between the salience approach and the feature model concerns polarity responses to negative initiatives. On the salience approach, polarity responses to such initiatives have access to two discourse referents, a negative discourse referent, $\overline{p}_{DR}$, introduced by the whole negative sentence, and a positive discourse referent, p_{DR}, introduced by the lower TP of the negative sentence. The overlap between *yes* and *no* in responses to negative initiatives in English (and similarly between *ja* and *nein* in German) is due, on this approach, to the availability of these two discourse referents as antecedents to polarity particles. Differences in particle preferences in such responses are taken to follow from the assumption that languages (or dialects) differ with respect to the relative salience assigned to these two discourse referents. Thus, for the *nein*-group in German, as well as for English, p_{DR} is assumed to be more salient than $\overline{p}_{DR}$, while for the *ja*-group, the relative salience of the two discourse referents is assumed to be reversed. (This assumption is not made in the original salience account of Krifka 2013, but is proposed by Claus et al. 2017 as a possible way to account for the novel experimental data within the salience approach).

strategy. On this account, both *nein* and *ja* are non-optimal but *nein* violates a lower-ranked constraint than *ja*. In Linear OT, this yields the prediction that *nein* is more acceptable than *ja*.

Under these assumptions, *yes/ja* are treated as always agreeing with the antecedent, while *no/nein* are treated as always rejecting it. Transposing this account in terms of the feature model, if both discourse referents introduced by negative sentences are available as antecedents to polarity particles, *yes/ja* and *no/nein* can be treated as realizing only relative features. The overlap in their distribution and interpretation in responses to negative initiatives is due solely to the availability of the two discourse referents that such initiatives introduce, while preferences are explained in terms of intrinsic salience differences between the two discourse referents.

In contrast, in the feature model it is assumed that particles in responses to negative initiatives can only access the negative discourse referent. The overlap in particle use in such responses is due to properties of the polarity particles themselves: English and German polarity particles are treated as being able to express both absolute and relative features.

A possible argument in favor of the salience approach runs as follows: by exploiting the presence of the two discourse referents introduced by negative initiatives, a simpler account of polarity particle responses in English is possible, since sensitivity to the absolute parameter can be dispensed with. This point is made explicitly in Goodhue and Wagner (2018).

We will argue, however, that this gain in simplicity disappears once we take further cross-linguistic data into account. More specifically, we will point out that there are certain mismatches between polarity particles and other propositional anaphors which are unexpected on the salience approach (§4.2), and that sensitivity to the absolute parameter is a necessary component of a theory of polarity particles cross-linguistically (§4.3). We end with a contrast in typological predictions made by the two approaches (§4.4).

4.2 Mismatches between polarity particles and other propositional anaphors

Essential to the salience approach is the assumption that both propositional discourse referents introduced by a negative initiative, p_{DR} and $\overline{p}_{DR}$, can in principle serve as antecedents to polarity particles. To motivate this assumption, Krifka (2013, p.6) considers examples like (17), which show that propositional anaphors like *this*, *that*, and *it* can target the positive discourse referent introduced by a negative sentence.

(17) The Incas didn't reach Tahiti, even though Heyerdahl claimed this. (Krifka, 2013)

The demonstrative *this* takes as its antecedent the discourse referent introduced by the TP *the Incas reached Tahiti*.

In order to rely on this fact in accounting for polarity particle responses to negative initiatives in the way the salience account does, one has to further assume that the anaphoric reach of polarity particles is similar to that of propositional anaphors such as *this*, *that*, and *it*. This assumption, upon closer inspection, appears to be problematic: the anaphoric potential of polarity particles does not always parallel that of other propositional anaphors.

A language in which this can be seen is Japanese, a two polarity particle language, where *hai* affirms the antecedent and *iie* reverses it, without regard to the sentence polarity of the antecedent or that of the response. In the feature model, Japanese polarity particles realize only relative features. In the salience approach, negative sentences in Japanese must be assumed to introduce a single propositional discourse referent corresponding to the whole negated sentence. Krifka (2013, §4.4) notes that this follows from the assumption that Japanese negation is always predicate negation rather than sentential negation.

Note, however, that under this assumption the salience account predicts that in Japanese, unlike in English, other propositional anaphors would also be unable to pick up the positive discourse

referent introduced by a negative sentence. This prediction appears to be problematic, as shown in (18):[9]

(18) Heierudaaru-wa soo syutyoo-siteita keredomo, (zissai) inka-zoku-wa tahiti-ni
Heyerdahl-TOP so claim-DO.PROG.PAST even though (in fact) the Incas-TOP Tahiti-to
tootatu-sinakatta.
reach-do.NEG.PAST
'Even though Heyerdahl said so, the Incas didn't reach Tahiti.'

The interpretation of this Japanese example is parallel to that of (17). Therefore, in Japanese, just like in English, negative sentences must be assumed to make a positive discourse referent available for propositional anaphora. This poses a challenge for the salience account but not for the feature model. To address this challenge one could assume that negative sentences in Japanese introduce a positive discourse referent which is inaccessible to polarity particles but can be accessed by other propositional anaphors. This assumption would remain, however, in need of independent justification.

In the feature model, on the other hand, all one has to say is that the negative discourse referent introduced by negative sentences is universally more salient than the positive one, to the extent that the positive discourse referent is never available as an antecedent for polarity particles.[10]

4.3 Sensitivity to sentence polarity

In the feature model, response particles are sensitive to both the relative and the absolute parameter. With respect to the latter, recall that absolute polarity features encode sensitivity to the sentence polarity of the prejacent. In interaction with relative polarity features, they indirectly encode sensitivity to the sentence polarity of the antecedent as well. Thus, in [AGREE] responses, the sentence polarity of the prejacent must be identical to that of the antecedent; in [REVERSE] responses, the sentence polarity of the prejacent must be the opposite of the sentence polarity of the antecedent.

In the salience account response particles themselves encode only information about the relative parameter: they either confirm or reject their antecedent. The overall system thus appears simpler. However, sensitivity to sentence polarity is introduced via presuppositions that may be associated with particular particles. For instance, *doch* is taken to presuppose a *negative* antecedent and rejects it. Once presuppositions making reference to sentence polarity are introduced in the salience account, the contrast between the two approaches is weakened since these presuppositions come close to mimicking the role of absolute polarity features in the feature model. Recall that in the feature model, *doch* signals that the response is positive, and that it reverses the antecedent,

[9] The reported judgment of this sentence, as well as the judgments reported below for Romanian, have been confirmed by two native speakers. Note that the embedded clause in (18) appears before the matrix clause, which means that we have a case of propositional cataphora rather than anaphora here. While this is the unmarked word order in Japanese, the opposite order is possible as well, resulting in a proper case of anaphora, without concomitant changes in interpretation. Instead of *soo*, it is also possible to use *aa* 'that' with the same interpretation.

[10] Note that this is compatible with the assumption that propositional discourse referents introduced by clauses embedded under certain attitude verbs (rather than negation) may, in certain contexts, serve as antecedents for polarity particles. Goodhue and Wagner (2018) illustrate that this is indeed possible by means of the following example:

(i) *Context: A finds B and C arguing about whether John is home, and decides to add her two cents.*
A: I know Mary believes John is home.
C: No, he isn't. / No, she doesn't. / No, you don't.

both in terms of content and in terms of polarity. From this it follows that the antecedent must be negative. The two analyses then are distinguished only in that on the salience account *doch* exclusively encodes information about the sentence polarity of the antecedent, while in the feature model it directly encodes information about the sentence polarity of the prejacent, and, indirectly, information about the sentence polarity of the antecedent.

We argue below that polarity particles in general need to be able to encode information about the sentence polarity of their prejacent. The argument is based on the polarity particle system of Romanian, discussed in detail in Farkas (2011) and Roelofsen and Farkas (2015).[11] Romanian is the opposite of Japanese in that in responses to negative initiatives, the negative particle *nu* is used in agreeing responses and the positive particle *da* is used in rejecting responses. A third particle, *ba*, occurs only in rejecting responses. We exemplify in (19):

(19) A: Petru nu a telefonat. 'Peter did not call.'
B: Nu, (nu a telefonat). 'No, (he) did not call.'
B: Ba da, (a telefonat). 'Yes, he DID.'

Going into further details, one finds that all responses must contain either one of the 'absolute particles' (*da* or *nu*) or an echo, i.e., a truncated form of the prejacent that contains the verb, which overtly marks the sentence polarity of the response. Furthermore, the absolute particle in the response must always agree with the sentence polarity of the prejacent. This is illustrated in (20) and (21):

(20) A: Petru a telefonat. 'Peter called.'
B: Ba nu, (nu a telefonat). / Ba, nu a telefonat / *Ba da, nu a telefonat. 'No he didn't.'

(21) A: Petru nu a telefonat. 'Peter did not call.'
B: Ba da, (a telefonat.) / Ba, a telefonat / *Ba nu, a telefonat.' 'Yes, he DID.'

The account of these facts in the feature model takes *da* and *nu* to realize the features [+] and [−] respectively, while the third particle, *ba*, realizes the feature [REVERSE]. In addition, in Romanian MAXIMIZE ABSOLUTE is highly ranked and therefore the absolute feature of the response must always be realized (by a particle or an 'echo').

The fact that the positive particle occurs in [REVERSE,+] responses, and the negative particle occurs in [REVERSE,−] responses indicates that these particles are sensitive to the sentence polarity of the *prejacent* rather than that of the antecedent. The fact that an 'echo' response obviates the necessity of an absolute polarity particle can naturally be explained under the assumption that the absolute polarity feature of a response has to be overtly realized in Romanian, and that both the use of an absolute polarity particle and that of an 'echo' are ways of satisfying this constraint. These facts support the assumption that polarity particles have to be allowed to encode information about the sentence polarity of their prejacent.

We turn now to the issue of how these facts could be captured in the salience account. The null hypothesis would be that *da* and *nu* are the equivalents of English *yes* and *no*, and German *ja* and *nein*, respectively, i.e., *da* affirms a salient propositional discourse referent, and *nu* rejects it. To account for the use of these two particles in responses to negative initiatives without making reference to sentence polarity, one would have to claim that in languages like Romanian negative initiatives introduce two discourse referents, just as in English and German, and that in these languages, p_{DR} is more salient than $\overline{p}_{DR}$.[12] This is so because in these languages only the positive

[11] The argument given here was already hinted at in footnote 19 of Roelofsen and Farkas (2015). For detailed discussion of the polarity particle system of Hungarian, which supports the same argument, see Farkas (2009).

[12] Krifka (2013) assumes that p_{DR} is more salient than $\overline{p}_{DR}$ in English and German as well, but Claus et al. (2017)

particle is used in [REVERSE,+] responses, and only the negative one in [AGREE,−] responses.

An immediate prediction this account would make is that other propositional anaphors in Romanian are sensitive to this salience difference as well. A contrast between Romanian and English is predicted with respect to examples such as (17), whose Romanian equivalent is given in (22).

(22) Incaşii nu au ajuns până la Tahiti, cu toate că Heyerdahl a afirmat asta.
Inca.Pl not have reached till at Tahiti with all that Heyerdahl has affirmed this
'The Incas did not reach Tahiti even though Heyerdahl claimed this.'

The prediction is that in Romanian, the propositional anaphor *asta* 'this' accesses the positive discourse referent introduced by the preceding negative sentence more readily than in English. This prediction awaits experimental support.

In (23) we give an example where *asta* can easily access the negative discourse referent:

(23) A: Petru nu a udat florile. 'Peter didn't water the plants.'
B: Maria i-a spus asta lui Ion? 'Did Maria tell this to Ion?'

The prediction this account makes is that interpreting *asta* as referring to Peter not having watered the plants is harder in Romanian than in the corresponding English sentence. More generally, the salience approach relies on the assumption that languages may differ with respect to the relative salience of the two discourse referents introduced by negative sentences, while the feature model does not. In the absence of experimental results providing independent support for this assumption, it remains stipulative.

Now, in order to account for the reversing nature of the particle *ba*, one would have to assume that this particle picks up a discourse referent and reverses it, and that in [REVERSE,+] responses the discourse referent it picks up is the non-salient negative discourse referent, while in [REVERSE,−] responses, it picks up the salient positive discourse referent. This is not a particularly attractive account. Alternatively, one could posit that *ba* simply marks a response as rejecting its antecedent, just like in the feature model but then the problem of differentiating *nu* and *ba* arises, since both would be rejecting particles under this account.

If one allows polarity particles to encode information about the sentence polarity of their antecedents in the salience model, a much simpler account becomes available. The particles *da* and *nu* could then be assumed to presuppose a positive discourse referent, and *da* would accept its antecedent, while *nu* would reject it. In the case of negative sentences, one could maintain the simplest assumption, namely that the negative discourse referent they introduce is more salient than the positive one. The particle *ba* then can be treated as picking up the most salient discourse referent and rejecting it. In a *ba da* response the first particle picks up the negative discourse referent and rejects it, while the second picks up the positive discourse referent and affirms it; in a *ba nu* response, both particles pick up the same positive antecedent and reject it. Something more would have to be said to rule out a simple *da* in [REVERSE,+] responses.

Note that the gain in simplicity just achieved is due precisely to the fact that particles are allowed to involve presuppositions concerning the sentence polarity of their antecedent. As a result, the relevant presuppositions come close to mimicking the role of the absolute polarity features in the feature model. These data then, as well as the existence of [REVERSE,+] particles like *doch*, show that sensitivity to sentence polarity is necessary for an account of polarity particles across languages.

assume the opposite for German.

We now return more specifically to the necessity of taking the sentence polarity of the *prejacent* into account. Recall the observation that in Romanian, polarity particle responses must always involve one of the absolute polarity particles *da* and *nu*, or an echo, or both. In the feature model both absolute polarity particles and echoes are taken to realize the absolute polarity feature of the response. Thus, the interaction between particles and echoes in Romanain can be accounted for by the requirement that the absolute polarity feature of a polarity particle response must be realized in this language.

On the other hand, it is unclear how this interaction could be captured in the salience model, where polarity particles do not encode information about the sentence polarity of the prejacent. Here, then, we have evidence that a theory of polarity particle responses that has cross-linguistic coverage must not only assume that such particles are sensitive to sentence polarity, but, more specifically, that they potentially encode information about the sentence polarity of their prejacent.

In sum, both the salience account and the feature model have to make reference to sentence polarity to account for data beyond English. The difference is that in the feature model, polarity features encode information about the sentence polarity of the prejacent as well as the antecedent, while in the salience account polarity particles only encode information about the sentence polarity of the antecedent. While this difference is subtle, taking the former view appears to have two advantages: (i) it allows a simpler account of systems like those found in Romanian, and (ii) it is better equipped to capture interactions between particles and echoes.

4.4 Further typological predictions

We now compare the two approaches with respect to languages that have three polarity particle systems. In the feature model, the existence of three particle systems like Romanian, French or German is expected. In Romanian, the third particle expresses [REVERSE], while in French and German, it expresses the feature combination [REVERSE,+]. If only one relative feature is expressed in a language, one expects it to be [REVERSE] rather than [AGREE]. In French and German, the third particle expresses the most marked feature combination, namely [REVERSE,+]. Having a designated particle that expresses this combination is not surprising. R&F argue that the second most marked feature combination, after [REVERSE,+], is [AGREE,−]. It is expected, then, that there are also languages with a dedicated particle for this feature combination. An example of such a language is the Ethiopian language Soddo, described in Leslau (1962). Besides a dedicated particle for [AGREE,−], this language has two particles which can be analyzed as realizing [AGREE] and [REVERSE], respectively. The former is blocked in [AGREE,−] responses by the particle that realizes the feature combination.

Under the salience account, a very natural polarity particle system is one with the two particles schematized in (24):

(24) a. p_1: affirms the most salient antecedent
b. p_2: rejects the most salient antecedent

Under the assumption that negative sentences introduce two discourse referents, and that there is an intrinsic salience difference between them, such a system is stable, and an additional particle is superfluous. If the two discourse referents were equally salient, however, reactions to negative sentences would be ambiguous. One could thus explain the three particle system of German as a way of getting around this situation. The use of *doch*, which on the salience account presupposes a negative antecedent and reverses it, marks the rejection of a negative antecedent in an unambiguous fashion.

Thus, under the salience account one could explain the existence of three-particle systems like that found in German in functional terms, arising under pressure to reduce possible ambiguity in case the positive and negative discourse referents introduced by negative sentences are equally salient. Note, however, that the actual account of German proposed in Claus et al. (2017) does not assume that the two discourse referents are equally salient. A difference in salience is needed in order to account for particle use in [AGREE,−] responses. Note also that this functional explanation cannot hold for the French particle system. In terms of the feature model, in French, the particle *oui* realizes [+] and the particle *non* realizes [−] (parallel to Romanian *da* and *nu*, respectively). In response to a negative initiative, *non* can only be interpreted as an [AGREE,−] response, and therefore *oui* would be unambiguously interpreted as a [REVERSE,+] response. Thus, from the perspective of the salience account, there is no functional need for an additional particle to be used in responses to negative initiatives. Nonetheless, French, just like German, has a dedicated [REVERSE,+] particle, *si*. The same point can be made based on the polarity particle system of Tigrinya, a language spoken in Ethiopia, which according to the description in Leslau (1962) is parallel to that of French. We conclude that the feature model is in a better position than the salience account to capture the properties of some attested polarity particle systems.

Conversely, note that the feature model differs from the salience account with respect to predictions concerning non-attested particle systems. To exemplify, consider the fictitious system described in (25):

(25) A fictitious particle system
  a. p_1 realizes [−]
  b. p_2 realizes [REVERSE]
  c. p_3 realizes [AGREE,+]

This system has a particle p_1 which realizes [−], just like French *non* and Romanian *nu*, a particle p_2 which realizes [REVERSE], just like Japanese *iie* and Romanian *ba*, and a particle p_3 which realizes the feature combination [AGREE,+]. If we assume that in this language the negative discourse referent introduced by a negative sentence is more salient than the positive one (as in the majority dialect of German according to Claus et al., 2017), then the system can be described in terms of the salience approach as follows.

(26)
  a. p_1 presupposes a positive antecedent and rejects it
  b. p_2 rejects the most salient antecedent
  c. p_3 presupposes a unique antecedent and confirms it

Note that from the perspective of the salience approach, p_3 is the mirror image of German *doch*: it presupposes a unique antecedent (which means that it can only be used in response to a positive sentence) and confirms it, while *doch* presupposes a negative antecedent (which means that it can only be used in response to a negative sentence) and reverses it.

In a language with this system, one would use p_3 to agree with a positive initiative, while to reject such an initiative, one could use either p_1 or p_2. To agree with a negative initiative, one would use p_1, while to reject a negative initiative one would use p_2.

Such a system is unobjectionable under the salience approach. In particular, there is no redundancy in the system: no strict subset of the three particles would be sufficient to cover all possible response types. Moreover, from the perspective of the salience approach, the presence of a dedicated particle for [AGREE,+] responses is just as natural as the presence of a dedicated particle for [REVERSE,+] responses in languages like German and French.

On the other hand, the particle system in (25) is highly unexpected from the perspective of the

feature model, precisely because it has a dedicated particle to realize [AGREE,+], which is the least marked feature combination of all. The feature model predicts that the existence of such a particle would be highly unusual, while systems with dedicated [REVERSE,+] particles are expected to be frequently found. Given the cross-linguistic data collected so far, this prediction is borne out.

5 Conclusion

The first aim of this paper was to account for the experimental findings on the interpretation of polarity particles in German reported in Claus et al. (2017) within the feature model of polarity particles developed in Roelofsen and Farkas (2015). The second aim was to compare the feature model with the salience approach proposed by Krifka (2013) and further refined in Claus et al. (2017). The core difference between the two approaches concerns the way responses to negative initiatives are treated. We argued that the assumption that polarity particles encode information about the sentence polarity of the prejacent as well as the antecedent (as in the feature model) rather than just that of the antecedent (as in the salience account) has advantages in accounting for the interplay between polarity particles and echo responses.

The two approaches further contrast in that the salience approach has to assume cross-linguistic differences in the salience of the two discourse referents introduced by negative sentences while the feature model does not. The differences stipulated on the salience approach make testable empirical predictions concerning the parallelism between polarity particle responses and the preferred anaphora resolution of propositional anaphors after negative sentences, which, if confirmed, would provide independent justification for this account. A third difference between the two approaches concerns the typological predictions they make concerning expected and unexpected polarity particle systems.

Work of the kind carried out by Claus et al. (2017), probing the details of preference patterns in the use and interpretation of polarity particles, is crucial in assessing and further refining theories of polarity particle systems across languages.

References

P. Boersma and B. Hayes. Empirical tests of the gradual learning algorithm. *Linguistic inquiry*, 32(1):45–86, 2001.

A. Brasoveanu, D. F. Farkas, and F. Roelofsen. N-words and sentential negation: Evidence from polarity particles and VP ellipsis. *Semantics and Pragmatics*, 6(7):1–33, 2013. doi: 10.3765/sp.6.7.

B. Claus, M. Meijer, S. Repp, and M. Krifka. Puzzling response particles: An experimental study on the German answering system. *Semantics and Pragmatics*, 10, 2017.

D. F. Farkas. Polarity particles in Hungarian. In *Approaches to Hungarian, volume 11: Papers from the 2007 New York Conference*, pages 95–118. John Benjamins Publishing Company, 2009.

D. F. Farkas. Polarity particles in English and Romanian. In J. Herschensohn, editor, *Romance Linguistics 2010. Selected papers from the 40th Linguistic Symposium on Romance Linguistics*, pages 303–328. John Benjamins, 2011.

D. F. Farkas and K. B. Bruce. On reacting to assertions and polar questions. *Journal of semantics*, 27:81–118, 2010.

D. Goodhue and M. Wagner. Intonation, yes and no. *Glossa: a journal of general linguistics*, 3(1), 2018.

A. Holmberg. The syntax of answers to polar questions in English and Swedish. *Lingua*, 128:31–50, 2013.

A. Holmberg. *The syntax of yes and no*. Oxford University Press, 2016.

L. R. Horn. Towards a new taxonomy of pragmatic inference: Q-based and R-based implicatures. In D. Schiffrin, editor, *Meaning, Form, and Use in Context*, pages 11–42. Georgetown University Press, 1984.

B. M. Jones. *The Welsh Answering System*. De Gruyter, 1999.

F. Keller. *Gradience in grammar: Experimental and computational aspects of degrees of grammaticality*. PhD thesis, University of Edinburgh, Edinburgh, 2000.

F. Keller. Linear Optimality Theory as a model of gradience in grammar. *Gradience in grammar: Generative perspectives*, pages 270–287, 2006.

P. Kiparsky. Blocking and periphrasis in inflectional paradigms. In G. Booij and J. van Marle, editors, *Yearbook of Morphology 2004*, pages 113–135. Springer, 2005.

R. Kramer and K. Rawlins. An ellipsis approach to answer particles in positive and negative contexts. Presentation at a Workshop on the Syntax of Answers to Polar Questions at Newcastle University, 2012.

M. Krifka. Response particles as propositional anaphors. In *Proceedings of Semantics and Linguistic Theory (SALT 23)*, pages 1–18, 2013.

S. Kuno. *The Structure of the Japanese Language*. MIT Press, 1973.

G. Legendre, Y. Miyata, and P. Smolensky. Harmonic grammar: A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. University of Colorado, Boulder, Department of Computer Science, 1990.

W. Leslau. Yes and no in the Ethiopian languages. *Language*, 38(2):147–148, 1962.

M. Meijer, B. Claus, S. Repp, and M. Krifka. Particle responses to negated assertions: Preference patterns for German ja and nein. In *Proceedings of the 20th Amsterdam Colloquium*, 2015.

J. Pater. Weighted constraints in generative linguistics. *Cognitive science*, 33(6):999–1035, 2009.

E. Pope. *Questions and Answers in English*. Mouton, The Hague, 1976.

F. Rainer. Blocking. In M. Aronoff, editor, *Oxford Research Encyclopedia of Linguistics*, pages 1–22. Oxford University Press, 2016.

F. Roelofsen and D. F. Farkas. Polarity particle responses as a window onto the interpretation of questions and assertions. *Language*, 91(2):359–414, 2015.

J. Sadock and A. Zwicky. Speech act distinctions in syntax. In T. Shopen, editor, *Language Typology and Syntactic Description*, pages 155–196. Cambridge University Press, 1985.

D. Wunderlich. Minimalist morphology: the role of paradigms. In G. Booij and J. van Marle, editors, *Yearbook of Morphology*, pages 93–114. Kluwer, Dordrecht, 1995.