

# Polarity particles revisited\*

Donka F. Farkas and Floris Roelofsen

Wordcount: 8089

August 21, 2018

## 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 three approaches to such responses, namely the *saliency account* in Krifka (2013), the *feature model* in Roelofsen and Farkas (2015), and the *ellipsis approaches* in Kramer and Rawlins (2012), Holmberg (2013), and Holmberg (2016). The authors then sketch a way to account for the data within the feature model, as well as a revised version of the saliency 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 saliency 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 particular analyses and general typological predictions. In Section 3 we turn to the findings in Claus et al. (2017) and develop an account of them within the feature model. To capture these data, we refine our earlier account of German while leaving the overall framework intact. After summarizing the revised version of the saliency account presented in Claus et al. (2017), Section 4 compares it with the feature model from a wider cross-linguistic perspective, and Section 5 concludes.

## 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 antecedents, 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 with or rejects the antecedent. The absolute parameter concerns the polarity of the response itself.

---

\*Acknowledgments to be added after the reviewing process.

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.<sup>1</sup>

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 propose implements this idea by proposing that polarity particles in all languages realize two types of polarity features, *absolute* polarity features and *relative* polarity features, which connect to the absolute and the relative parameter, respectively.

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

These features are taken to be universal in that they characterize polarity responses across languages. Polarity responses are taken to involve a polarity head that hosts an absolute and a relative polarity feature, and a prejacent clause, which may be fully or partially elided.

The semantic contribution of the polarity features is as follows (see R&F for formal details).

- (3) *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.
- (4) *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.<sup>2</sup>  
 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.

Given their presuppositions, 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) that can be realized by that particle. In these terms, Japanese has mapping rules connecting particles to relative features but no rules that connect particles to absolute ones. In principle, it is possible to have the reverse situation, i.e., a language that has rules connecting particles to absolute features but no particles connected to relative ones.

<sup>1</sup>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.

<sup>2</sup>R&F assume that sentences introduce propositional discourse referents, and that these discourse referents are marked for polarity, i.e., they are marked as − if the clause that contributes them is negative, and + otherwise.

Second, as discussed below, the form of polarity responses may be affected by language specific *realization rules*, which specify that certain features must be realized obligatorily, or that there is a preference for overtly realizing one type of feature over another.<sup>3</sup>

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, i.e., 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 just two polarity particles. For instance, a language may have two particles realizing the two absolute features, and in addition one or two 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 actual form of polarity responses in a particular language, as well as the possible typological space of such systems is delimited by a series of markedness distinctions concerning polarity features. To exemplify, markedness considerations constrain feature-particle mappings in the case of a multi-functional particle, i.e., a particle that may realize either an absolute or a relative feature. Given that [+] and [AGREE] are the unmarked absolute and relative feature, respectively, and [-] and [REVERSE] the marked ones, in cases of multifunctionality [+] and [AGREE] will be connected, by harmonic alignment, to one particle, and [-] and [REVERSE] to another. What is unexpected is a language with one particle that can realize [AGREE] and [-], and another particle that can realize [REVERSE] and [+].

**The English polarity particle system** In R&F, English is taken to exemplify a ‘mixed’ two particle system with the feature-particle mapping in (5):

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

This account correctly predicts that only *yes* can be used in [AGREE,+] responses, and only *no* can be used in [REVERSE,-] ones. 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.<sup>4</sup>

- (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 two particles is allowed in [AGREE,-] responses, each realizing one of the two features carried by the response, markedness considerations may be used to explain preference patterns. Other things being 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. Thus, given that in

<sup>3</sup>A terminological note: in Roelofsen and Farkas (2015) we used the term ‘realization rule’ as a general cover term for rules specifying the feature-particle mapping in a given language and rules specifying that certain features must be realized obligatorily, or preferences for realizing one type of feature over another. Here, we use the term only for the latter type of rule, and refer to the former as ‘feature-particle mappings’.

<sup>4</sup>For the role of intonation in [REVERSE,+] responses in English, see Goodhue and Wagner (2018).

[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, 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 already 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.

**Languages with three particles** 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.<sup>5</sup>

The two languages in the [REVERSE] third particle group for which R&F propose a detailed account are Hungarian and Romanian. To account for the complex pattern of polarity responses in these two languages, R&F make two further assumptions. First, it is proposed that 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 1).

Second, as already mentioned, it is proposed that universal markedness considerations are not alone in determining which features are overtly realized in particular languages. Languages may also involve language specific realization rules which require the realization of a particular type of feature in all polarity responses, or a preference for realizing one type of feature over another. 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. [AGREE] and [+] can be realized by *ja*
  - b. [REVERSE] and [−] can be realized by *nein*
  - c. [REVERSE,+] can be realized by *doch*

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 main predictions of this account 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 the latter case, there is a preference for *nein* over *ja*.

There is one further aspect of the feature model that is relevant to the discussion below. Namely, R&F assume that speakers generally tend to *avoid pernicious ambiguity* in polarity particle responses, in line with Grice’s maxim of Manner. That is, speakers will tend to prefer response forms

<sup>5</sup>For details of the special nature of [REVERSE,+] combination, see F&R.

that allow the interlocutor to unambiguously recover the intended meaning. This is particularly relevant to bare particle responses where the whole communicative burden is borne by the particle.

**Summary** Whether a feature is realized or not in a particular situation depends on five factors. The first is whether the language has a means of realizing that feature or not. Recall that Japanese has no rules mapping absolute features to a particle, and therefore absolute features will not be realized by particles. The second factor is whether there is a realization rule requiring that feature to be overtly realized. Recall that Romanian requires the overt marking of absolute features in all responses. The third 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]. Fourth, 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 [+]. Finally, the form of polarity particle responses is further constrained by a general pressure against pernicious ambiguity.

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

Below we review the novel experimental results in Claus et al. (2017) that probe the details of polarity particle usage in German, and then extend our earlier account of this language so as to overcome the challenges they pose.

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 further 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 is unexpected in the saliency 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. argue that German speakers in fact fall into two groups, a large *ja*-group, which shows a preference for *ja* over *nein* in [AGREE,−] responses, and a smaller *nein*-group which prefers *nein* over *ja* in such responses.

Finally, Claus et al. found that in [REVERSE,+] responses *doch* is the favored response but *nein* is judged relatively acceptable as well, while *ja* is the least acceptable. 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), the median rating for *nein* was between 3.5 and 4, while the median rating for *ja* was 1. The prediction in (9c) does not capture this gradient pattern.

**Account of [AGREE,−] responses** The starting point of the account to be worked out below is the proposal sketched in Claus et al. (2017, §4.1.1).<sup>6</sup>

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 rules in (8), the grammar of these speakers

---

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

contains a realization rule favoring the realization of relative features, which outranks the preference for realizing marked features over unmarked ones.

- (10) *Realization rule for ja-group speakers*  
 Prefer the realization of relative features over absolute features.

In [AGREE,–] responses then, the pressure to realize relative features will lead speakers in the *ja*-group to prefer *ja* over *nein*.

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 the case of English, there is no operative preference for realizing one feature over the other, and therefore particle choice in [AGREE,–] responses is sensitive to markedness alone, 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. There are two differences between them, however. First, the relevant constraint targets absolute features in Romanian and relative features in German; second, the constraint in Romanian is categorical while in German it manifests itself as a preference. Neither difference is unexpected. If the theory has to include a realization rule targeting the realization of absolute features in one language, there is no reason to assume that a similar rule cannot target relative features in another language. Second, categorical requirements of the type we have in Romanian may well start out as preferences that get strengthened over time.

In Optimality Theoretic terms, particle choice in [AGREE,–] responses in German is subject to two conflicting faithfulness constraints:

- (11) REALIZE MARKED  
 Realize marked polarity features or feature combinations.
- (12) REALIZE RELATIVE  
 Realize relative polarity features.

REALIZE MARKED is rooted in a universal 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, REALIZE RELATIVE militates for the realization of relative features, resulting in a preference for *ja*. German speakers in the *ja*-group rank REALIZE RELATIVE over REALIZE MARKED, while speakers in the *nein* group have the opposite ranking. In English and Romanian REALIZE RELATIVE is inoperative, i.e., ranked so low as not to have an effect.

To account for the preferential rather than categorical judgments, we assume that acceptability ratings are affected not only by the subject's own grammar, but also by her awareness of the behavior of others in her language community. More specifically, we assume that the level of acceptability assigned by a subject to the use of a certain polarity particle in a certain configuration is determined by two factors, namely (i) whether this use of the particle is licensed by her own grammar, and (ii) the likelihood that the subject assigns to the use of the particle in the given configuration by other speakers in her language community. Thus, a subject from the *ja*-group will rate *ja* higher than *nein* in [AGREE,–] responses, but will not find *nein* completely unacceptable as long as there are some speakers from the *nein*-group in her language community. And similarly, a subject from the *nein*-group will rate *nein* higher than *ja* but will not find *ja* completely unacceptable if her language community is heterogeneous.

**Account of [REVERSE,+] responses** Recall that the experimental evidence in Claus et al. (2017) shows that in [REVERSE,+] responses in German, speakers show a marked preference for *doch* but find *nein* quite acceptable as well, while *ja* in such responses is rated very low.

Recall also that in R&F, *doch*, the particle that realizes the feature combination [REVERSE,+], was claimed to block *nein*, which in these responses realizes [REVERSE], and *ja*, which in these responses realizes [+]. Without further qualifications this account does not predict the gradient pattern found by Claus et al. (2017). However, we show below that under a more sophisticated conception of blocking, motivated elsewhere on independent grounds, this pattern is not problematic for the feature model.

Blocking is generally seen as a mechanism that “adjudicates between those outputs which express either all of the input meaning (feature content) or some subpart of it” (Kiparsky, 2005). If the input in question 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.

However, blocking is not always absolute—its force depends on a variety of factors. Especially relevant for us is the hypothesis that if an expression  $e_1$  blocks an expression  $e_2$ , the more frequent  $e_1$  is relative to  $e_2$ , the stronger its blocking force will be (Rainer, 1989, 2016, among others). In our case, both *ja* and *nein* are more frequent than *doch* in polarity responses. This is expected to temper the blocking force of *doch*.

Thus, the blocking constraints relevant here are given in (13) and (14):

(13) EXPRESSIVENESS  
Express feature content as much as possible.

(14) FREQUENCY  
Prefer the use of frequent forms.

The feature model account of [REVERSE,+] responses in German goes as follows. Given the realization rules of German, all three polarity particles may in principle be used in [REVERSE,+] responses: *doch* realizes both features, *nein* realizes [REVERSE], and *ja* realizes [+]. Assuming that for all speakers of German, EXPRESSIVENESS is ranked above all the other relevant constraints, the preference for *doch* over the other two particles is explained. To account for the median rating of *nein*, a Stochastic OT framework can be adopted, in which constraints are ranked along a continuous scale and in which the relative ranking of constraints that are close to one another can be perturbed (Boersma and Hayes, 2001). We assume in particular that this holds for our two blocking constraints, EXPRESSIVENESS and FREQUENCY. Note that *nein* obeys REALIZE MARKED in a [REVERSE,+] response, while *ja* does not. Thus, if the relative ranking between EXPRESSIVENESS and FREQUENCY is perturbed, *nein* becomes optimal while *ja* remains non-optimal. On this account, then, *doch* is optimal under the default constraint ranking, *nein* is optimal under a perturbed constraint ranking, while *ja* is never optimal. We assume here, as above, that the acceptability rating assigned by a subject to a certain polarity particle in a certain configuration is not only determined by the grammar of that specific subject but also by the likelihood that she assigns to the use of that particle in that configuration by other speakers in her language community. This, then, accounts for the high acceptability of *doch*, the median acceptability of *nein*, and the unacceptability of *ja*.

In bare particle responses to negative antecedents, the pressure to avoid ambiguity is relevant as well. In such responses, *doch* conveys the intended interpretation unambiguously, while both *nein* and *ja* are ambiguous between an [AGREE,-] and a [REVERSE,+] response, given that in German both particles are possible in [AGREE,-] responses. Given the overall higher frequency of *ja* in [AGREE,-] responses, its use as a bare particle response in [REVERSE,+] responses is predicted to

be highly degraded since this choice goes against all relevant constraints.

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 saliency approach based on cross-linguistic considerations.

## 4 Comparison with the saliency approach

On the saliency approach, polarity particles are propositional anaphors that affirm or reject their antecedent discourse referent. Just like in the feature model, the discourse referents are taken to be marked as being positive or negative. Positive sentences introduce a positive discourse referent. 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 saliency approach and the feature model concerns polarity responses to negative initiatives. On the saliency approach, polarity responses to such initiatives have access to two discourse referents, a negative discourse referent,  $\bar{p}_{DR}$ , introduced by the whole negative sentence, and a positive discourse referent,  $p_{DR}$ , introduced by the lower TP of the negative sentence. The overlap of positive and negative polarity particles in responses to negative initiatives in both English and German is due, on this approach, to the availability of these two discourse referents as antecedents to polarity particles. Differences in particle preference in such responses are claimed 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  $\bar{p}_{DR}$ , while for the *ja*-group, the relative salience of the two discourse referents is reversed.

In this account, positive polarity particles are treated as always agreeing with the antecedent, while negative particles 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 such initiatives introduce, while preferences are explained in terms of intrinsic salience differences between the two discourse referents.

In contrast, recall that in the feature model, 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 saliency 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.<sup>7</sup>

We will argue, however, that this gain in simplicity vanishes once further cross-linguistic data is taken into account. More specifically, we will argue that there are certain mismatches between polarity particles and other propositional anaphors which are unexpected on the saliency approach (§4.1) and that sensitivity to the absolute parameter is a necessary component of a theory of polarity particles cross-linguistically (§4.2). We end with a contrast in typological predictions (§4.3).

---

<sup>7</sup>This point is made in Goodhue and Wagner (2018).



## 4.1 Mismatches between polarity particles and other propositional anaphors

Essential to the saliency approach is the assumption that negative sentences introduce two discourse referents,  $p_{DR}$  and  $\bar{p}_{DR}$ , both of which can in principle serve as antecedents to polarity particles. To motivate this assumption, Krifka (2013, p.6) considers examples like (15), which show that propositional anaphors like *this*, *that*, and *it* can target the positive discourse referents introduced by negative sentences.

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

In order to rely on this fact in accounting for polarity particle responses to negative initiatives in the way the saliency 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.

We exemplify with 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 saliency 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 saliency 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 (16):<sup>8</sup>

- (16) 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  
'The Incas didn't reach Tahiti, even though Heyerdahl claimed so.'

The interpretation of this Japanese example is parallel to that of (15), i.e., (16) can be interpreted as saying that the Incas didn't reach Tahiti even though Heyerdahl claimed that they did. 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 saliency account but not for the feature model. The challenge is not unsurmountable, since one can 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.<sup>9</sup>

---

<sup>8</sup>The reported judgment of this sentence, as well as the judgments reported below for Romanian, have been confirmed by two native speakers.

<sup>9</sup>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:

## 4.2 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 saliency 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 saliency 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 this model, *doch* signals that the response is positive, and that it reverses the antecedent, 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 saliency account *doch* exclusively encodes information about the sentence polarity of the antecedent, while in the feature model it encodes information about the sentence polarity of the prejacent as well as that 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).<sup>10</sup> 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.<sup>11</sup> We exemplify in (17):

- (17) A: Petru nu a telefonat. / Petru nu a telefonat? ‘Peter did not call. / Did Peter 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 overtly marks the sentence polarity of the response. Furthermore, the absolute polarity particle in the response must always agree with the sentence polarity of the prejacent. This is illustrated in (18) and (19):

- (18) 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.’
- (19) A: Petru nu a telefonat. ‘Peter did not call.’  
B: Ba da, (a telefonat.) / Ba, a telefonat / \*Ba nu, a telefonat.’ ‘No, he did.’

- 
- (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.

<sup>10</sup>For detailed discussion of the polarity particle system of Hungarian, which supports the same argument, see Farkas (2009).

<sup>11</sup>For details concerning when this third particle may, must or cannot be used in reversing responses, see the papers cited above.

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, the realization rules of Romanian require the absolute feature of the response to be always 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 is taken as an indication 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—both the use of an absolute polarity particle and that of an ‘echo’ ensure this. 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 saliency 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, but that in these languages,  $p_{DR}$  is more salient than  $\bar{p}_{DR}$ . This is so because in these languages only the positive 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 saliency difference as well. A contrast between Romanian and English is predicted with respect to examples such (15), whose Romanian equivalent is given in (20).

- (20) Incașii nu au ajuns până la Tahiti, cu toate că Heyerdahl a afirmat asta.  
 ‘The Incas did not reach Tahiti even though Heyerdahl claimed this.’

The prediction would be 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 (21) we give an example where *asta* can easily access the negative discourse referent:

- (21) 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?’

Again, the prediction would be that interpreting *asta* as referring to Peter not having watered the plants is harder in Romanian than in the corresponding English sentence. More generally, note that the saliency approach relies on the assumption that languages may differ with respect to the relative saliency 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 saliency 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.<sup>12</sup>

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 feature 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.

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. Note that in the feature model both absolute polarity particles and echoes realize the absolute polarity feature of the response. Thus, the interaction between particles and echoes in Romanian 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 saliency 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 cross-linguistically 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*.

Polarity particles interact with ‘echo’ responses in other languages as well. In Irish, for instance, there is no *yes* particle, i.e., no particle realizing either [AGREE] or [+]. The feature model correctly predicts that in such a language some form of an echo response will be used in [AGREE,+] responses. The data from Catalan and Russian discussed in González-Fuente et al. (2015) also shows a strong interconnection between polarity particles and echo responses. The two languages contrast in that in Catalan the positive particle *sí* may realize both [AGREE] and [+], while in Russian, the positive particle *da* can only realize [AGREE]; [+] in this language is realized by echo responses only. These interactions between polarity particles and echo responses find a natural account in the feature model but pose a challenge for the saliency account.

In sum, both the saliency 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 saliency account polarity particles only encode information about the sentence polarity of the antecedent. While this difference is subtle, taking the former view has two advantages: (i) it allows a simpler account of systems like those found in Romanian, and (ii) it is better equipped to explain the interplay between particles and echoes across languages.

---

<sup>12</sup>Something more would have to be said to rule out a simple *da* in [REVERSE,+] responses.

### 4.3 Typological predictions concerning languages with three polarity particles

The two approaches compared here also differ with respect to expectations about languages with three polarity particles. In the feature model, the existence of systems with three particles, where the third particle expresses [REVERSE] or [REVERSE,+], is expected. The feature [REVERSE] is more marked than the feature [AGREE], and having a particle realizing it is natural; the combination of [REVERSE] and [+] is the most marked feature combination, and therefore it is expected that languages may have a designated particle for it.

Under the saliency account, the least surprising polarity particle system is one with just two particles, schematized in (22):

- (22) a.  $p_1$ : affirms most salient antecedent  
b.  $p_2$ : rejects most salient antecedent

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

Thus, under the saliency account one could explain the existence of three-particle systems in functional terms, arising under pressure to reduce 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, so as to be able to account for [AGREE,-] responses. Note also that the functional explanation does not hold for the system of three particles found in French. In this language, *oui* and *non* are parallel to Romanian *da* and *nu*, while the third particle, *si* is parallel to German *doch*. Under the assumptions of the saliency account, *oui* targets a positive discourse referent and agrees with it, while *non* targets a positive discourse referent and rejects it. Thus, *oui* as a response to a negative initiative would be unambiguously interpreted as confirming the positive discourse referent, i.e., as a [REVERSE,+] response, while *non* would unambiguously be interpreted as rejecting the positive discourse referent, i.e., as an [AGREE,-] response. The functional need for a third particle does not arise. Nonetheless, French, just like German, has a [REVERSE,+] particle, *si*. We conclude that the feature model is in a better position than the saliency account to explain the existence of certain attested three particles systems.

Now let us consider an *unattested* three particle system which is highly unexpected in the feature model and unobjectionable under the saliency account. This system would be like German in that it has a particle  $p_1$ , just like *ja*, marking agreement with the antecedent, and a particle  $p_2$ , just like *nein*, marking rejection of the antecedent, as well as a third particle  $p_3$  which presupposes a positive antecedent and confirms it, as schematized in (23):

- (23) An unattested three particle system  
a.  $p_1$  confirms the most salient antecedent  
b.  $p_2$  rejects the most salient antecedent  
c.  $p_3$  presupposes a positive antecedent and confirms it

The third particle here is the opposite of *doch*, which on the saliency account presupposes a negative antecedent and reverses it. Let us further 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). In this language, then, one would use  $p_1$  or  $p_3$  to agree with a positive antecedent, while to reject such an antecedent, one would use  $p_2$ . To agree with a negative antecedent, one would use  $p_1$ , while to reject a negative antecedent one would use  $p_2$ . Thus, in this language one has two ways of marking agreement with a positive antecedent; blocking makes  $p_3$  preferable over  $p_1$ .

While such a system is unobjectionable under the saliency account given its similarity to German, it is unexpected under the feature model. In the feature model,  $p_1$  would be analyzed as realizing [AGREE],  $p_2$  would be analyzed as realizing [REVERSE], and  $p_3$  would be analyzed as realizing [AGREE,+]. Having a particle dedicated to realizing [AGREE,+], which is the least marked feature combination of all, is highly unexpected in this model. The feature model constrains the type of polarity particle systems that we expect to find cross-linguistically via markedness considerations. In the absence of constraints on the type of presuppositions polarity particles may involve, the typological predictions of the saliency approach are less clear.

## 5 Conclusion

The first aim of this paper was to account for the experimental findings on polarity particle usage 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 saliency 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 concluded that ultimately, both systems need to make reference to sentence polarity. Furthermore, the assumption that polarity particles encode information about the sentence polarity of the pre-jacent as well as the antecedent (as in the feature model) rather than just that of the antecedent (as in the saliency account) has advantages in accounting for the interplay between polarity particles and echo responses.

The two approaches further contrast in that the saliency approach has to stipulate cross-linguistic differences in the saliency of the two discourse referents introduced by negative sentences while the feature model does not. The differences assumed on the saliency 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 support for this account. A third difference between the two approaches concerns the typological predictions they make concerning systems with three polarity particles that we should or should not expect to find 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. 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.
- González-Fuente, S. Tubau, M. T. Espinal, and P. Prieto. Is there a universal answering strategy for rejecting negative propositions? typological evidence for the use of prosody and gestures. *Frontiers in Psychology*, 2015.
- 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.
- B. M. Jones. *The Welsh Answering System*. De Gruyter, 1999.
- 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.
- 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.
- E. Pope. *Questions and Answers in English*. Mouton, The Hague, 1976.
- F. Rainer. Towards a theory of blocking: The case of Italian and German quality nouns. In G. Booij and J. van Marle, editors, *Yearbook of Morphology 1988*. Foris, Dordrecht, 1989.
- 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.