

Polarity particles revisited

Donka F. Farkas and Floris Roelofsen

March 28, 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) (henceforth R&F), 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.

The first goal of this paper is to work out in full detail an account of the new German data in the feature model. In the process, we will clarify and better articulate those aspects of the feature model that are responsible for how preference patterns are to be explained. Our second goal is to deepen the comparison between this model and the saliency account 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 the new 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

R&F propose a general approach to polarity particle responses, exemplified in (1), whose cross-linguistic predictions are checked against data from English, Romanian, Hungarian, French and German.

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

The earliest work on polarity particle responses (Kuno, 1973; Pope, 1976; Sadock and Zwicky, 1985) concentrated on cross-linguistic differences that are detectable in polarity particle usage in agreeing responses to negative antecedents. 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* parameters. 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. 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.¹

Building on suggestions made in Farkas and Bruce (2010), Roelofsen and Farkas (2015) 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 assuming 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 assumed to be made up of a polarity head that hosts an absolute and a relative polarity feature, and a prejacent clause, which may be fully or partially elided. In (1), *yes* and *no* realize the polarity features of the response and *he is/he isn’t* are the prejacent.

The semantic contribution of the polarity features is as follows (see Roelofsen and Farkas 2015 for a precise formulation).

- (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.²
- 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.

Polarity features are realized morphologically by polarity particles. Which features are realized by which particles is a language specific matter, captured in the model proposed in R&F by two factors. First, a language specific *feature-particle mapping* associates each particle in the language to the feature(s) that can be realized by that particle. Second, the form of polarity responses in each language may be affected by language specific *realization rules*, which may specify that certain

¹Sadock 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.

²R&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. The assumption that propositional discourse referents are marked for polarity is analogous to the assumption that individual discourse referents are marked for gender. For arguments supporting the latter assumption, see Farkas and Zec (1995).

features must be realized obligatorily, or that there is a preference for overtly realizing one type of feature over another.³

In these terms then, a purely ‘truth based’ language, such as Japanese, is one that always favors the realization of relative features over the realization of absolute features. Such a language will have two particles only, one realizing [AGREE], and the other realizing [REVERSE]. A purely ‘polarity based’ language, on the other hand, is one that always favors the realization of absolute features over relative ones. Such a language will also have just two particles, one realizing [+], and the other realizing [−].

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 one 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 feature model account of English 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,+] responses⁴:

- (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,+]

³A 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’.

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

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 marking of marked elements. Thus, given that in [AGREE,−] responses [−] is marked and [AGREE] is unmarked, in the absence of other constraints, markedness considerations would 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 is not the only factor in accounting for preference patterns in case the grammar allows a choice of particles for a particular response type.

Three particle systems in the feature model 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. In the latter case, the form of the particle is expected to be either a special positive or a special adversative particle.

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) preadjacent as long as this ensures that the absolute polarity of the response is recoverable. 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 may require the realization of a particular type of feature in all polarity responses, or a preference for realizing one type of feature over another. More specifically, 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*

Furthermore, it is assumed, without further discussion, that *doch* blocks both *ja* and *nein* in [REVERSE,+] responses. Thus, on this account German is like English in having two multifunctional particles, and it is like French in having a third [REVERSE,+] particle. However, unlike in French, the third particle is realized by an adversative morpheme rather than an emphatic positive morpheme.

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

Avoid Ambiguity There is one further aspect of the feature model that is relevant to the discussion below. Namely, R&F assume that the form of polarity responses is subject to an AVOID AMBIGUITY constraint which militates for choosing a form that allows the interlocutor to recover the intended message.

- (10) AVOID AMBIGUITY
 Avoid expressions that are perniciously ambiguous.

Note that this principle, which echoes a subpart of Grice’s Manner Maxim, is not an ad hoc addition to the feature account, though it is particularly relevant to bare particle responses where the whole communicative burden is borne by the particle. R&F rely on it to explain experimental data in English in Kramer and Rawlins (2012), involving contrasts between bare particle responses and responses with an overt preajcent. Note that when responding to a positive initiative, it is crucial to differentiate [AGREE,+] from [REVERSE,−] responses; when responding to a negative initiative, it is crucial to differentiate [AGREE,−] from [REVERSE,+] responses. This is so because in polarity responses, the communicative goal reduces to signaling commitment to the antecedent or commitment to its complement.

Summary In the feature model sketched above, the language specific aspect of the grammar of polarity particles concerns the connection between particular features and particular particles in the language (feature-particle mappings), as well as possible constraints encoding the obligatory realization of certain features or preferences for the realization of certain features over that of others (realization rules). Universal markedness considerations constrain both of these language specific components. They also play a role in accounting for preference patterns in responses where a choice is allowed. A general pressure against pernicious ambiguity further constrains the form of polarity responses.

Whether a feature is realized or not in a particular situation therefore depends on four factors. The first is whether the language has a particle realizing that feature or not. Thus, Japanese has only particles realizing relative features, 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. Thus, Romanian requires the overt marking of absolute features, while Hungarian has this requirement only for the marked absolute feature [−]. The third factor concerns the markedness of the feature relative to the other feature in the response. For instance, in [AGREE,−] responses in English realizing the marked feature [−] is preferred, other things being equal, over realizing the unmarked feature [AGREE]. 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 extended account of German in the feature model

New data Claus et al. (2017) report a series of experiments meant to probe the details of polarity particle usage in German. We review these results here, and extend our earlier account of German

presented above so as to overcome the challenges they pose.

Confirming the prediction in (9d) Claus et al. found that both *ja* and *nein* are indeed 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, an issue we will return to below.

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 preajcent, *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) is too blunt if one wants to account for the fact that *nein*, while not perfect, is not completely unacceptable in [REVERSE,+] responses either.

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).⁵ We take it that German speakers do indeed fall into two groups with respect to their preferred particle in [AGREE,-] responses, a *ja*-group, who prefer *ja* in such responses, and a *nein*-group, who prefer *nein*. To account for the *ja*-group, we propose that, in addition to the rules in (8), the grammar of these speakers contains a realization rule favoring the realization of relative features, which outranks the preference for realizing marked features over unmarked ones. 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, the analysis in R&F accounts for their judgments 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 that 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.

To sum up so far, the feature model can account for the new experimental data concerning [AGREE,-] responses by assuming that the grammar of polarity particles for the speakers in the *nein*-group is as in the original account, while the speakers in the *ja* group have, in addition, the

⁵As 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).

realization rule in (11):

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

In more general Optimality Theoretic terms, particle choice in [AGREE,−] responses in German is subject to two conflicting constraints: one general constraint that militates for the realization of the marked feature [−], resulting in a preference for *nein*, and another language specific constraint that militates for the realization of the relative feature, resulting in a preference for *ja*. In Optimality Theoretic terms markedness constraints and realization rules can be formulated as faithfulness constraints, as in (12) and (13), where the former is a constraint rooted in a universal pressure for the overt marking of marked features, and the latter is a narrow constraint favoring the realization of one type of feature over another.

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

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, a Stochastic OT framework would have to 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).

Note that cross-linguistic and dialectal variation based on relative ranking of constraints is precisely what one would expect. From a wider cross-linguistic perspective what we see is the need to generalize the language specific realization rules discussed in R&F for Romanian and Hungarian. In our earlier proposal, a categorical constraint requiring the realization of absolute features was needed for Romanian, and a categorical language specific constraint requiring the realization of the marked absolute feature was needed for Hungarian. Markedness considerations delimit the language specific realization requirements: one would not expect a constraint requiring the realization of [+] but not of [−], or the realization of [AGREE] but not of [REVERSE].

The novel experimental data on German shows that a constraint concerning the realization of relative features is needed as well, a fact that is perfectly in line with the general properties of the feature model. One only needs to allow realization rules to function as preferences in some languages and absolute requirements in others. Preferences in particle choice then can be sensitive either only to general markedness considerations, as in English and in the *nein*-group in German, or additionally to language (or dialect) specific preferences for the realization of particular features, as in Romanian, Hungarian, and the *ja*-group in German. In OT these differences are accounted for by variation in the ranking of the relevant constraints.

Account of [reverse,+] responses Recall that the experimental evidence in Claus et al. (2017) shows that in [REVERSE,+] responses 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

data we now have experimental evidence for. In order to account for these data we refine our blocking approach. The basic idea is that *doch* is best because it expresses both features, *nein* is worse because it only expresses [REVERSE] but acceptable because the feature expressed is the marked one; *ja*, on the other hand, is worst because it has nothing to recommend it: it is less expressive than *doch* and the feature it expresses is the unmarked one.

Blocking, in some form or another, has been present in linguistics since Pānini (5th or 4th century BC), who argued that of two competing rules, the more restricted one generally has precedence. Aronoff (1976) and many others have invoked blocking with respect to word formation, explaining the non-occurrence of certain forms as a consequence of the existence of other, competing forms expressing the same meaning. Most relevant for our present concerns is that blocking may “adjudicate 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 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.

Moreover, it is important to take into account that blocking is not always absolute, something that has been recognized at least since Paul (1886). Rather, 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 that are relevant to our present concerns are given in (14):

- (14) *Blocking constraints*
- a. EXPRESSIVENESS: Express meaning (feature content).
 - b. FREQUENCY: Prefer the use of frequent forms.

These constraints interact with REALIZE MARKED, the constraint that militates for the realization of marked features over unmarked ones, and AVOID AMBIGUITY, the constraint that militates for the choice of expression that conveys the intended interpretation in an unambiguous form.

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 thus explained. Now, within a Stochastic OT framework we may assume that FREQUENCY is ranked just below EXPRESSIVENESS and that this ranking can be perturbed. 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. This, then, explains the medium acceptability of *nein* and the unacceptability of *ja*.⁶

In bare particle responses to negative antecedents, AVOID AMBIGUITY becomes 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

⁶We make a further subtle prediction, namely that for speakers of the *ja*-group, the difference in rating between *nein* and *ja* in [REVERSE,+] responses will be greater than for speakers of the *nein*-group, because for the former, *nein* obeys REALIZE RELATIVE, while *ja* does not. Data supporting this prediction was indeed reported in Meijer et al. (2015): “participants with a preference for *ja* as an affirming response [i.e., speakers from the *ja*-group] rated rejecting responses with *nein* higher than did participants with a preference for *nein* as an affirming response [i.e., speakers from the *nein*-group] (M=5.00 vs 3.47), with the difference being especially pronounced with bare particles (M=4.49 vs 2.01).” This finding is not discussed in Claus et al. (2017).

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 use in German within the feature model. To account for the fact that German speakers fall into two groups with respect to their preference for *nein* or *ja* in [AGREE,−] responses, we have proposed that the grammar of the speakers in the *ja*-group involves a constraint favoring the realization of relative features over absolute ones, which diminishes the effect of REALIZE MARKED for this dialect in [AGREE,−] responses. This type of language specific constraint favoring the realization of particular features was already proposed in R&F for Romanian and Hungarian. To account for the preference patterns in German [REVERSE,+] responses we refined the blocking account we suggested earlier, basing it on assumptions that are needed elsewhere as well.

4 Synopsis of the saliency approach

The saliency approach has focused on accounting for polarity particle responses in English (Krifka, 2013) and German (Claus et al., 2017). Before summarizing these specific accounts we characterize here the main aspects of the 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, *ja* and *yes* pick up the positive discourse referent and affirm it, while *nein* and *no* 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.

Thus, in this model, positive polarity particles can be treated as always agreeing with the antecedent, and negative particles can be 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 *nein/no* 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.

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.⁷

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

⁷Another difference between the saliency approach and the feature model is that polarity particles are treated as propositional anaphors in the former, with the ‘prejacent’ treated as an appositive clause, while in the feature model the prejacent, when present, forms a constituent with the particle. Roelofsen and Farkas (2015) and Goodhue and Wagner (2018) present several arguments in favor of the second view. However, this view may be adopted rather straightforwardly on the saliency approach as well; it seems this would not require any changes in its core assumptions.

with. The saliency approach would become even more attractive if one could maintain that reference to the absolute parameter (or an absolute polarity feature) is superfluous in other languages as well.

We will argue, however, that this gain in simplicity is only apparent, and vanishes once further data is taken into account. After summarizing the saliency account of German and English in Claus et al. (2017) and Krifka (2013) below, in Section 5 we compare this account to the feature-based one, and then turn to a comparison of the saliency approach and the feature model from a wider cross-linguistic perspective. The conclusion we will reach is that sensitivity to the absolute parameter is a necessary component of a theory of polarity particles cross-linguistically; exploiting the two discourse referents introduced by negative initiatives does not allow us to dispense with this parameter. An account of English that does not use it is not, therefore, necessarily preferable to one that does. On the other hand, we will see that the saliency approach, in order to capture certain cross-linguistic differences, must assume that the relative salience of the two discourse referents introduced by negative sentences varies across languages or dialects. Whether this assumption receives independent motivation remains an open issue. For the time being, the fact that the feature model can account for cross-linguistic differences without assuming language-specific salience-rankings is, in our view, an advantage.

The saliency account of German We summarize below the saliency account of German given in Claus et al. (2017). Just as in the feature model, this account assumes that polarity particles need a propositional antecedent that introduces a discourse referent marked as positive or negative. Following Krifka (2013), polarity particles are treated as propositional anaphors; *ja* affirms the propositional discourse referent it targets, while *nein* affirms the negation of that discourse referent. The third particle, *doch*, presupposes that the discourse referent it targets is negative, and it rejects it, i.e., it asserts its negation.

As already mentioned, crucial to this account is that negative initiatives introduce two discourse referents, \bar{p}_{DR} and p_{DR} , both possible antecedents to subsequent response particles. Claus et al. assume that the relative salience of these two discourse referents is a matter of cross-dialect, and, presumably, cross-linguistic variation.⁸ In particular, Claus et al. assume that for the *ja*-group, \bar{p}_{DR} is more salient than p_{DR} , while for the *nein*-group, the reverse salience relation obtains. How particles are interpreted, as well as the various preference patterns we find are regulated in this account by a set of ranked constraints, which we summarize below.

A general highly ranked constraint, PRES, penalizes violations of presuppositions. Its effect with respect to particle choice is to ensure that *doch* is used only in rejecting responses to negative initiatives. The second operative constraint is *NONSAL, relevant to anaphoric expressions generally, which penalizes the choice of a less salient antecedent when a more salient one is available. The third relevant constraint is BLOCK, a meta-constraint “by which optimal form-meaning pairs suppress the expression of the same meaning by a different form, or the use of the same form to express a different meaning” (Claus et al., 2017, p.10). The relevant application of BLOCK here is meant to ensure that the pragmatic rule of Maximize Presupposition is obeyed.⁹ The relative ordering of these constraints is as in (15):

(15) PRES > *NONSAL > BLOCK

In responses to a positive initiative, which introduces a single positive discourse referent, this account correctly predicts that *ja* will be interpreted as agreeing with the unique positive discourse

⁸In Krifka (2013) it was assumed that the relative salience of these two discourse referents is determined by context, but the results reported in Claus et al. (2017) disconfirmed this hypothesis.

⁹Claus et al. do not go into further details concerning BLOCK or how it interacts with Maximize Presupposition.

referent, *nein* will be interpreted as reversing it, and *doch* will be ruled out because it violates PRES:

- (16) A: Bill raucht. ‘Bill smokes’.
B: Ja. (= He does.)
B: Nein. (= He does not.)
B: #Doch. (= He does not.)

The picture is more complex in responses to negative initiatives, which introduce two discourse referents, \bar{p}_{DR} and p_{DR} . The polarity particle pattern in responses to such initiatives is sensitive to the relative salience of these two discourse referents.

For the speakers in the *ja*-group, \bar{p}_{DR} is more salient than p_{DR} . Thus, interpreting *ja* as an agreeing, i.e., as an [AGREE,-], response, obeys all relevant constraints because under this interpretation the particle selects the most salient antecedent, namely \bar{p}_{DR} and affirms it. Interpreting *ja* as a rejecting, i.e., as a [REVERSE,+], response, on the other hand, violates the highly ranked *NONSAL constraint because to get this interpretation, the selected antecedent would have to be the less salient p_{DR} .

Turning now to a *nein* response to a negative initiative, for speakers in the *ja*-group, an agreement interpretation violates the highly ranked *NONSAL constraint because for this interpretation, *nein* has to choose the less salient discourse referent p_{DR} and reverse it. Interpreting *nein* as a rejecting response, i.e., as a [REVERSE,+], response, on the other hand, obeys *NONSAL because in this case the selected antecedent is \bar{p}_{DR} . This interpretation, however, violates BLOCK, which militates for the use of *doch* in such responses.

Under this account, for speakers in the *ja*-group, the optimal particle choice in agreeing responses to negative initiatives is *ja*. The optimal particle choice in rejecting responses to such initiatives is *doch*. The choice of *nein* in such a response violates BLOCK while the choice of *ja* violates the higher ranked *NONSAL. Claus et al. make the assumption that such a situation leads to gradient judgments which rank *doch* highest, followed by *nein*, followed by *ja*.¹⁰

The crucial difference between the speakers in the *ja*-group and those in the *nein*-group is that for the latter, the saliency of the two discourse referents introduced by negative sentences is reversed: for the speakers in the latter group, p_{DR} is more salient than \bar{p}_{DR} . As a result, these speakers will prefer *nein* to *ja* in agreeing responses to negative initiatives. In the case of rejecting responses to such initiatives, *doch* is the optimal particle choice for these speakers too, with *ja* violating BLOCK and *nein* violating *NONSAL. For the speakers in this group then, *ja* should be better than *nein* in rejecting responses to negative initiatives because the former violates a lower ranked constraint while the latter violates a higher ranked one. This issue is not discussed in Claus et al. so we leave it open here.

The saliency account of English In Krifka (2013), English *yes* and *no* are treated on a par with German *ja* and *nein*, as accepting and reversing their antecedent, respectively. Claus et al. do

¹⁰Note that it is not quite straightforward to motivate this assumption. In standard OT, the optimal expression is predicted to be acceptable and all non-optimal expressions are predicted to be unacceptable. To account for gradient judgments, one has to work within a Stochastic OT framework. But, even if the account under consideration here would be implemented in a Stochastic OT framework, it would not immediately predict the gradient judgments. This is because *doch* does not violate any of the constraints considered, so under any possible ranking of the relevant constraints, *doch* would come out as the optimal expression. In order to derive the desired predictions, Claus et al. could take FREQUENCY into account, as we did in our own proposal above. Then, competition between PRES and FREQUENCY could be held responsible for the finding that, even though *doch* is the optimal choice, *nein* is marginally acceptable as well in [REVERSE,+], responses.

not discuss how the proposed modified account of German would extend to English. The simplest account is to assume that the grammar of speakers of English is the same as that of the German speakers in the *nein*-group in that the negative discourse referent introduced by a negative sentence is less salient than the positive discourse referent. In English then, *no* will be preferred over *yes* in agreeing responses to negative antecedents. In rejecting responses to negative antecedents, English has no particle equivalent to *doch*, and therefore BLOCK will not be relevant. In such responses *yes* targets the more salient discourse referent and affirms it, while *no* targets the less salient discourse referent and rejects it. Without further stipulations, the account predicts then that in rejecting responses to negative antecedents *yes* will be optimal while *no* will not be. Under this account the predicted preference patterns for polarity particles in English are as in (17):

- (17) English preference patterns in responses to negative initiatives in the saliency account
- a. Agreeing responses to negative initiatives, i.e., [AGREE,−] responses
no > *yes*
 - b. Rejecting responses to negative initiatives, i.e., [REVERSE,+] responses
yes > *no*

These predictions are identical to those made by the feature model for [AGREE,−] responses. For [REVERSE,+] responses, the feature model predicts the opposite: *no*, realizing [REVERSE], is predicted to be preferred over *yes*, which realizes [+]. Which of these predictions is correct remains an open empirical issue.

5 The two models compared

5.1 Comparison of the two accounts of English and German

We have seen above that the details of the German data, as well as the basic data in English, can be captured in both approaches considered here. The details of accounting for the interplay between bare and non-bare particle responses in English is left for future research in the saliency account, and for both approaches, the issue of the details of particle choice in agreeing responses to negative sentences with quantifiers remains an open issue. Before going to a comparison beyond English and German, we consider here the two theories from a more general perspective.¹¹

Starting with what the two approaches have in common, note that in both the feature model and the saliency approach, polarity particles are anaphoric to a propositional discourse referent introduced in the immediately preceding discourse. In both approaches, these discourse referents are marked as + or − depending on the polarity of the sentence that introduces them.¹² In both approaches, polarity responses commit the speaker to the propositional antecedent discourse referent or its complement.

Turning to the main differences, on the saliency approach polarity particles are not sensitive to the polarity of the response, while in the feature model they are. Note, however, that in the saliency approach, absolute polarity is nonetheless relevant to polarity particles, given that in this approach, polarity particles may involve presuppositions concerning the polarity of their antecedent. Thus, in the saliency approach, *doch* presupposes that the discourse referent it picks up is −, and the

¹¹One difference between the two proposals that we will leave out of consideration here is the coverage of polarity particle responses to *disjunctive* questions. Such cases have received much attention within the feature model, and provide crucial support for some of its core assumptions, while in the saliency account they have received much less attention. Of course, this does not mean that the saliency account could not be extended in a suitable way so as to deal with them, but doing so does not seem to be a trivial task.

¹²R&F go into details concerning how this is done, details that can easily be adopted by the saliency approach.

effect of the particle is to reverse this discourse referent. Given the way [AGREE] and [REVERSE] are defined in the feature model, this amounts to having the effect of expressing [REVERSE,+]: if the antecedent is $-$ and the particle reverses it, the response will be $+$. So absolute polarity, though not that of the response but rather that of the antecedent, does play a role in the saliency approach as well.

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. As a result, polarity particle responses to negative sentences may target either one of these two discourse referents. This assumption is motivated by the fact that discourse anaphors other than polarity particles may target the positive discourse referent introduced by a negative sentence, as exemplified in (18):

- (18) A: Tom didn't steal the painting.
 B: Many people believe *that*.
 They have great faith in his innocence. /
 They think he is guilty because of his past behavior.

The propositional anaphor *that* in (18) can be interpreted either as referring to the proposition expressed by *Tom didn't steal the painting*, i.e., \bar{p}_{DR} , or to the proposition expressed by *Tom stole the painting*, i.e., p_{DR} .¹³

Given these observations, it is indisputable that the TP embedded under negation can serve as an antecedent to certain propositional anaphors. It would appear simplest, then, to rely on this fact in the account of polarity particles in responses to negative sentences in precisely the way the saliency account does. In order to do so, one has to further assume that the anaphoric reach of polarity particles is similar to that of other propositional anaphors such as *that* or *it*.

Now recall that the saliency account assumes that for speakers in the *ja*-group, the negative discourse referent introduced by sentences like (18) is more salient than the positive discourse referent, while for speakers in the *nein*-group it is the other way around. Furthermore, it is assumed that the relative saliency of these two discourse referents is not affected by the larger context. This account predicts then that German speakers in the two groups will differ with respect to how they interpret examples such as (18), with speakers in the *ja* group preferring \bar{p}_{DR} as the antecedent of *that*, and speakers in the *nein* group preferring p_{DR} . Under the assumption that the saliency account of English is parallel to that of the *nein* group, English speakers are also predicted to prefer p_{DR} rather than \bar{p}_{DR} as the antecedent for propositional anaphors in examples like (18), other things being equal. These predictions await experimental verification.

Note that in order to account for particle preferences in German and English responses to negative initiatives, the saliency account has to stipulate cross-linguistic (and cross-dialectal) differences in the relative saliency of the two discourse referents introduced by negative sentences. In the absence of independent evidence for this stipulation, this is a drawback. The absence of contextual effects on the relative salience of these discourse referents is also surprising on this account.

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.¹⁴

¹³Note that the latter interpretation is practically the only one possible if an adversative connective such as *but* or *although* is used in front of the sentence. Which interpretation is preferred in the absence of such a connective is an open question that can be answered only after further experimental investigations.

¹⁴Note 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

We now turn to a comparison of the two accounts based on a broader empirical basis.

5.2 Going beyond English and German

The empirical focus of the saliency approach to polarity particles has been on English and German. While it is a risky business to engage in speculation on how a theory would deal with data it has not explicitly addressed yet, we will engage in this very endeavor in this subsection. We do so because when theories are compared based on how they deal with data from a narrow set of languages, one approach may appear simpler and more elegant than another, but this difference may turn out to be illusory when the empirical lens is widened. We have already seen that when the saliency theory is applied to English it appears simpler than the feature model applied to the same data: there is no need to make reference to the absolute parameter. Once the scope of the inquiry is minimally widened to German, however, complications have to be introduced, such as the assumed inherent salience differences, and reference to the absolute polarity of the antecedent. In what follows we widen the lens further, considering languages that are not closely related to either English or German.

Mismatches between polarity particles and other propositional anaphors One challenge for the saliency account comes from the fact that the anaphoric potential of polarity particles does not always match that of other propositional anaphors. This is the case, for instance, in Japanese. Recall that Japanese is a two polarity particle language, where *hai* affirms the antecedent and *iie* reverses it, without regard to the 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 must be assumed to introduce a single propositional discourse referent corresponding to the whole negated sentence. The saliency account predicts that Japanese will differ from English, then, also in that propositional anaphors following a negative sentence will be unable to pick out the positive discourse referent, and thus be interpreted unambiguously as referring to the negative sentence. This prediction appears to be falsified, as shown in (19):

- (19) a. Edo-wa kukki-o nusumanakatta kedo, ooku-no hito-wa soo sinziteita.
Ed-TOP cookies-ACC steal-NEG-PAST but many-GEN people-TOP so believe-PAST
'Edo didn't steal the cookies but many people believed it [that he did].'
b. Edo-wa kukki-o nusumanakatta kedo, minna-wa soo sinzitei-nai.
Edo-TOP cookies-ACC steal-NEG-PAST but everyone-TOP(wa) so believe-NEG
'Edo didn't steal the cookies but not everyone believes it [that he didn't].'

As seen in (19), Japanese appears to be parallel to English in that both the positive and the negative discourse referent can serve as antecedent to subsequent propositional anaphors. This poses a challenge for the saliency account but not for the feature model. The challenge is not unsurmountable, since one can assume that the positive discourse referent is not salient enough to be accessible to polarity particles but salient enough to be accessible to other types of anaphora, as assumed by the feature model for all languages, but the assumption that there would be intrinsic saliency differences across languages would be in need of independent justification.¹⁵

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.

¹⁵We will return to another possible account of the Japanese data in an expanded saliency account below.

The role of absolute polarity: back to Romanian and Hungarian Recall that an advantage of the saliency account of polarity particles in English is that by taking advantage of the positive discourse referent introduced by negative sentences it can treat polarity particles in English as being sensitive only to the relative parameter. Flexibility is introduced by assuming that the relative saliency of the two discourse referents introduced by negative sentences varies across languages. The strongest theory in this spirit would be one in which the absolute parameter would not be relevant to polarity particle systems at all.

In contrast, the feature model assumes that polarity particles may be sensitive to both parameters, even within a single language. With respect to negative sentences, it is assumed that only the negative discourse referent they introduce is salient enough to serve as antecedent for polarity particles. The way the polarity features interact in this system, information about the absolute polarity of the response also encodes information about the absolute polarity of the antecedent: if the relative feature of the response is [AGREE], the absolute polarity of the response must be identical to that of the antecedent; if the relative feature is [REVERSE], the absolute polarity of the response must be the opposite of the absolute polarity of the antecedent.

In the saliency account of German, sensitivity to the absolute parameter is introduced via presuppositions concerning the absolute polarity of the antecedent: *doch* presupposes a negative antecedent and rejects it. Note, however, that once presuppositions making reference to absolute 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. In the feature model, *doch* signals that the response is positive, and that it reverses the antecedent, both in terms of content and in terms of polarity. Since from this it follows that the antecedent is negative, the two analyses are distinguished only in that on the saliency account *doch* exclusively encodes information about the absolute polarity of the antecedent, while in the feature model it encodes information about the absolute polarity of the response as well as the antecedent.

Note that once particles may encode presuppositions concerning the polarity of the antecedent, a different account of Japanese is possible in the saliency model as well. Instead of relying on the salience difference between the two antecedents introduced by negative sentences, one can assume that in Japanese both particles presuppose a positive antecedent and *ie* affirms it while *hai* reverses it. This account avoids the problem discussed in (19), but is more complex than the feature account of Japanese, which simply says that *hai* realizes [AGREE] and *ie* realizes [REVERSE].

We now turn to polarity systems in Romanian and Hungarian, languages with a system of three polarity particles that differ from what we find in German. These languages, discussed in detail in Farkas (2009, 2011) and Roelofsen and Farkas (2015), provide further support for the need to allow polarity particles to make reference to absolute polarity and have played an important role in the development of our theoretical proposal. We review the facts that lend support to treating these particles as encoding information about the absolute polarity of the response rather than that of the antecedent.

Romanian and Hungarian are the opposite of Japanese in that in responses to negative initiatives, the negative particle *nu* (R)/ *nem* (H) is used in agreeing responses and the positive particle *da* (R)/ *igen* (H), accompanied by a reverse particle, is used in rejecting responses. We exemplify with Romanian in (20):

- (20) 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.’

The third particle, *ba* (R)/ *de* (H), occurs only in rejecting responses, which is why R&F assume

that it realizes the feature [REVERSE].¹⁶

Going into the details of the Romanian system, 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 polarity of the response. Furthermore, the absolute polarity particle in the response must always agree with the absolute polarity of the prejacent. This is illustrated in (21) and (22):

- (21) 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.’
- (22) A: Petru nu a telefonat. ‘Peter did not call.’
B: Ba da, (a telefonat.) / Ba, a telefonat / *Ba nu, a telefonat.’ ‘No, he did.’

The facts are parallel in Hungarian, except that in reversing responses to negative antecedents, the absolute particle *igen* accompanying the reverse particle *de* is optional.

The account of these facts in the feature model takes *da/igen* and *nu/nem* to realize the features [+] and [−] respectively, while the third particle (*ba/de*) 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), while the realization rules of Hungarian require only the marked absolute feature [−] to be obligatorily realized. The fact that the positive particle occurs in [REVERSE,+] responses, and the negative particle occurs in [REVERSE,−] responses is taken as an indication that the particle is sensitive to the absolute polarity of the *response* rather than that of the antecedent. The fact that an ‘echo’ response may replace the absolute polarity particle can naturally be explained under the assumption that both the particle and the echo overtly mark the absolute polarity of the response. These facts then motivated us to assume that the absolute polarity of the response is relevant to polarity particles, not just that of the antecedent.

We turn now to the issue of how these facts could be captured in the saliency account. The null hypothesis is to treat *da* and *nu* as the equivalents of English and German *yes/ja* and *no/nein* respectively. This means that *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 absolute 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, the positive discourse referent is salient, while the negative discourse referent is not. 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 the account makes is that other propositional anaphors in these languages can access the positive discourse referent introduced by a negative sentence but not the negative one. This prediction is false, as illustrated by the Romanian example in (23):

- (23) A: Petru nu a furat tabloul. ‘Peter did not steal the painting.’
B: Mulți cred asta. ‘Many believe this.’

The anaphor *asta* ‘this’ can be interpreted either as referring to the negative discourse referent, committing B to the claim that many believe that Peter did not steal the painting, or to the positive discourse referent, committing B to the claim that many believe Peter did steal the painting. This

¹⁶For supporting data, and details concerning when the reverse particle may, must or cannot be used in reversing responses, see the papers cited above. In brief, marking [REVERSE] with this particle is obligatory in [REVERSE,+] responses, optional in [REVERSE,−] responses to declarative initiatives and impossible in [REVERSE,−] responses to interrogative initiatives.

shows that the negative discourse referent introduced by negative sentences is salient enough to be accessible to sentential anaphors in Romanian, just like in English.

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 information about absolute polarity to be made relevant in the saliency model, a much simpler account is 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.

Note that the gain in simplicity just achieved is due precisely to the fact that we have allowed particles to involve presuppositions concerning the absolute polarity of their antecedent. As a result, the system has become sensitive to the absolute parameter. Once this is done, 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 information concerning absolute polarity is necessary for an account of polarity particles across languages. In the end, the saliency approach has to resort to information about absolute polarity, while the feature model does not have to make any particular assumptions about saliency differences of discourse antecedents beyond claiming that in all languages, the positive discourse referent introduced by a negative sentence is not salient enough to function as an antecedent to polarity particles.

Now let us return to the empirical observation that in Romanian reverse responses the reverse particle *ba* must be accompanied either by *da/nu*, or by an echo, or by both. Note that under the assumption that both absolute polarity particles and echoes realize the absolute feature of the response, this observation is accounted for by the requirement that the absolute polarity of the response must be overtly realized in Romanian. On the other hand, it is unclear how to account for this observation in the saliency model, where polarity particles do not encode information about the absolute polarity of the response. Here, then, we have evidence that a theory of polarity particle responses cross-linguistically must not only assume that such particles are sensitive to the absolute parameter, but also, more specifically, that they potentially encode information about the absolute polarity of the response, not just that of the antecedent.

In languages other than Romanian and Hungarian, we also find interaction between polarity particles and echo responses. 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 absolute polarity to account for data beyond English. The difference is that in the feature model, the absolute polarity feature encodes information about the absolute polarity of the response, while in the saliency account the absolute polarity information concerns the polarity of the antecedent. While this is not a major difference, taking the former view has two advantages: (i) it allows a simpler account of systems like that in Romanian and Hungarian, and (ii) it is better equipped to explain the interplay between particles and echoes across languages.

In more general terms, once the need to make reference to absolute polarity is recognized in both accounts, the saliency account involves an extra source of possible cross-linguistic variation, namely differences in the relative saliency of the two discourse referents introduced by negative sentences. In the feature model on the other hand no such cross-linguistic saliency difference needs to be assumed. If further research shows that the saliency differences needed to account for polarity particle distribution and interpretation are independently motivated, the overall approach would receive strong support. In the absence of such independent motivation, these assumptions remain stipulative.

Why would languages have systems with three particles? The two approaches compared here also differ with respect to expectations about languages that have three polarity particles rather than two. 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 choose to dedicate a particle to it.

Under the saliency account, the least surprising polarity particle system is one with just two particles, schematized in (24):

- (24) 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 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. 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*; the only difference is that in French the form of this particle is not adversative but rather, it is an emphatic positive particle. Note that in the feature model, [REVERSE,+] particles are expected to be either adversative in character (connected to the feature [REVERSE]) or positive (connected to the feature [+]).

We conclude then that the feature model leads us to correct expectations concerning systems with three particles discussed above, while such systems pose a challenge to the saliency account.

Finally, we will sketch below an unattested three particle system which is highly unexpected in the feature model and unobjectionable under the saliency account. Such a system would be just 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 (25):

- (25) 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 would be the opposite of *doch*, which, in 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). 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 would have two ways of marking agreement with a positive antecedent, with Blocking (or Maximize Presupposition) making the choice of p_3 preferable over p_1 .

While such a system is expected under the saliency account, 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.

More generally, while the feature model constrains the type of polarity particle systems that we expect to find cross-linguistically via markedness considerations involving both relative and absolute polarity features, no such constraints arise from the saliency approach. In this sense, the feature model is more predicative from a typological perspective.

6 Conclusion

The starting point of this paper was an account of the recent 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). We then compared this approach to the saliency account of Claus et al. (2017), which in turn is based on Krifka (2013). 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 absolute polarity. The assumption that the absolute polarity information concerns the polarity of the response (as in the feature model) rather than the polarity of the antecedent (as in the saliency account) has positive consequences concerning 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 predictions they make concerning expected/unexpected systems with three polarity particles. Finally, note that we have set aside here issues concerning the role of intonation. For a thorough discussion of this issue and a comparison between the saliency and the feature models, see Goodhue and Wagner (2018).

There are a host of open issues in this area, some of which we list below. One understudied question concerns the possibility of embedding polarity particles. In English embedded clauses *so* and *not* are preferred to bare *yes* and *no* though there are circumstances in which the bare particles start being acceptable. In many languages, such as those in the Romance group, bare polarity particles occur as embedded clauses though there are minor variations here as well. For instance, while *si* may occur in embedded contexts in French, *ba da/nu* in Romanian cannot. Finally, another issue that is in need of an explanation is the fact that in at least some languages, English among them, embedded particles cannot occur with a prejacent.

Yet another understudied issue is the ability of embedded clauses to serve as antecedents to polarity particles. Intuitively, the closer the connection between the question under discussion and the proposition expressed by the embedded clause, the easier it can serve as an antecedent to a polarity particle response (Goodhue and Wagner, 2018) but more data is needed to make this intuition more precise. Similarities and differences between polarity particles and other propositional anaphors cross-linguistically should also be investigated in depth. Finally, turning back to preference patterns in polarity particles responses to English negative sentences, the role of quantificational DPs in influencing the preference for *yes* over *no* reported in Brasoveanu et al. (2013) is still awaiting explanation.

References

- M. Aronoff. *Word formation in generative grammar*. Linguistic Inquiry Monographs. MIT Press, Cambridge, Massachusetts, 1976.
- 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.
- D. F. Farkas and D. Zec. Agreement and pronominal reference. In *Advances in Romanian Linguistics*, pages 83–101. John Benjamins, 1995.
- 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.
- H. Paul. *Prinzipien der Sprachgeschichte*. Niemeyer, Tübingen, 1886.
- 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.
- 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.

Wordcount: 12393