Fabian Brackhane

Observations on the frequency and function of *ja* in spontaneous spoken German

A large number of studies have already dealt with the description of this particle. However, most of them covered single functional aspects. A comprehensive synoptic analysis of the functional and formal parameters of *ja* has not yet been carried out. Accordingly, the present article intends to supplement the existing literature, aiming to verify and build on the findings of earlier studies on the basis of a much more extensive data sample of several thousand records. To this end, the absolute and relative frequency of particle use is described on the one hand and the functional spectrum of particles in the data used is analysed on the other. It was possible to show that previous studies of *ja* considered a large number of functional aspects but that there are additional uses in the analysed corpus which have not been described in detail so far.

In addition, exemplary illustrations are provided of speaker-specific usages, which show that the complex usability of the particle is strongly influenced by speakers' individual communication strategies. This article aims to give an initial overview of how complex the phenomenon under investigation is while at the same time showing a very high frequency of use in the data in question. A detailed presentation of the phonetic empirical findings is reserved for a separate publication.

Dictionaries take a very differentiated approach to *ja*, and the problem of non-unified terminology becomes apparent, making comparisons difficult. In addition to lexicographical descriptions, since the 1970s a large number of publications have dealt partly or exclusively with the particle *ja*. Their approaches and perspectives are quite heterogeneous. However, only Heringer (1988) and Imo (2013) attempted to catalogue all of the functions found in the samples concerned. While Heringer presented a functionally non-demarcating continuum of (at least) 31 different usage variants of *ja* (especially in the area of modalizing communication strategies), Imo categorically divided the instances he describes into six functions (plus particle combinations) (Table 1), but, like Meer (2007) and Weidner (2015), emphasized the general tendency of *ja* to be multifunctional.

The databases for the aforementioned studies were always quite limited. For example, Imo was only able to draw on 197 instances from a single telephone conversation for his description of the functional spectrum of *ja* (Imo 2013, p. 159). Heringer (1988) even constructed his catalogue of 31 functional aspects of *ja* seemingly without any reference to "real" speech data. Against the background of such "data poverty", this study aimed to review the previous findings on the use of *ja* with a holistic approach. All instances of *ja* in the selected corpus were to be examined and classified by function.

The present study arose within the framework of an investigation that originally intended to focus on phonetic aspects of the particle *ja* on the basis of existing interactional linguistic work. The starting point was an exploratory question addressing which formal variations could be observed in the functional spectrum of the particle in a homogeneous corpus of spontaneous German. For this study, the GECO ("German conversation") corpus developed by the Institut für maschinelle Sprachverarbeitung (IMS), Stuttgart, was selected (Schweitzer et al. 2015). This corpus comprises 46 dialogues exclusively involving female

speakers, each with an average duration of approx. 25 minutes. Twenty-two of these dialogues with 12 female speakers were conducted in a monomodal setting; the speakers could, therefore, only hear each other but not see each other. Only these monomodal recordings were included in the analysis since they were the only ones for which a partial flow of information via the visual channel could be excluded.

For the analyses presented here, only those instances were used which were phonetically an unambiguous "pure" *ja* – [ja]. All instances were excluded that are either particle combinations (*ah ja*, *oh ja*, *naja* or *haja*) or instances with phonetic extensions (*nja*, *jam*...) for which it cannot be assumed beyond doubt that they still have the same semantics as *ja*. This left 3,758 instances (78.3% of the raw hits) for further analysis. All instances were analysed by the author for the described aspects and the annotations were checked for consistency and validity by a student assistant. For each analysed instance, the author determined its communicative function against the background of its context and with reference to Selting/Couper-Kuhlen (2000) for its interactional linguistic aspects. This was done by a combination of substitution and paraphrase, for which the categories proposed by Imo (2013) served as a basis (cf. Table 1). These annotations were checked again with a certain time lag and then evaluated in data sessions with colleagues (all German native speakers).

The functional catalogue developed by Imo (2013) initially seemed sufficiently differentiated: In the course of the annotation work it became clear, however, that the functional spectrum of *ja* in the sub-corpus in question turned out to be more diverse and complex than in the 197 instances from a single dialogue analysed by Imo. Moreover, applying some categories to the data raised the question of their definitional solidity. The annotation scheme, therefore, had to be modified and extended considerably in order to be able to cope with the spectrum of usage of *ja* which was found. This resulted in a final functional spectrum of 12 categories:

Category	**Definition**
Response particle	An autonomous reaction to an utterance by the interlocutor. This can be a decision relating to a question posed by the interlocutor, but it can just as well be a wording that can be agreed upon
Response particle, phrase-final*	Phrase-final responsives often occur in situations where, during the turn that has just ended, the interlocutor formulated an interjection that is positively addressed with this phrase-final *ja*.
Modal particle	Suggests to the recipient that he or she should regard a fact mentioned in the same phrase as already known.
Discourse marker*	A turn or intonation phrase as initial discourse marker (discourse marker in the narrow sense, obligatory position before the prefield (cf. Imo 2013)). Likewise, *ja* in this function can be used to introduce a reply that factually contradicts the interlocutor's preceding statement but tries to bring this contradiction constructively into the discussion.
Hesitation signal*	Displays uncertainty, hesitation, planning of the following turn or segment. Some similarity to the usage of *ja* as a discourse marker. One characteristic here, however, is an often elongated and/or glottalized articulation with prosodic disintegration.

Category	**Definition**
Quotation marker*	Introduces the rendition of a quote, for example when a third person's statement is reproduced as verbatim speech in the course of a longer narrative. In this case, the *ja* is not part of the transmitted statement but is used to mark the beginning of this statement. In this respect, it is a specialized variant of the function as discourse marker.
Termination signal+	A signal from the speaker that they have said everything necessary on the immediately preceding topic and that they consider their turn to be finished. An implicit request to the interlocutor to continue with the dialogue.
Hearer signal	Supportive; indicates to the speaker that they should continue with their turn, the right to speak is explicitly not claimed by the user of the *ja* but continues to be attributed to the interlocutor; alternative to „mhm".
Self-affirmation	The speaker marks the scope of their immediately adjacent utterance as being actually true after having previously signalized uncertainty about it. Variant on "explicit inner discourse".
Newsmark signal*	Indicates either surprise at the interlocutor's immediately preceding statement and/or a request for a further explanation of that immediately preceding information.
Attention control*	Calling for sustained attention. Often used in the context of a more extensive turn to mark one particular aspect as being especially relevant.
Discourse structuring particles+	Uses that could not be assigned to any of the other categories.

Table 1: Designation and description of the pragmatic categories used in the research described. Categories marked with * have been modified in definition compared to Imo (2013), those marked with + have been completely new added.

In the corpus used, the particle *ja* was the most frequent lexical unit in absolute terms, at over 4%. This finding also applies – sometimes even more clearly – in other, very differently composed corpora of spontaneous German. More than three quarters (77%) of the 3,757 analysed instances of *ja* in the monomodal part of the GECO corpus were distributed over only four of the 12 categories (Table 1). Instances of *ja* with the function of a "classical" responsive (RP) represented the most frequent category at 26%. "Classical" discourse markers in a narrow sense of this term introducing a turn or phrase accounted for 20% of the data while listener signals, which can also be analysed as a variant on discourse markers, since they are semantically also strongly bleached, were in fourth place at 14%. The modalizing use of *ja* was the third most frequent category at 18%. This hierarchy shows both similarities and differences to the much smaller data sample in Imo (2013): Here, the four most frequent functions were responsive particle, hearer signal, modal particle and discourse marker/hesitation signal.

Contrasting with the affirmative use of *ja* is its semantically-functionally strongly divergent use in a discourse-organizing function. In the data at hand, this class was much more strongly represented and differentiated than that of the responsive particle. Although it cannot be ruled out that some of the differences observed here are due to the framework conditions of the corpus surveys, these are unquestionably not artefacts since it can be assumed that the speakers only used communication strategies that they were familiar with anyway.

For reasons of space, the factor of speaker-specific uses of *ja* could not be dealt with in depth here. It should be noted, however, that considerable differences can be observed. Apart from purely idiosyncratic factors, this could also be due to aspects affected by the course of the conversation. It is striking that hardly any speaker used the full range of functions in a dialogue. Instead, clear preferences for certain interactional strategies linked to *ja* could be identified for each speaker. Only two speakers used *ja* in at least 11 out of the 12 functions examined here in one of their dialogues.

References[1]

Heringer, Hans Jürgen (1988): Ja, ja, die Partikeln! Können wir Partikelbedeutungen prototypisch erfassen? In: Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung 41, pp. 730–754.

Imo, Wolfgang (2013): Sprache in Interaktion. Berlin: De Gruyter.

Meer, Doris (2007): „ja er redet nur MÜLL hier.“ – Funktionen von ‚ja‘ als Diskursmarker in Täglichen Talkshows. (= gidi Arbeitspapierreihe Nr. 11).

Schweitzer, Antje/Lewandowski, Natalie/Duran, Daniel/Dogil, Grzegorz (2015): Attention, please! Expanding the GECO database. In: The Scottish Consortium for ICPhS 2015 (ed.): Proc. 18th ICPhS, Paper 620. Glasgow, UK: the University of Glasgow. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0620.pdf (Stand: 14.1.2021).

Selting, Margret/Couper-Kuhlen, Elizabeth (2000): Argumente für die Entwicklung einer ‚interaktionalen Linguistik‘. In: Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion 1/2000, pp. 76–95.

Weidner, Beate (2015): Das funktionale Spektrum von *ja* im Gespräch – Ein Didaktisierungsvorschlag für den DaF-Unterricht. In: Imo, Wolfgang/Moraldo, Sandro M. (eds.): Interaktionale Sprache und ihre Didaktisierung im DaF-Unterricht. (= Stauffenburg Deutschdidaktik 4). Tübingen: Stauffenburg, pp. 165–195.

[1] In the referenced given here only those sources were listed that are explicitly mentioned in the abstract.