

Acoustic, Morphological, and Functional Aspects of “yeah/ja” in Dutch, English and German

Jürgen Trouvain¹ and Khiet P. Truong²

¹Phonetics, Saarland University, Saarbrücken, Germany

²Human Media Interaction, University of Twente, Enschede, The Netherlands

trouvain@coli.uni-saarland.de, k.p.truong@utwente.nl

Abstract

We explore different forms and functions of one of the most common feedback expressions in Dutch, English, and German, namely “yeah/ja” which is known for its multi-functionality and ambiguous usage in dialog. For example, it can be used as a yes-answer, or as a pure continuer, or as a way to show agreement. In addition, “yeah/ja” can be used in its single form, but it can also be combined with other particles, forming multi-word expressions, especially in Dutch and German. We have found substantial differences on the morpho-lexical level between the three related languages which enhances the ambiguous character of “yeah/ja”. An explorative analysis of the prosodic features of “yeah/ja” has shown that mainly a higher intensity is used to signal speaker incipency across the inspected languages.

Index Terms: feedback, yeah, ja, dialog act, prosody, cross-linguistic, speaker incipency

1. Introduction

One of the most typical and frequent feedback expressions in English is “yeah” (e.g. [1] [2] [3] [4] [5] [6]) which also has corresponding expressions in other languages, usually with a different spelling such as “ja” in Dutch and in German. Especially in Dutch and German, “ja” also frequently occurs in reduplicated forms such as “jaja” or in multi-word expressions such as “ja genau” in German or “nou ja” in Dutch. In addition, there is a huge diversity of possible meanings and functions of “yeah/ja” which is additionally enhanced by its morpho-lexical variability as explained above. This variability in meanings and functions may also affect the possible phonetic productions of “yeah/ja” (e.g., [7]). All together, these aspects make “yeah/ja” a highly ambiguous and complex feedback expression that is interesting to study from cross-linguistic, dialog-interactive, and phonetic point of views. The current study investigates the highly frequent feedback expression “yeah/ja” in conversational speech corpora of three languages (Dutch, English, German) with a special interest (i) in morpho-lexical variability and (ii) in prosodic differences between “yeah/ja” tokens showing speaker incipency and those showing passive recipency (i.e., the intention to commence speakership).

Phonetically “yeah/ja” is usually an opening diphthong, starting with a palatal glide and ending in the area between an open, unrounded and central vowel and an open-mid vowel. The phonetic make-up and hence the spelling of “yeah” and “ja” in Dutch, English, and German can be seen as standardized.

In addition to its morpho-lexical variability, the production of “yeah/ja” can also differ across languages. In Swedish for example, “yeah/ja” can occur in various reduplicated forms such “jaja” or “jajaja” similar to Dutch and German. However, the

airstream mechanism differs since in Swedish “ja” is often produced with an ingressive airstream [8].

It has been argued that multiple sayings of “ja” uttered in the same intonation phrase are not just intensifications of a single “ja” [9]. Additionally multiple sayings can bear different meanings depending on the intonation contour used (cf. [10] and [11] for German). Thus, the morpho-lexical variability, functional variability, and phonetic variability of “yeah/ja” are all related to and affect each other.

Jurafsky et al. [4] point out that “yeah/ja” is highly ambiguous in terms of function in dialog. “Yeah/ja” can be used in the backchannel [12] as a continuer and additionally to signal agreement with a yes-answer as a particular case, and it can be used to provide assessment. Although the multi-functionality of “yeah/ja” is acknowledged there is no generally accepted standard set of functions (or dialogue acts) of “yeah/ja” in dialogues. Table 1 lists four similar but only partially compatible approaches of labeling the multi-functionality of response tokens such as “yeah/ja” in English and German. It should be clear that these are just four out of several labeling schemes. Obviously there is no standard labeling scheme as the variability in labeling functions of response tokens Table 1 illustrates. It also illustrates the range of possible ambiguity of “yeah/ja”.

As shown in Table 1, it has been suggested for English that “yeah” apart from its function as continuer also signals a certain level of speaker incipency, i.e. starting a longer discourse unit with “yeah” ([1] [2]). In contrast to backchannel utterances featuring neutral nasal consonants (often transcribed as “m” or “hm”) “yeah” can indicate that the speaker is prepared to shift from recipency to incipency [1]. This pivotal mechanism makes the change from active listener to active speaker easier and thus conversations more fluent. In order to process this fluency in-time we would expect that speaker incipency is also prosodically marked beyond syntax. This could be done by a higher intensity at the turn beginning signalling the planning of a longer stretch of speech to follow (e.g. [13]).

The cross-linguistic aspect not only plays a role in the morpho-lexical and phonetic variability, it also comes into play when we are looking at the function of “yeah/ja” in dialog. In German for example, “ja” can have the lexical meaning of “yes” and it can be used as a modal particle signalling common ground (e.g. [16]). In contrast to other cross-linguistic studies on feedback signals which focused on their frequency of occurrence and the prosody of phrases preceding the feedback signal, such as Levow et al. [17], we concentrate on only one token which mostly, but not exclusively, is used as a feedback token. Unlike [17], where productions of Chinese, English and Spanish were investigated, we are dealing here with differences and parallels of closely related languages which in our case all be-

Table 1: Possible functions of response tokens taken from four labeling schemes.

Jurafsky et al. [4]	Benus et al. [5]	AMI [14]	Buschmeier et al. [15]	
continuer (backchannel)	backchannel	backchannel	“please continue”	
agreement	acknowledgement/ agreement	assessment	“I understand”	
assessment			“I agree”	
yes-answer			“I disagree”	
			attitude	
incipient speaker (incl. pivot/latching)	beginning discourse seg- ment		new discourse segment	
	pivot: acknowledgement + beginning disc. segment			
	ending discourse segment		ends discourse segment	
	acknowledgement + ending disc. segment			
	question		“What are you saying?”	
			“I do not understand”	
	stall/filler		stall	
	literal modifier			
	back from a task			
	cannot decide			unresolved
		inform		
		fragment		
		other		

long to the Western branch of the Germanic languages.

Summarizing, “yeah/ja” is a multi-faceted feedback expression. We aim to explore the morpho-lexical variability of “yeah/ja”, and possible different phonetic realizations of speaker incipient and passive recipient “yeah/ja” in Dutch, German, and English. Section 2 presents the data and our methods for analysis. The results are shown in Section 3 and we discuss our findings in Section 4.

2. Method

2.1. Data

For our analysis, we used three conversational speech corpora: the Lindenstrasse corpus [18] for German, the Diapix Lucid corpus [19] for English and the Dutch Spoken Corpus (CGN) [20] for Dutch. We expected to see annotations of “yeah/ja” which are possibly not consistent within one corpus and/or are not comparable between the different corpora (cp. [21]). For none of the corpora a clear function for “yeah/ja” was annotated. We decided to manually re-annotate selected functions (see below) and segmentation boundaries of the annotated tokens in question. Manual re-labeling was performed by both authors independent of each other. For this reason we restricted ourselves to 100 “yeah/ja” tokens per corpus. For each corpus, 3 female-female and 3 male-male conversations were randomly selected. From these 6 conversations per corpus, 100 yeah-tokens (50 from female-female and 50 from male-male conversations) were selected in a random manner. Among the selected tokens we concentrated on single and turn-initial tokens of “yeah/ja” and thus excluded those “yeah/ja” where it occurs in combination, e.g. German “naja” or “jaja”, or where it occurs in a medial or final turn position.

2.2. Labeling

After reviewing previous work on the ambiguous functions of “yeah/ja” (e.g., [4]) and after several attempts to label these, we decided to focus on the labeling of speaker incipency (SI) vs. passive recipency (PR) according to the set of categories shown in Table 2. The operationalization of speaker incipency can take several forms. In Drummond and Hopper [2], a speaker incipient “yeah/ja” was initially defined as a “yeah/ja” token that is immediately (< 200 ms) followed by same-speaker speech. In Truong and Heylen [6], speaker incipency was (automatically) defined as the number of ‘conversational states’ that has passed until the current speaker starts a new full turn. Their definition takes into account the preparedness aspect of speaker incipency (see Jefferson [1]). For the current study, we use a similar operationalization of speaker incipency as described in [2]. Since it is imaginable that “yeah/ja” can be followed by same-speaker speech without constituting a bid for speakership, the distinction between a minimal (label A2) and full turn (label B) was made (following [2]). Although A2 could be considered to have a higher (gradual) level of speaker incipency, we make a binary distinction and consider both A1 and A2 as forms of passive recipency and B as a form of speaker incipency.

Table 2: Labeling of “yeah/ja”

Label	SI/PR	Description: “Yeah/ja” is ...
A1	PR	freestanding
A2	PR	the first part of a minimal turn
B	SI	the first part of a full turn
C	N/A	none of the above (for example, not in turn-initial position)

In addition to labeling speaker incipency, we also marked whether “yeah/ja” was used in a single form or as a multi-word

expression. Only single tokens of “yeah/ja” were taken for subsequent acoustic analysis.

2.3. Acoustic analysis

For each token we automatically measured its duration, mean intensity, mean F0, and F0 range ($F0_{max} - F0_{min}$) using Praat [22]. All measurements were transformed to z-scores ($z = (x - \mu) / \sigma$) per speaker where the mean (μ) and standard deviation (σ) were taken over all single “yeah/ja” tokens uttered by that speaker. The main expectation is that speaker incipient “yeah/ja” have a higher intensity than passive recipient “yeah/ja” (cf. [13]).

3. Results

3.1. Morpho-lexical variability of “yeah/ja”

We counted each occurrence of “yeah/ja” and looked at whether it was used as a single token or in combination with the same or other ‘particles’ creating new multi-word expressions. Figure 1 shows that there are substantial differences in the morpho-lexical variability across the three languages. Although the single form of “yeah/ja” is the predominant form in all languages it is used only 65% in Dutch in contrast to 89% in English (which is in line with the numbers in [4] for English).

There are also differences regarding the possibility of combinatory forms including multiple sayings and other multi-word expressions such as “ja genau” (=“yeah exactly”) in German and “uh ja” and “oh ja” in Dutch. For Dutch we count more than 60 combinatory forms compared to around 20 combinations in English and in German.

Summarising it can be noted that for Dutch and German there is a substantial degree of morpho-lexical variability, i.e. the usage of new combinatory forms such as “jaja” or “nou ja”/ “naja”. This finding was not expected when we take English as the baseline. This large lexical variability may in turn increase the variability in function: some of these new multi-word expressions are more used as idiomatic expressions such as “jaja” or “nouja” in Dutch, and some may carry an affective meaning.

3.2. Functional variability of “yeah/ja”

As stated above “yeah/ja” in Dutch and German are not exclusively used as a feedback utterance, be it a simple continuer or be it a continuer including some assessment or further additional functions. In the German data “ja” in the usage of a modal particle rather than a discourse particle lies around 5%. Additionally there were several unclear cases of “ja” where it was used in indirect speech or as a self-comment. Likewise, among the Dutch combinations “jaja”, “uh ja”, “oh ja”, “nou ja”, or “maar ja” not all are used as feedback signals but as fillers.

Thus, the ambiguity of “yeah/ja” illustrated in [4] for English is enhanced by further meanings in Dutch and in German which go beyond ‘pure’ feedback.

3.3. Acoustic variability

The acoustic analysis of the single tokens among the selection of 100 “yeah/ja” per corpus revealed that for all three languages intensity plays an important role for distinguishing speaker incipency in contrast to passive recipiency, see Table 3. Although there was a tendency for all three languages to have a higher mean fundamental frequency and a narrower F0 range for speaker incipency the differences between recipiency and incipency found for fundamental frequency (mean and range)

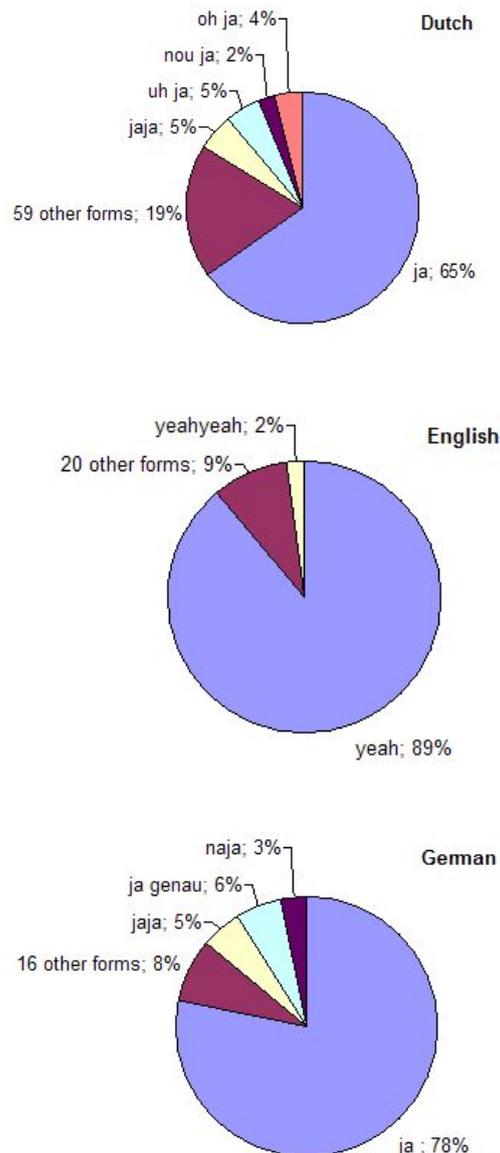


Figure 1: Wording of “yeah/ja” in Dutch, English, and German.

were statistically not significant. The duration of tokens of “yeah/ja” was longer when used for recipiency. However, this difference was only significant for German.

As expected speaker incipency is acoustically signalled mainly by intensity (see also [13]) but in general, has no prosodically marked forms.

4. Discussion and conclusions

In contrast to English, “yeah/ja” shows a substantial variability in their morpho-lexical forms in German and particularly in Dutch. In English, “yeah/ja” is mostly used in its single form while in Dutch and German, it is more often used in multi-word expressions that may have different dialog functions. It makes it clear that we have to acknowledge that a common feedback expression such as “yeah/ja” can have different functions in dif-

Table 3: Averaged acoustic measurements given in z-scores. Significant differences with p-values below .05, tested with t-tests, are marked with an asterisk.

	Feature	A1+A2	B	signific.
Dutch	Dur	-0.02	-0.37	
	F0	-0.26	0.05	
	F0range	-0.16	-0.18	
	Intens	-0.08	0.46	*
English	Dur	0.09	0.00	
	F0	0.40	0.55	
	F0range	0.41	0.22	
	Intens	-0.10	0.56	*
German	Dur	0.07	-0.37	*
	F0	-0.20	0.17	
	F0range	-0.07	-0.28	
	Intens	-0.13	0.66	*

ferent languages even if these languages are closely related such as the three West-Germanic languages analysed here. Given the growing amount of cross-cultural human-human and human-machine communication, more attention should be paid to these cross-linguistic aspects of feedback expression. Future work has to show for instance how far multi-word expressions based on "ja" differ in function and meaning to single tokens of "ja", especially regarding their intonation structure.

What else is clear, is that there is a large variety of dialog act labeling approaches to feedback expressions which also reflects the multi-functionality of feedback expressions. For future research, it would be interesting to perform a more thorough meta-analysis of possible functions and meanings of "yeah/ja" and other feedback expressions. It would also be an asset to work with speech data of various languages but elicited via the same task as performed by Levow et al. [17].

Finally, although we did not find a *clear* prosodically marked form for speaker incipient "yeah/ja", prosodic measurements, as illustrated in previous work, *can* and *should* be used to help disambiguate (other) dialog act functions of feedback expressions. In this connection, prosody should also include features of voice quality such as creaky voice which has been shown to signal passive reciprocity [23]. To explore the fine phonetic detail of these functions across languages remains a further task for the future, as well as the automatic processing of these functions which will help make human-machine interactions more fluent.

5. Acknowledgements

Thanks to Eva Lasarczyk and one anonymous reviewer for their feedback. This work was partly supported by the EU 7th Framework Programme (FP7/2007-2013) under grant agreement no. 231287 (SSPNet) and the UT Aspasia Fund.

6. References

- [1] G. Jefferson, "Notes on a systematic deployment of the acknowledgement tokens 'Yeah' and 'Mm hm'," *Tilburg Papers in Language and Literature*, 1984.
- [2] K. Drummond and R. Hopper, "Back channels revisited: acknowledgement tokens and speakership incipency," *Research on Language and Social Interaction*, vol. 26, pp. 157–177, 1993.
- [3] —, "Some uses of *yeah*," *Research on Language and Social Interaction*, vol. 26, pp. 203–212, 1993.
- [4] D. Jurafsky, E. Shriberg, B. Fox, and T. Curl, "Lexical, prosodic, and syntactic cues for dialog acts," in *Proceedings of ACL/COLING Workshop on Discourse Relations and Discourse Markers*, 1998, pp. 114–120.
- [5] S. Benus, A. Gravano, and J. Hirschberg, "The prosody of backchannels in american english," in *Proceedings of the 16th International Congress of the Phonetic Sciences (ICPhS)*, 2007, pp. 1065–1068.
- [6] K. P. Truong and D. Heylen, "Disambiguating the functions of conversational sounds with prosody: the case of yeah," in *Proceedings of Interspeech*, 2010, pp. 2554–2557.
- [7] T. Stockmeier, S. Kopp, and D. Gibbon, "Synthesis of prosodic attitudinal variants in german backchannel 'ja'," in *Proceedings of Interspeech*, 2007, pp. 1290–1293.
- [8] R. Eklund, "Pulmonic ingressive phonation: Diachronic and synchronic characteristics, distribution and function in animal and human sound production and in human speech," *Journal of the International Phonetic Association*, vol. 38, pp. 235–325, 2008.
- [9] T. Stivers, "'No no no' and other types of multiple sayings in social interaction," *Human Communication Research*, vol. 30, pp. 260–293, 2004.
- [10] A. Golato and Z. Fagyal, "Comparing single and double sayings of the German response token 'ja' and the role of prosody: A conversation analytic perspective," *Research on Language and Social Interaction*, vol. 41, pp. 241–270, 2008.
- [11] D. Barth-Weingarten, "Response tokens in interaction – prosody, phonetics and a visual aspect of German JAJA," *Gesprächsforschung*, vol. 12, pp. 301–370, 2011.
- [12] V. H. Yngve, "On getting a word in edgewise," in *Papers from the Sixth Regional Meeting of Chicago Linguistic Society*. Chicago Linguistic Society, 1970, pp. 567–577.
- [13] A. Hjalmarsson, "The vocal intensity of turn-initial cue phrases and filled pauses in dialogue," in *Proceedings of SIGdial*, 2010, pp. 225–228.
- [14] J. Carletta, "Unleashing the killer corpus: experiences in creating the multi-everything AMI meeting corpus," *Language Resources and Evaluation*, vol. 41, pp. 181–190, 2007.
- [15] H. Buschmeier, Z. Malisz, M. Włodarczyk, S. Kopp, and P. Wagner, "'are you sure you're paying attention?' - 'uh-huh' communicating understanding as a marker of attentiveness."
- [16] E. Karagjosova, "Modal particles and the common ground: meaning and functions of German 'ja', 'doch', 'eben'/'halt' and 'auch'," in *Perspectives on Dialogue in the New Millennium*, P. Kühnlein, H. Rieser, and H. Zeevat, Eds. Amsterdam: John Benjamins, 2003, pp. 2–11.
- [17] G.-A. Levow, S. Duncan, and E. King, "Cross-cultural investigation of prosody in verbal feedback in interactional rapport," in *Proceedings of Interspeech*, 2010, pp. 286–289.
- [18] IPDS, *Video Task Scenario: LINDENSTRASSE The Kiel Corpus of Spontaneous Speech, Volume 4, DVD*. Institut für Phonetik und Digitale Sprachsignalverarbeitung Universität Kiel, 2006.
- [19] R. Baker and V. Hazan, "DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs," *Behavior Research Methods*, vol. 43, pp. 761–770, 2011.
- [20] N. Oostdijk, "The Spoken Dutch Corpus. Overview and first evaluation," in *Proceedings of the International Conference on Language Resources and Evaluation (LREC2000)*, 2000, pp. 887–894.
- [21] J. Trouvain and K. P. Truong, "Comparing non-verbal vocalisations in conversational speech corpora," in *Proceedings of the 4th International Workshop on Corpora for Research on Emotion Sentiment & Social Signals*, 2012, pp. 36–39.
- [22] P. Boersma and D. Weenink, "Praat, a system for doing phonetics by computer," *Glott International*, vol. 5, no. 9/10, pp. 341–345, 2001.
- [23] T. Grivicic and C. Nilep, "When phonation matters: The use and function of 'yeah' and creaky voice," *Colorado Research in Linguistics*, vol. 17, 2004.