

Euh as cue for speaker confidence and word searching in human spoken answers in French

Anne Garcia-Fernandez^{1,2}, Ioana Vasilescu², Sophie Rosset²

¹University Paris-Sud, Orsay, F-91405

²LIMSI-CNRS, Orsay, F-91403

annegf, vasilescu, rosset@limsi.fr

Abstract

This paper deals with the contextual analysis of the vocalic hesitation *euh* in French in a corpus of human elicited answers. Through the analysis of the contextual combinatorial patterns, the new information introductory role of this vocalic hesitation is investigated. Observations supports trends noticed in other languages and suggest potential optimization for question answering automatic systems.

Index Terms: vocalic hesitation, feeling of knowing, rephrasing, interaction management, QA systems.

1. Introduction

Spoken disfluencies have been widely investigated in computational linguistics. Approaches focused on the automatic classification and identification of the various disfluent phenomena, their acoustic/prosodic cross-language patterns or the impact on the automatic performance in various spoken language processing frameworks. The initial position adopted in processing such spoken events consisted in an in-depth description to efficiently *clean* the speech, that is the lexical level of the oral message [1]. However, more recently it has been argued that most spoken disfluencies are not problems in speaking but *solutions* to problems in speaking [2].

The current study is concerned with automatic question/answering (QA) dialog systems: in this framework making an effective usage of such disfluent devices is an actual objective which aims to improve and make more natural the interaction. In particular, detecting and efficiently exploiting the various disfluent-like items produced by the human interlocutor and delivering answers with appropriate corresponding items is still a going concern. The stress is put here on the automatic *answer* modeling with the purpose of (i) providing more accurate answers in terms of grammatical structure and information content proper, but also of (ii) systems ability to evaluate their *confidence* in the answer. The latter is highly concerned with the disfluency level: psycholinguistic studies on both the speakers and the listeners feelings about the answer delivered/received within a spoken interaction have shown that paralinguistic features of utterances such as pauses, intonation, and interjections play a compelling role [3, 4].¹ In line with these findings, computational studies focused on the speakers feelings monitoring

¹The meta-cognitive state of the speaker displayed in an answer has been associated with the *feeling of knowing* (FOK), that is speakers' ability to accurately monitor their knowledge about the information delivered. Listeners in turn are able to infer about such states by making smart usage of paralinguistic cues (i.e. *feeling of another's knowing* - FOAK) [4].

(such as uncertainty) with the aim of improving system performance [5, 6].

We focus here on the distribution of combinatorial patterns of the vocalic hesitation *euh* in French in a corpus of human elicited answers. Our working hypothesis is that *euh* may point as well on the answer structure by signaling the new provided information. The longer term objective is to model such items in a natural language generation answers framework.

2. Data description and annotation methodology

2.1. The MACAQ corpus

MACAQ corpus (Multi-Annotated Corpus of Answers to Questions) [7] is composed of answers provided by humans. They have been elicited through a number of questions similar to the ones usually addressed to an automatic system operating in open domain. Questions with various controlled linguistic forms have been manually generated beforehand to illustrate a potential range of interrogative structures a human may address to an automatic system. Then questions have been addressed to humans who provided answers.² MACAQ has been built with the aim of portraying typical answers to serve as model for further automatic answers generation in natural language.

Duration	1h10
# Answers	1,044
# Lexical items	6,472
# Lexical types	657
# Utter. w. <i>euh</i>	244 (23.4%)

Table 1: General description of the MACAQ corpus.

The table above details the content of the corpus: more than 23% of the utterances contains at least a vocalic hesitation, that is, according to [4], about one in four speakers display some doubt or at least need of longer time to formulate the answer.

2.2. Annotation methodology

The adopted manual annotation strategy is described in [7]. An annotation of the elements of the question reused in the answer was done. Those elements annotated with the tag **QUE** correspond to an *old* information. The information-answering the question (as defined in [7]) which correspond to a *new* information was annotated using the tag **ANS**. A difference has been made between the expected answer and some additional

²Both spoken and written modality have been elicited but we consider only the speech corpus in this paper.

developments which detail the answer: **ADD** tag corresponds to such completions. Furthermore, pragmatic and question markers (**PM**) as well as discourse markers (**DM**) were annotated. Figure 1 shows examples of annotation.

<DM> alors </DM> <PM> je dirais qu' </PM> <QUE> un bébé </QUE> <QUE> mesure </QUE> euh <ANS> 50 centimètres </ANS>.

(so I'd say a baby measures umm 50 centimeters)

euh <REP> 28 jours </REP> <PM> ou </PM> <REP> 29 </REP> <ADD> pendant les annes bissextiles </ADD> (umm 28 days or 29 in leap years)

Figure 1: Example of annotated answers

The table below depicts the percentage of different annotations tags according to adopted classes of spoken items. Relevant information chunks are predominant in particular when considering ANS and QUE together (45.8%). *Euh* occurrences are counted according to their position in the utterance: the great majority of the vocalic hesitations occurs then as utterance-opening elements (**INIT**ial position). Utterance internal vocalic hesitations may be encountered as well (**MED**ial position), however final position (**END**ing position) seems to generally avoid such spoken items.

Tag	#occ.	%
ANS	340	27.6
QUE	225	18.2
PM	189	15.2
EUH_INIT	172	13.9
ADD	149	11.8
EUH_MID	94	7.6
DM	68	5.5
EUH_END	4	0.3
Total	1218	100

Table 2: Distribution of annotation tags in the MACAQ corpus.

3. Contextual analysis

The table 3 shows *euh* bigrams combinatorial patterns according to position in the utterance. The position has been selected as salient parameter: observing elements on the left (-1) and on the right (+1), it may indicate the role of the *euh* disfluency. From the *vocalic hesitation as rephrasing work indicator* perspective, the position signals the utterance level concerned with rephrasing: utterance (INIT) or utterance-internal region (MID).

The main pattern is that *euh* occurs prior to the new information (ANS) wherever placed within the utterance (INIT or MID positions). *Euh* indicates both the *place* of the new (relevant) information (the ANSwer) and the speakers' *feeling* about the delivered information (as shown in [4]).

The medial position exhibits *euh* occurrences flanked by QUE, PM or ANS (left, -1), that is old information and ANS (right, +1) suggesting that speakers' potential uncertainty or rephrasing intention concern the new information.

	AFTER				BEFORE		
	Tag	#occ.	%		Tag	#occ.	%
INIT(+1)	ANS	126	73.3	MID(-1)	QUE	34	36.2
	QUE	19	11.0		PM	25	26.6
	PM	17	9.9		ANS	18	19.1
	DM	7	4.0		DM	7	9.5
	Other	3	1.8		Other	8	8.6
MID(+1)	ANS	42	44.7	END(-1)	ANS	2	50.0
	PM	20	21.3		PM	1	25.0
	QUE	14	14.9		ADD	1	25.0
	ADD	12	12.8				
	Other	6	6.3				

Table 3: 2-gram distribution of *euh* occurrences according to position in the utterance.

4. Conclusions

This work focused on the *euh* contextual analysis in a corpus of elicited natural human answers in French. The work hypothesis concerned: the vocalic hesitation as indicator of (i) rephrasing work and (ii) speakers' feeling about the new information delivered in the answer, that is his/her potential (un)confidence. Current data indicate that *euh* occurs prior to the new information and has a rephrasing role. This analysis suggests that using *euh* in NLG component of interactive QA systems could increase the user confidence in the system or at least give him/her enough information to understand the confidence the system has in its own answer.

5. Acknowledgements

This work has been partially financed by OSEO under the Quaero program.

6. References

- [1] Adda-Decker, M., Habert, B., Barras, C., Adda G., Boula de Mareüil, Ph., Paroubek, P., "A disfluency study for cleaning spontaneous speech automatic transcripts and improving speech language models", Disfluency In Spontaneous Speech (DISS) Workshop, Gothenburg, Sweden, 2003.
- [2] Clark, H. H., "Speaking in time", in Speech Communication, 36/1-2, 5-13, 2002.
- [3] Smith, V. L., Clark, H. H., "On the course of answering questions", in Journal of Memory and Language, 32, 25-38, 1993.
- [4] Brennan, S. E., Williams, M., The Feeling of Another's Knowing: Prosody and Filled Pauses as Cues to Listeners about the Metacognitive States of Speakers, in Journal of Memory and Language, Elsevier, 34/3, 1995.
- [5] Litman, D., Forbes-Riley, K., Spoken tutorial dialogue and the feeling of another's knowing, in Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue, ACL, 286-289, 2009.
- [6] Tsukahara, W., Ward, N., "Responding to subtle, fleeting changes in the user's internal state", in Proceedings of the SIGCHI conference on Human factors in computing systems, ACM, 77-84, 2001.
- [7] Garcia-Fernandez, A., Rosset, S., Vilnat, A., "MACAQ: A Multi Annotated Corpus to study how we adapt Answers to various Questions", LREC 2010, Valletta, Malta, 2010.