Proceedings of DiSS-LPSS Joint Workshop 2010,
25-26 September 2010, Tokyo, Japan, pp. 111-114.

On the functions of the vocalic hesitation *euh* in interactive man-machine question answering dialogs in French

Ioana Vasilescu, Sophie Rosset, Martine Adda-Decker

LIMSI-CNRS F-91403 Orsay Cedex
ioana, rosset, madda@limsi.fr

Abstract

This paper deals with the functions of the French vocalic hesitation *euh* in interactive speech of man-machine question answering dialogs. The present analysis suggests that the vocalic hesitation *euh* may carry various properties in speech, both *disfluent* signaling the speakers' efforts to put the intended message under production into appropriate words, and *fluent*, as markers of discourse structure. Moreover, *euh* seems to play a role in bracketing lexical units, pointing to the informative content within an utterance. This bracketing may favour intelligibility or *decoding fluency* on the listener's side. The potential contribution of the vocalic hesitation *euh* to lexical information bracketing is investigated with the goal of improved information processing by QA systems. Future objectives include a smarter interaction capacity by an appropriate usage of such *euh* items.
Index Terms: disfluency, fluency, vocalic hesitation, French, discourse markers, Q/A, dialog corpus

1. Introduction

Automatic question/answering (QA) dialog systems deal with real-time, social convention-based interactions. In this framework, two different levels of spoken utterances need to be processed: discourse structure and lexical information proper. Automatic extraction of the truly informative words out of the utterances may benefit from a smart handling of spontaneous speech specific items such as *vocalic hesitations*.

Unprepared spoken utterances and more generally oral conversations include a variety of speech events which do not directly contribute to the final message under elaboration as conveyed at the lexical level. Such speech phenomena are of two kinds: *(i)* truly *disfluent events* such as splutters, slips of the tongue, and other hesitations[1] of various types; and *(ii)* the so-called "little words", e.g. *enfin, bon, ben, eh bien* ("well"), *donc, alors* ("so"), typically discourse markers. The latter relate to discourse level: they mark the discourse structure and act as "instructions" (from the speaker to the listener) on how to process the unit they frame [1]. However, analysing and classifying such elements is not straightforward as inter-class boundaries are permissive and taxonomy is context and/or corpus-dependent. For instance, vocalic hesitations when associated with spluttered speech regions may be categorized as disfluent, whereas within a dialog and in particular in utterance-initial position the same vocal item may play a role in interaction management and structuring the discourse over the time. The vocalic hesitation then behaves as a classical discourse marker [2].

Vocalic hesitations have been initially counted as disfluency marks within an analysis framework which stressed the opposition between *fluent* and *disfluent* speech, the latter being characterized by various non-lexical and irregular spoken events [4]. Despite their intrinsic *disfluent* nature, vocalic hesitations also carry a widely recognized pragmatic function: to mark the speaker's intention to keep the floor during his effort of building the lexical message. More recent works underlined the lexical status of the American English vocalic hesitations *uh* and *um*, and confirmed previous assumptions on their dialogic role [5]. The vocalic hesitations are included in a broad class of interjections and as conventional words of English they signal a delay in speaking. Even though their lexical status as interjections has been reconsidered since the seminal work of [5], other studies validated the basic meaning of *uh* and *um* as words which indicate various delays in spoken English [6]. Similar discursive behavior and interaction with the neighbouring lexical region have been noticed in other languages [2, 3], whereas studies on French pointed out some cross-language shared properties of the vocalic hesitation *euh* [7].

Finally, during the last decades, a significant body of work has been dedicated to the vocalic hesitations in the computational area [8, 9]. Approaches focused on their acoustic/prosodic and linguistic discriminant features, on the automatic identification of such events or on their impact on automatic speech recognition performance.

The present work is exploratory and deals with interactive speech in the framework of man-machine QA dialogs. In this framework, the behavior of the French vocalic hesitation *euh* is investigated with the particular aim to point out potential discursive functions in speech. The objective is to investigate contextual discourse marker-like properties with a particular stress on the propensity to initiate rephrasing and to bracket units of speech. Such properties may contribute to improve information processing by Q/A systems. A further objective includes their appropriate generation within oral responses by QA systems along with other discourse markers to ease natural interaction.

The following section presents the working hypothesis. The analyzed corpus, the proposed annotation scheme and resulting *euh* distributions are described in section 3. Section 4 deals with the contextual functions and the *disfluent* vs. *fluent* behavior of *euh* as revealed by the contextual combinatorial patterns. Observations and conclusions are summarized in section 5.

2. Working hypothesis

The motivation of the present study is to check whether French vocalic hesitations may carry different functions in spoken in-

[1] *Hesitation* is a generic term in corpus linguistics: various non-lexical events, occuring in oral (non-prepared) utterances, are included in a broad class of hesitation marks, from filled or silent pauses to complex rephrasing phenomena such as repetitions or restarts. In French, vocalic hesitation refers to the filled pause *euh*.

raw	*euh Allen A deux L E N* ("*uh Allen A two L E N*")
tagged	<EUH> *euh* </EUH> <RIC> *Allen* </RIC> <RIC> *A deux L E N* </RIC>
tag seq.	EUH RIC RIC
raw	*donc qui a réalisé* ("*so who directed" (...)*)
tagged	<CDM> *donc* </CDM> <PM> *qui* </PM> <RIC> *a réalisé* </RIC> (...)
tag seq.	CDM PM RIC
raw	*euh j'ai pas compris ce que vous avez dit* (*uh I did not understand what you said*)
tagged	<EUH> *euh* </EUH> <PM> *j'ai pas compris ce que vous avez dit* </PM>
tag seq.	EUH PM
raw	*des températures euh maximum* ("*maximum uh temperatures*")
tagged	... <RIC$_{par}$> *des températures* </RIC$_{par}$> <EUH> *euh* </EUH> <RIC$_{par}$> *maximum* </RIC$_{par}$> ...
tag seq.	RIC$_{par}$ EUH RIC$_{par}$

Table 1: Examples of RITEL user turns: `raw` and `tagged` transcripts, as well as corresponding `tag` sequence patterns.

teractions. Our working hypothesis is that beyond the *disfluent euh* signaling the speakers' lexical formulation trouble, *euh* may also carry other functions, contributing to discourse and/or lexical information structures. These structuring functions are considered to contribute to the overall fluency of the dialog. Hence, the corresponding *euh*s are termed *fluent*.

Concerning our previous work on the role of some classical discourse markers (such as *alors*, "so"; *bon, ben*, "well" etc.) in interaction in French man-machine dialogs, we showed that vocalic hesitations, along with the mentioned discourse markers, may bear rephrasing functions and in particular may highlight local (utterance-internal) word searching [10]. Furthermore, *euh* often occurred to bracket informative speech chunks.

The current work is then dedicated to the analysis of the contextual combination patterns of *euh* to highlight its different properties, *fluent* and *disfluent*, in speech. A better knowledge of *euh* realisations and their functions will contribute to improve spontaneous speech processing in question/answering systems. A long term aim of this work is then to make automatic use of the bracketing role of *euh*, i.e. as an indicator of spoken regions of interest, subject to rephrasing.

3. Data, annotation and processing

3.1. Data description

We make use of the **RITEL** (Recherche d'Information par TELéphone, "Information retrieval by phone") corpus. RITEL is composed of spontaneous utterances of humans who question an open-domain automatic dialog system. The RITEL system aims at providing userfriendly access to open-domain information via spoken interaction [11, 12]. This system integrates a spoken language dialog system and an open-domain information retrieval system in order to enable human users to ask general question and to refine their search for information interactively. Data have been collected during different evaluations of the RITEL system. The RITEL corpus contains 652 dialogues with more than 6 hours of user turns. There are 6 720 user turns comprising a total of 71 089 lexical items (tokens), with a total of 3 434 distinct lexical words (types). Data are summarized in Table 2.

Total user duration	6h40
# Dialogs	652
# Lexical tokens	71,089
# Lexical types	3,434
# User turns	6,720
# U. turns w. CDM + *euh*	1,718 (25.6%)

Table 2: General description of the RITEL corpus.

3.2. Annotation

The RITEL corpus has benefited from a preliminary annotation and analysis of rephrasing strategies involving classical discourse markers and the vocalic hesitation *euh* [10]. In the present study, we propose a new annotation scheme aiming at *(i)* highlighting both (*fluent* and *disfluent* properties of the vocalic hesitation in human spoken utterances and *(ii)* highlighting speech regions of interest, termed *relevant information chunks* and the potential bracketing or *framing role* of the vocalic hesitation.

The adopted strategy is based on semantic and pragmatic considerations and aims at dividing utterances according to two message levels: the lexical level proper and the discourse structure. The proposed annotation scheme uses four classes:

RIC *relevant information chunks*, *named-entity*-like words and phrases carrying salient information[2];

PM other words or phrases mainly pragmatic and question markers;

CDM classical discourse markers: in our data classical discourse markers correspond to *bon, ben alors, bah, ah, enfin, mais enfin/m'enfin*;

EUH the vocalic hesitation.

After the annotation process, the entire corpus is partitionned according to these classes. The former (**PM** and **RIC**) classes correspond to the lexical information level to be processed by the QA system. The latter two include items which play a role in structuring the discourse setting over time (CDMs and potentially *euh*) or in signaling disfluent speech regions (*euh*).

For (**RIC** and **PM**) two sub-categories were defined: *autonomous* and *partial*. By default, tags are *autonomous*. *Partial* tags were introduced to process: *(i)* complex chunks assembling words which belong to both **RIC** and **PM** classes (e.g. *quel acteur* "which actor" consists in a *partial* **PM** and a *partial* **RIC**) and *(ii)* **RIC** or **PM** chunks splitted by *euh* (in which

[2]**RIC** correspond to pertinent information chunks as described in [11]. Current annotation has been performed by one French native annotator.

case the two parts are annotated as *partial*). *Partial* tags may also indicate some *disfluent* use of the vocalic hesitation, e.g. *euh* flanked by *partial* **RIC** and/or **PM** may be characterized as disfluent. Table 1 gives annotation examples.

3.3. *Euh* distribution

Among the user turns, 25.6% exhibit at least one extra-lexical level item (see Table 2), that is a classical discourse marker or a vocalic hesitation. Vocalic hesitations prevail: they represent almost 3% of the transcribed tokens (only 0.7% for CDM) and they can be found in about 85% of the user turns containing such items. To check whether these *euh* items were more frequent at the beginning or at the end of the dialogs, Figure 1 plots the average number of *euh* per turn rank. Most dialogs remain relatively short: only 30% (resp.10%) of the dialogs have more than 10 (resp. 20) turns. Figure 1 shows the distribution of the vocalic hesitation across dialogs, i.e. as a function of turn rank k, $k = 1..K$. The curve gives the average number of *euh* items per turn rank k (normalised by the total number of turns of rank k). One may hypothesize that the more the speakers negociate with the automatic system (i.e. the turn rank increases) the more they produce hesitations. However, this pattern does not apply to the current data: higher ranks correlate with fewer hesitations (except ranks below 5 and above 32). Speakers seem to learn how to efficiently interact with the system in longer dialogs.

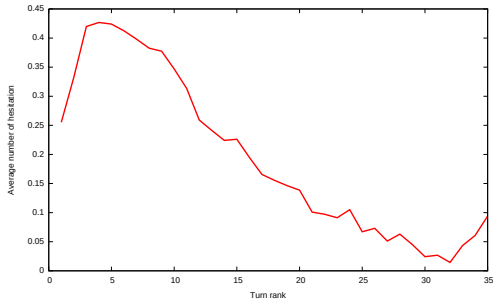

Figure 1: Avrg. number of *euh* as a function of dialog turn rank.

4. Combinatorial patterns

This section aims at describing the constraints which govern the cooccurence of the vocalic hesitation with the dialog tags (**RIC** and **PM**, *autonomous* and *partial*). The underlying hypothesis is that the current exploratory analysis may inform us on its changing nature whithin the spoken utterances, that is on its *disfluent* or *fluent* nature. The latter is particularly interesting in the framework of automatic QA systems and concerned with the potential bracketing functions of the vocalic hesitation.

The position of the vocalic hesitation within the speaker turn has been selected as a relevant factor: initial (**init**), medial (**mid**) and final (**end**). Table 3 summarizes the tag frequencies in the RITEL subset including **CDM** and **EUH** concerned with the combinatorial analysis.

Table 3 shows that more than 20% of the tags correspond to vocalic hesitations and that they are relatively homogeneously distributed with respect to the different utterance positions (**init**, **mid**, **end**). The overall *euh* rate tends to decrease from 9% in initial to 5.1% in final positions. It is interesting to observe that

Tag type	#occ.	%
RIC	2327	24.3
PM_{par}	2045	21.4
EUH	2032	21.2
RIC_{par}	1504	15.7
PM	1160	12.1
CDM	502	5.2
Total	9570	100.0
EUH_{init}	865	9.0
EUH_{mid}	679	7.1
EUH_{end}	488	5.1
CDM_{mid}	267	2.8
CDM_{init}	227	2.4
CDM_{end}	8	0.1

Table 3: Distribution of tags in the RITEL sub-corpus with CDM + *euh* (25.6%); *par*=partial. In the two lower parts, EUH and CDM tags are subdivided according to initial, medial or final positions within a turn.

*euh*s are more frequent in turn-initial than internal positions, although all but the 2 boundary positions are internal. Aside from indicators of speakers' hesitation in lexical selection at both utterance and phrase/word levels (i.e. *disfluent* behavior), one may hypothesize that *euh* occurence correlates with other potential (*fluent*-like) properties. In particular, two potential functions are considered here: turn-starters and local (word/phrase level) rephrasing. In a previous analysis [10], we associated the propensity of the vocalic hesitation to close the utterance to a speaker affect carrying role: *euh* closes the (failed) interaction and express a potential embarrassement. The following paragraphs will focus on the most frequent combinatorial patterns of the vocalic hesitation *euh* with *autonomous* and *partial* **PM** and **RIC**.

4.1. Bigram combinatorial patterns

bigram type	Tag **X**	#occ.	%
<**EUH-X**	PM_{par}	395	45.7
	PM	285	32.9
	RIC	80	9.2
	CDM_{mid}	44	5.1
	Others	61	7.1
-EUH-X	PM_{par}	211	31.1
	RIC	185	27.2
	RIC_{par}	183	27.0
	PM	49	7.2
	Others	51	7.5
X-EUH-	PM_{par}	249	36.7
	RIC_{par}	183	27.0
	RIC	99	14.6
	PM	79	11.6
	Others	69	10.2

Table 4: *euh* bigram patterns in the RITEL sub-corpus (*par*=partial). The first block corresponds to turn *initial* bigrams (< : turn start). The 2nd and 3rd blocks show turn internal bigrams.

Table 4 summarizes the bigram combinatorial patterns of the vocalic hesitation in the RITEL sub-corpus. Three factors are considered in the analysis: the contextual class, its position with respect to the vocalic hesitation (right/left), and the posi-

tion of the vocalic hesitation within the speaker turn. We consider only initial and medial positions in the following. Concerning the initial position, **EUH** cooccurs with **PM** (either autonomous (45.7%) or partial (32.9%)) in almost 80%. In the role of *turn-starters* the vocalic hesitation tends to be followed by **PM** structures which set the questions. One may associate this pattern with both *fluent* and *disfluent*-like features of *euh*: *fluent* as they act as turn-initiator devices[3], and *disfluent*, as speakers by using *euh* express their need to phrase or rephrase the utterance and hesitate in their lexical selection process. Medial vocalic hesitations introduce or occur most often (54.2%) prior to **RIC**[4]. Medial *euh* followed by **RIC** may then be associated either with *rephrasing*: it acts as a cue for word searching, or *bracketing*, i.e. pointing to the salient information of the utterance. The third block of Table 4 shows that medial *euh* more often follows **PM** tags (47.3%) than **RIC** tags (41.6%).

4.2. utterance-internal 3-gram combinatorial patterns

In the following, we focus on combinatorial patterns which correspond to medial occurences of the vocalic hesitations. This choice is motivated by the objective to highlight *rephrasing* and *bracketing properties* of relevant information chunks. Table 5 shows the amount of *euh* which are concerned with the most frequentleft/right medial combinatorial patterns in analyzed data.

trigram type			#occ.	%
PM_{par}	- **EUH** -	PM_{par}	121	17.8
RIC_{par}	- **EUH** -	RIC_{par}	87	12.8
PM_{par}	- **EUH** -	RIC_{par}	85	12.5
RIC	- **EUH** -	RIC	75	11.0
Others			311	45.8

Table 5: 3-gram patterns in the RITEL sub-corpus (*par*=partial).

Concerning these 3-grams, one may notice that most of the utterance-internal contexts involve *partial* tags. These configurations including partial tags mainly highlight disfluent phrases. The occurence of the medial vocalic hesitation surrounded by partial tags may thus be correlated with a *disfluent*-type feature: the vocalic hesitation splits unfulfilled phrases as mark of the speakers' difficulty to put in apropriate words the speech region to rephrase. Pragmatic markers are more often split by the vocalic hesitation than **RIC**. **PM** seem to be subject to rephrasing at least as much as **RIC**. However, *euh* occurs more often in a **RIC** environment and we found 11% of *euh* items surrounded by autonomous **RIC** tags. In this situation, *euh* tends to play a bracketing role.

5. Conclusion

The work proposed here puts a new slant on the various roles the French vocalic hesitation *euh* may carry in speech. Speaker-turn analysis conducted on man-machine question answering dialogs points out that the analyzed item helps in starting the speaker turns and serve in initiating the rephrasing of the utterance or of some particular spoken regions within an on-going utterance. Such properties allow defending the inclusion of the vocalic hesitations in a broad class of discourse markers. In the same way as classical discourse markers, *euh* allows then relevant information framing, in particular when occuring in utterance-internal position. Turn medial vocalic hesitations are efficient cues for word searching and simultaneously for relevant information chunks framing. The *fluent* nature of the vocalic hesitation as highlighted by the current data does not exclude though its role of signaler of speakers' effort to put the lexical message under construction in appropriate words, that is the *disfluent* nature of *euh*. This role may correspond to both initial and utterance-internal *euh* even though the latter seem to be more easily concerned with this property. Our on-going research concerns current findings validation on larger data. A particular emphasis will be put on the various context dependent functions of the vocalic hesitations, that is as disfluent regions of speech indicator vs. truly marker of the discourse structure, keeping in mind our final objective which is the automatic exploitation of such properties.

[3]Confirming classical assumptions on the vocalic hesitations as cues for speakers' intention to efficiently manage the spoken interaction.

[4]When summing up *autonomous* and *partial* realizations

6. Acknowledgements

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

7. References

[1] Mosegaard Hansen, M.-B., "The semantic status of discourse markers", in Lingua, 104, 3/4, 1998.

[2] Swerts, M., "Filled pauses as markers of discourse structure", in Journal of Pragmatics, 485-496, 30, 1998.

[3] Watanabe M., Fillers as Indicators of Discourse Segment Boundaries in Japanese Monologues, Disfluency In Spontaneous Speech (DISS) Workshop, Aix-en-Provence, France, 2002.

[4] Maclay H., Osgood Ch. E., "Hesitation phenomena in spontaneous English speech", 19-44, 15, 1959.

[5] Clark, H., Fox Tree J., " Using *uh* and *um* in spontaneous speaking", in Cognition, 84:1, 2002.

[6] O'Connel, D. C., Kowal, S., *Uh* and *Um* revisited: Are they interjections for signaling delay?, in Journal of Psycholinguistic Research, 34:6, 2005.

[7] Vasilescu I., Adda-Decker, M., Nemoto R., "Caractristiques acoustiques et prosodiques des hsitations vocaliques dans trois langues", in Traitement automatique des langues, 199-298, 49:3, 2009.

[8] Shriberg, E., "To "errrr" is human: Ecology and acoustics of speech disfluencies", in Journal of International Phonetic Association, 153-169, 31:1, 2001.

[9] 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.

[10] Vasilescu, I., Rosset, S., Adda-Decker, M., "On the role of discourse markers in interactive spoken question answering systems", LREC 2010, Valletta, Malta, 2010.

[11] Rosset, S., Galibert, O., Illouz, G., Max, A., "Integrating Spoken Dialog and Question Answering: the RITEL Project, InterSpeech'06, Pittsburgh, USA, 2006.

[12] Toney, D., Rosset, S., Max, A., Galibert, O., Bilinski, E., "An Evaluation of Spoken and Textual Interaction in the RITEL Interactive Question Answering System, LREC'08, Marrakech, Morocco, 2008.