

On the role of discourse markers in interactive spoken question answering systems

Ioana Vasilescu, Sophie Rosset, Martine Adda-Decker

LIMSI-CNRS

B.P. 133 - F-91403 ORSAY Cedex

ioana@limsi.fr, sopia.rosset@limsi.fr, madda@limsi.fr

Abstract

This paper presents a preliminary analysis of the role of some discourse markers and the vocalic hesitation *eu*h in a corpus of spoken human utterances collected with the RITEL system, an open domain and spoken dialog system. The frequency and contextual combination patterns of classical discourse markers and of the vocalic hesitation has been studied. This analysis highlights some specificities in terms of combination patterns of the analyzed items. The classical discourse markers seem to help initiating larger discursive blocks both at initial and medial positions of the ongoing turns. The vocalic hesitation also stand for marking the user's embarrassments and wish to close the dialog.

1. Introduction

Most often oral messages do not perfectly transcribe to written language. When interacting by speech, humans are involved in a complex process including both message generation proper and interaction handling. As a result, they may produce disfluencies, make mistakes during production and include signs of dynamic re-planning in the middle of an utterance; acknowledge or ask questions for grounding purposes; notice that conversation partners become uncomfortable, upset, argumentative; and take or concede the floor (Traum and Hinkelman, 1992; Heeman and Allen, 1999).

It is widely acknowledged that unprepared spoken utterances and more generally oral conversations include a variety of speech events which do not contribute directly to the final message under elaboration as conveyed at the lexical level. However, the conversational level is highly concerned with such events as they contribute to efficiently implement the interaction and carry out interpersonal information. Such items may be qualified as words (for example, ok, well, then, in English; *alors well, donc now* in French, etc.) or non-words (for example, oh, eh, mmhh, uh, um, etc.). They share some common contextual features, and are designed as discourse markers (Schiffrin, 1987), *grunts* (Ward, 2006b; Campbell, 2007) or filled pauses (Shriberg, 1994; Shriberg, 2001).

This paper deals with the role of some discourse markers and of the vocalic hesitation *eu*h in French man-machine spoken dialogs.

According to (Schiffrin, 1987), discourse markers can be defined as set of linguistic expressions that brackets units of talk. (Schiffrin, 1987) includes in a broad class of discourse markers conjunctions (e.g. and, but, or), interjections (oh), adverbs (now, then), and lexicalized phrases (y'know, I mean). As for the filled pauses, they have been initially counted as disfluencies marks (Maclay and Osgood, 1959). More recently (Clark and Tree, 2002) underlined the lexical status of the American English vocalic hesitations *uh* and *um*, and confirmed previous assumptions on their

interaction management functions.

Studies on French confirm cross-language shared properties of discourse markers and vocalic hesitation *eu*h (Ducrot, 1980; Hansen, 1995; Candea, 2000; Vasilescu et al., 2009). Finally, during the last decades, a significant body of work has been dedicated to both discourse markers and filled pauses in the computational area, ((Hovy, 1995; Shriberg, 2001; Litman et al., 2005; Litman and Hirschberg, 1990; Adda-Decker et al., 2003), etc.). Studies focused on their acoustic/prosodic and linguistic discriminant features, on the automatic identification and classification of such events for the automatic speech recognition or dialog modeling purposes, or to study their impact on automatic speech recognition performance.

What about *man-machine conversations*? Generally, automatic generation systems aim at producing straightforward answers to questions. When in trouble, they either ask for missing information or tend to close the interaction on failure.

Earlier research on man-machine dialog for travelling applications (Lamel et al., 2000) highlighted the need for more flexible interaction strategies. In the field of human-machine interface research, building dialog systems that can handle social conventions is one of the ultimate objectives. Most work on real-time social conventions concentrates on turn-taking (Ward, 2006a; Edlund et al., 2005) and responsiveness (Ward and Tsukahara, 2003). In order to improve the naturalness of human-machine interaction, some studies investigate the recognition and generation of emotions. Most of these studies have been conducted in the field of multimodality (Cowie et al., 2001) and of animated conversational agents (Cole et al., 2003).

However, in the field of man-machine conversation, up to now few systems make a smart usage of such discourse structuring items to facilitate the interaction¹.

The major aim of this study is then to improve the interaction possibilities in man-machine dialog by an appropriate usage of such discourse markers, including items clas-

¹This aim has been expressed by (Swerts, 1998)

sically defined as discourse markers, but also filled pauses. Section 2. depicts the architecture of the RITEL system and the data processing. Section 3. focuses on the discourse markers and filled pause in the RITEL data. We then present the working hypothesis (section 4.), followed by the data analysis and preliminary results (section 5.). Findings are summarized in section 6.

2. The RITEL system

Spoken interaction has the potential to be very helpful to humans in information retrieval applications. However, a lack of proactivity and perceived cooperativity was judged as stifling in earlier implementations. To mimic the human behavior and to keep up the naturalness of the dialog with the human user, dialog systems must reproduce the linguistic phenomena corresponding to discourse markers including filled pauses and react efficiently both to the semantic content of the question as carried by the lexical level and to the additional information conveyed by the discourse markers. Moreover, systems need to introduce such items in appropriate places within their responses to ease communication and receptivity by human listeners. The RITEL project aims at providing userfriendly access to open-domain information via spoken interaction (Rosset et al., 2006; van Schooten et al., 2007; Toney et al., 2008). 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.

2.1. System architecture

We provide a brief overview of the RITEL system architecture. A more detailed description can be found in (Rosset et al., 2006). The system architecture (Figure 1) is highly distributed and based on servers and specialized modules that can exchange messages.

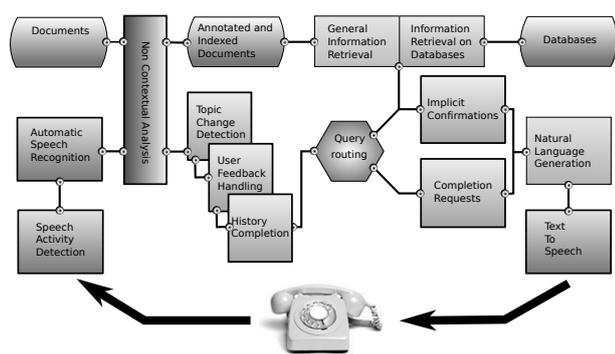


Figure 1: System architecture

Some aspects of this architecture make the system differ from systems with similar objectives: a unique analysis module is used for document indexing and user query analysis; user inputs are handled by specific registered modules according to their type; dialogue management is not centralised as in other approaches (Zue and Glass, 2000) but rather distributed over the various components of the platform. The Speech Activity Detection and the Automatic

Speech Recognition components are fully described in (van Schooten et al., 2007).

Analysis of both indexed documents and user utterances are handled by the same module which is called Non Contextual Analysis (NCA) because no information from the dialogue or previous utterances is used. The Question Answering system is described in (Rosset et al., 2008). The indexing server's main role is to retrieve snippets, i.e. lines of documents corresponding to a given query. Queries take the form of a list of named entities and answer types. Candidate answers are ranked according to the scoring mechanism detailed in (Rosset et al., 2008).

2.2. Data processing

The data collected during different evaluations of the RITEL system ((Rosset and Petel, 2006; Rosset et al., 2006; van Schooten et al., 2007; Toney et al., 2008)) has been used in this study. The overall corpus has been tagged with the Ritel analyzer. The analysis is *non-contextual* because each sentence (or turn or speech segment) is processed in isolation. The general objective of this analysis is to find the bits of information that may be of use for search and extraction, which we call *relevant information chunks*. These can be of different categories: named entities, linguistic entities (e.g. verbs, prepositions), or specific entities (e.g. scores). The entity definition on which the system is based is hierarchical. This system is rule-based and uses *WMatch* (Galibert, 2009) This engine matches (and substitutes) regular expressions using words as the base unit instead of characters. This property enables the use of classes (lists of words) and macros (sub-expressions in a larger expression). *WMatch* includes also NLP-oriented features like strategies for prioritizing rule application, recursive substitution modes, word tagging (for tags like noun, verb, etc.), word categories (number, acronym, proper name, etc.). Analysis is multipass, and subsequent rule applications operate on the results of previous rule applications which can be enriched or modified. The RITEL analyzer provides more than 300 different types classes of words and multi-word expressions, most of them being semantic and close to some NE definition. The figure 2 shows an example of such an analysis.

3. Discourse markers in the RITEL corpus

In the following paragraphs, several items which act as markers of discourse structure in the RITEL corpus are examined: *bon*², *alors*, *bah*, *ben*, and the filled pause *ehh*. As one can notice, the broad class of *discourse markers* includes quite heterogeneous *lexical* and *non-lexical* items, from purely lexical such as *bon* and *alors*, to vocalic hesitation *ehh* and to the shorter forms of the former, i.e. *bah* and *ben*. This selection is both based on the literature and em-

²Contextual classification of *discourse markers* vs lexical (adjectival, adverbial) use of ambiguous items such as *bon* and *alors* is an essential step in dialog understanding which has been widely considered in computational studies. In the current work a manual annotation of the *discourse markers* use of such ambiguous items has been conducted by the second author. The adjectival and adverbial use of *bon* and *alors* has been then excluded.

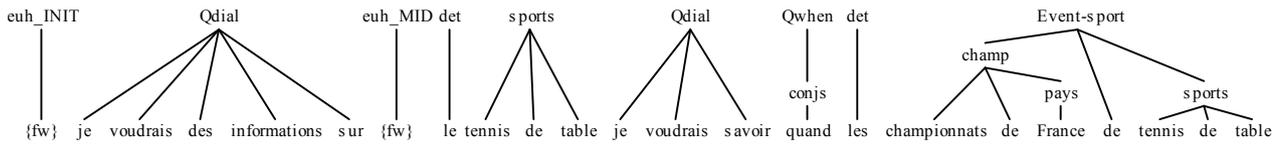


Figure 2: Example of a RITEL analysis: *uh I would like some information about uh table tennis I would like to know when the France championship of table tennis...*

pirical corpus-dependent observation as illustrated by the examples discussed in the next paragraph.

3.1. Examples of discourse markers in the RITEL corpus

Whereas the RITEL system provides answers with different levels of accuracy and without any use of *discourse markers/hesitations*, human users show variable propensity to negotiate with the system by rephrasing up to a dozen of times the initial question. They make use of different rephrasing techniques and employ such items with different aims.

The examples below illustrate for instance typical occurrences of discourse markers in the RITEL corpus. In the **Example 1** *alors* and *ehh* are employed as discourse markers. The two items behave as turn-initiators, the turn-opening function being frequently carried out by the analyzed items (almost half of the discourse markers occur at the beginning of a speaker turn, as detailed in section 5.). In this example, the speaker makes use of the rephrasing strategy four times to question the RITEL system. S/he starts with a question in which the only discourse marker is the discourse particle *alors* (now) which stands here for a mark of turn-initiation. Then, s/he uses rephrasing four times, each time employing the filled pause *ehh* as turn-initiator but also to search for a phrase (Q_2^2) or only for a word (Q_3^2). Finally, in Q_3^2 , s/he employs the same filled pause with a different function: as turn and dialog-ending initiator.

• Example 1

Q_1^2 : *alors j' aimerais savoir qui a inventé la théorie des avantages comparatifs (I would like to know who invented the comparative advantages theory)*

Q_2^2 : *ehh je cherche le nom d' un économiste qui a inventé la théorie des avantages comparatifs (uh I'm looking for the name of the economist who invented the theory of comparative advantages)*

Q_3^2 : *ehh je souhaiterais ehh connaître le nom de celui qui fut l' inventeur de la théorie des avantages comparatifs (uh I would like uh to know the name of the person who invented the theory of the comparative advantages)*

Q_4^2 : *j' ai déjà répondu à cette question je voudrais savoir de quand date la théorie des avantages comparatifs et qui l' a mise au point (I already answered to this question and I would like to know when the theory of the comparative advantages was invented and who invented it)*

Q_5^2 : *ehh non ça n' est pas ehh ça n' est pas ehh le sujet de (uh/um no it's not uh/um it's not uh/um the topic)*

In **Example 2**, the first question (Q_1^1) does not exhibit discourse markers, but then as the human speaker is confronted with a wrong answer, h/she makes use of a rephrasing strategy in order to get a more appropriate reaction from

the automatic system. Q_2^1 and Q_3^1 illustrate this strategy (rephrasing). The filled pause *ehh* stands for turn-initiation and for speaker's effort to put the first question in other words.

• Example 2

Q_1^1 : *je voudrais savoir combien il y a d' aéroports internationaux à Marseille (I would like to know how many international airports are there in Marseilles)?*

Q_2^1 : *ehh j' aimerais connaître le nombre de d' aéroports internationaux dans la ville de Marseille (uh I would like to know the the number of international airports around the city of Marseilles)?*

Q_3^1 : *ehh je voudrais savoir combien d' aéroports compte la ville de Marseille s'il-vous-plaît (uh I would like to know how many international airports are located around the city of Marseilles please)?*

The examples discussed here suggest that the filled pause *ehh* carries the following functions: turn initiation, and also phrase search within a rephrasing strategy. In addition, the user employs of discourse particles to initiate the dialog such as *alors* in Example 1.

3.2. Selected data description

In this study we make use of the data obtained through the methodology described in section 2. The RITEL corpus contains 652 dialogues with more than 6 hours from user turns. There are 6720 user turns comprising a total of 71 089 lexical items (tokens), with a total of 3434 distinct lexical words (types). Discourse markers are then present in 1718 that is 25.56% of the utterances of the RITEL corpus.

The Table 1 gives a summary of the RITEL corpus.

Duration	6h40
#dialogues	652
#turns	6720
#turn/dial	10
#turns w. DM	1718 (25.56%)

Table 1: General description of the RITEL corpus

Among the 1718 utterances containing at least one discourse marker the filled pause *ehh* prevails. *Ehh* is then present in 1520 (88.2%) of the utterances containing at least a discourse marker.

4. Working hypothesis

The RITEL corpus is a good candidate to study human rephrasing strategies. The general hypothesis of this work is that in the RITEL corpus discourse markers keep their

"native" role of "bracketing" units of speech and that this structuring is useful -in particular- when rephrasing. As for the hesitations, the filled pause *eu*h may also indicate the words or phrases prone to rephrasing and which are not optimally processed by the RITEL system. That means if the speaker rephrases the question or the particular regions of a question using a classical discourse marker or a hesitation as initiators, the broad lexical context of such elements may point out the problematic speech region for the RITEL system. Discourse markers may stand for rephrasing plans of the human speakers, namely to put in other words the utterance or only a salient region with the final aim of getting more responsiveness from the RITEL system.

In order to highlight the spoken regions of interest, we conducted a study on the discourse markers combination patterns with different dialog tags. Different lexical contexts prior to/following the discourse markers have been considered. In this purpose, the word and multi-word tags obtained through the methodology described in section 2.2. have been clustered in two main classes. The obtained classes gather two types of lexical items and multi-word expressions, that is:

PM pragmatics markers which group speech act markers (e.g. *<Dialog-opening> hello </dialog-opening>*, *<Qdial> I'm looking for </Qdial>*, *<Qnegdial> I'm not looking for </Qnegdial>*) and question markers (e.g. *<Qwho> who </Qwho>* wrote..., *<Qmeasure> how many </Qmeasure>*)

RIC relevant information chunks, that is linguistic chunks which regroup extended named entities (e.g. *<loc>La Valette</loc>*, *<pers>Victor Hugo</pers>*) and other linguistic chunks (i.e. lexical items/syntagms which are not in one of the previous categories).

The rationale behind the clustering is that classes of words and multi-word expressions may be categorized in the RITEL corpus according to the final aim of the study, which is to delimit the relevant information chunks prone to rephrasing, e.g. **RIC**. Then words and multi-word expressions may belong to the class of **RIC**, or serve other purposes, mainly interactional. The global class of pragmatic markers assemble all lexical items and expressions which are not salient linguistic chunks or discourse markers.

As illustrated by Figure 3, discourse markers may initiate rephrasing at two major points of a speaker turn: at the beginning of the utterance as a *global* initiation of the whole sentence; or within the utterance pointing on a *local* region to rephrase (word or phrase). As for Figure 2, it shows the output of the RITEL analyzer after clustering the different dialog tags as pragmatic markers **PM** or relevant information chunks **RIC**.

The cooccurrence of discourse markers *versus* **PM** or **RIC** has been considered in different combination patterns within a maximum scope of two tags. Table 2 show examples of left and right combination patterns.

5. Combinatorial analysis

This section aims at explaining the constraints which govern the cooccurrence of different types of discourse markers with the dialog tags with the aim that this analysis may inform us on their potential rephrasing functions. The position of discourse markers in the speaker turn has been selected as a relevant factor: initial (**INIT**), medial (**MED**) and final (**END**). Table 3 summarizes the lexical entries concerned with the combinatorial analysis.

One may notice that the great majority of the data is represented by relevant information chunks **RIC** (65.2 %). Delimiting such chunks with the help of discourse markers is the final objective of this work.

Then 17.7 % of the data are represented by the pragmatic markers **PM**.

Both **PM** and **RIC** may occur in the neighborhood of discourse markers, however the position of discourse markers in the speaker turn is expected to correlate with a particular initiation role, e.g. global (utterance) or local (word/phrase) initiation. Finally, discourse markers represent 17.1 % of the data and have been splitted here in classical discourse markers (*CDM*) *versus* filled pause *eu*h.

Type	nbr. occur.	%
RIC	9764	65.2
PM	2646	17.7
EUH_INIT	939	6.3
CDM_INIT	202	1.3
EUH_MID	689	4.6
CDM_MID	99	0.7
EUH_END	623	4.2
Total	14962	100

Table 3: Pragmatic markers (**PM**), relevant information chunks (**RIC**), classical discourse markers *CDM* and filled pause *eu*h in the RITEL corpus.

One may notice that the great majority of *CDM* and *eu*h occurs at the beginning of the speaker turns suggesting similar functions of the two classes of items. *CDM* are not encountered in final positions. Filled pauses are particularly frequent as turn-initiators, however they are also employed within and at the end of an utterance leading to believe that they have at least one different function as compared to *CDM* (i.e. turn initiators). This distribution is highlighted by the Table 4.

	INIT	MID	FIN
<i>EUH</i> (88.2%)	41.7%	30.6%	27.7%
<i>CDM</i> (11.8%)	67.1%	32.9%	-

Table 4: Distribution of the two categories of discourse markers *eu*h and *CDM* according to the position in the speaker turn

The following paragraphs will focus on the combination patterns of *discourse markers* *CDM* and *eu*h with **PM** and **RIC**.

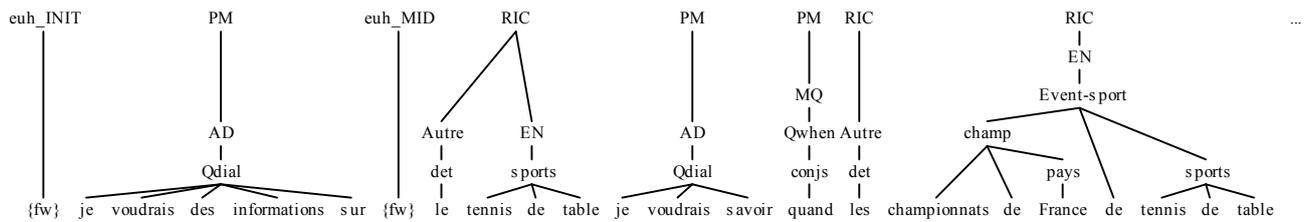


Figure 3: Example of Ritel analysis after clustering: *uh I would like some information about uh table tennis I would like to know when the France championship of table tennis...*

Combination	tag example	utterance example
DM tag	EUH PM	<EUH> euh </EUH> <PM> je ne veux rien savoir sur </PM> ...
DM tag tag	EUH PM RIC	<EUH> euh </EUH> <PM> qui </PM> <RIC> a réalisé Titanic </RIC>
tag DM	PM CDM	<PM> oui </PM> <CDM> alors </CDM> ...
tag tag DM	PM RIC DM	<PM> je voudrais savoir </PM> <RIC> le tennis de table </RIC> <EUH> euh </EUH> ...

Table 2: Examples of left and right combination patterns

5.1. 2-gram patterns

Table 5 summarizes the 2-gram combination patterns of discourse markers. Three factors are considered in the analysis: the contextual class **PM** and **RIC**, its position against the discourse marker **right/left** (+1/-1), and the position of the discourse marker within the speaker turn: initial (**INIT**), medial (**MID**) and final (**END**).

		CDM		EUH	
		# occ.	%	# occ.	%
INIT (+1)	PM	114	56.7	587	66.9
	RIC	64	31.9	271	30.9
	DM	23	11.4	19	2.2
	Total	201	100.00	877	100.00
MID (-1)	PM	25	25.3	191	27.7
	RIC	21	21.2	449	65.2
	DM	53	53.5	49	7.1
	Total	99	100.00	689	100.00
MID (+1)	PM	44	44.5	150	21.8
	RIC	43	43.4	451	65.4
	DM	12	12.1	88	12.8
	Total	99	100.00	689	100.00
END (-1)	PM	-	-	96	15.4
	RIC	-	-	454	72.9
	DM	-	-	73	11.7
	Total	-	-	623	100.00

Table 5: 2-gram combination patterns for *CDM* and *euh*.

Table 5 shows that both *CDM* and *euh* are followed in their great majority by pragmatic markers when in initial position (INIT(+1)). On the reverse, the medial position favours the occurrence of relevant information chunks after *euh* (65 %) versus *CDM* (43 %). The right context patterns are somehow different for classical discourse markers and for the vocalic hesitation. The latter is more often followed by relevant information chunks whereas the classical discourse markers equally occur next to pragmatic markers and relevant information chunks. This tendency supports the hypothesis of the word/phrase search marker of the vocalic hesitation: this function is generally acknowledged and in

this paper we look for evidence to support similar behavior of the analyzed items. As for the *CDM*, the distribution in the neighborhood of both **PM** and **RIC** is to confirm in further studies: such items seem to be specifically prone to initiate global rephrasing than the vocalic hesitation. When occurring in the initial position, the turn-initiation function is non-ambiguous for both *CDM* and vocalic hesitations. The left context patterns of the discourse markers follow the same tendency, i.e. the vocalic hesitation being surrounded by **RIC**, whereas the *CDM* occurring in both **RIC** and **PM** contexts. Finally, only the vocalic hesitation may fill the final position of the speaker turn, preceded in majority by **RIC**. This distribution supports the idea that *CDM* are obviously to be related to rephrasing strategies, whereas the vocalic hesitation may fill other pragmatic functions. Speakers seem to make use of vocalic hesitations to notify turn or dialog ending and to possibly mark an embarrassment due to the unsuccessful interaction with the system.

5.2. 3-gram patterns

Tables 7 and 6 below summarize the 3-gram combination patterns. The same factors as for the 2-grams are considered to display the data. Tables show the three most frequent patterns, whereas the remaining combinations are listed as "Others". The differences observed for *CDM* versus *euh* patterns according to the position in the speaker turn (initial, medial and final) are supported by the 3-grams specificities. Whereas both *CDM* and *euh* are followed by a pragmatic marker and a relevant information chunk in initial position, the medial position correlates with different patterns. The most frequent right (n+2) context for *euh* is two relevant information chunks (59.32 %), whereas *CDM* may be followed equally by two relevant information chunks or a pragmatic marker and a relevant information chunk (30.8 % of combinations).

Concerning the left bigrams, *euh* distribution confirms the global tendency of being in a **RIC** context if the speaker hesitates in the middle of the utterance. For *CDM* though, the limited number of occurrences available for the current analysis does not allow inferring on a clear tendency. Finally, the final *euh* is mainly preceded by two **RIC**, sug-

Position	Combi.	nbr. occ.	%
INIT (+2)	PM RIC	346	39.6
	RIC RIC	202	23.1
	PM PM	136	15.6
	Other	189	21.7
MID (+2)	RIC RIC	350	59.3
	PM RIC	111	18.8
	RIC PM	46	7.8
	Other	83	14.1
MID (-2)	RIC RIC	324	55.0
	PM RIC	80	13.6
	RIC PM	52	8.8
	Other	133	22.6
END (-2)	RIC RIC	398	67.1
	EUH PM	47	7.9
	RIC EUH	47	7.9
	Other	101	17.1

Table 6: 3-gram combination patterns for *euh*.

Position	Combi.	nbr. occ.	%
INIT (+2)	PM RIC	69	35.2
	RIC RIC	34	17.3
	PM PM	31	15.8
	Other	62	31.7
MID (+2)	PM RIC	29	30.8
	RIC RIC	29	30.8
	RIC PM	12	12.8
	Other	24	25.6
MID (-2)	RIC RIC	8	24.2
	CDM RIC	5	15.2
	EUH EUH	4	12.1
	Other	16	48.5

Table 7: 3-gram combination patterns for *CDM*.

gesting the medial and final vocalic hesitations may serve as right and left boundaries of **RIC**.

6. Conclusion

We presented here a preliminary analysis of the role of some discourse markers and the vocalic hesitation *euh* in the RITEL corpus. The RITEL corpus has been obtained through an open-domain question/answering system and consists in speaker turns from human interlocutors who question the automatic system. The relatively poor responsiveness of the early stage of the RITEL system leads to the systematic use of rephrasing strategies and humans are willing to put the same question in other words up to dozen of times. The corpus was annotated to delimit strings of words mainly serving interactional purposes, termed pragmatic markers (PM), from the relevant information chunks (RIC) typically prone to rephrasing. The occurrence of discourse markers, both classical items and filled pauses *euh*, and their contextual combinations with PM and RIC have been examined with the long term objective improving the naturalness of the RITEL system in two possible ways: (i) an improved spotting of relevant information chunks within the questions; (ii) introducing discourse marker generation and hesitation capacities within the RITEL response

module. The proposed analysis highlighted combinatorial specificities of the discourse markers and hesitations with respect to the considered PM and RIC classes. Concerning the general rephrasing, both classical discourse markers and filled pauses are used to initiate the speaker turn, whereas *euh* is more frequent to initiate local word/phrase search. The classical discourse markers, apart from being less frequent in our corpus than the vocalic hesitation, seem to help initiating larger discursive blocks both at the initial and medial positions of the ongoing speaker turn. Finally, the relative frequency of the vocalic hesitation may be linked to other contextual functions than rephrasing: as the human speaker is confronted with unsuccessful answers in the course of the dialog, vocalic hesitations may also stand for marking his/her embarrassment and wish to close the dialog. Discourse markers and vocalic hesitations will further be examined in relation with the speaker turn position within the dialog. Future work will address the corpus-related functions of different items acting as markers of discourse and dialog structure in the RITEL system, as well as implementations to improve the capacities of the RITEL system .

7. Acknowledgements

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

8. References

- M. Adda-Decker, B. Habert, C. Barras, G. Adda, Ph. Boula de Mareüil, and P. Paroubek. 2003. A disfluency study for cleaning spontaneous speech automatic transcripts and improving speech language models. In *Proceedings of Disfluency In Spontaneous Speech (DISS) Workshop*, Gothenburg, Sweden.
- N. Campbell, 2007. *Verbal and non verbal communication behaviours*, volume 4775/2007 of *Lecture notes in computer science*, chapter On the use of non verbal speech sounds in human communication, pages 117–128. Springer Berlin Heidelberg.
- M. Candea. 2000. Les *euh* et les allongements dits "d'hésitation": deux phénomènes soumis à certaines contraintes en français oral non lu. In *Proceedings of XXIIIèmes Journées d'Etude sur la Parole*, Aussois, France.
- H. Clark and J. Fox Tree. 2002. Using *uh* and *um* in spontaneous speaking. *Cognition*, 84:1.
- R. Cole, S. Van Vuuren, B. Pellom, K. Hacioglu, J. Ma, J. Movellan, S. Schwartz, D. Wade-stein, W. Ward, and J. Yan. 2003. Perceptive animated interfaces: First steps toward a new paradigm for human computer interaction. *Proceedings of the IEEE, Special Issue on Multimodal Human Computer Interface*.
- R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, W. Fellenz, and J. Taylor. 2001. Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*, 18(1):32–80.
- O. Ducrot. 1980. *Les mots du discours*. Les Éditions de Minuit, Paris.
- J. Edlund, M. Heldner, and J. Gustafson. 2005. Utterance segmentation and turn-taking in spoken dialogue

- systems. *Computer Studies in Language and Speech*, 8(6):576–587.
- O. Galibert. 2009. *Approches et méthodologies pour la réponse automatique à des questions adaptées à un cadre interactif en domaine ouvert*. Ph.D. thesis, Université Paris-Sud 11, Orsay, France.
- A. B. Hansen. 1995. Marqueurs métadiscursifs en français parlé : l'exemple de "bon" et de "ben". *Le français moderne*, 63:1:20–41.
- P. A. Heeman and J. F. Allen. 1999. Speech repairs, intonational phrases and discourse markers: Modeling speakers' utterances in spoken dialog. *Computational Linguistics*, 25(4):527–571.
- E. Hovy. 1995. The multifunctionality of discourse markers. In *Proceedings of the Workshop on Discourse Markers*, pages 1–11, Egmond-aan-Zee, The Netherlands.
- L. Lamel, S. Rosset, J.-L. Gauvain, S. Bennacef, M. Garnier-Rizet, and B. Prouts. 2000. The limsi arise system. *Speech Communication*, 31(4):339–354.
- D. Litman and J. Hirschberg. 1990. Disambiguating cue phrases in text and speech. In *Proceedings of the 13th conference on Computational linguistics*, pages 251–256, Helsinki, Finland.
- D. Litman, J. Hirschberg, and M. Swerts. 2005. Characterizing and predicting corrections in spoken dialogue systems. *Computational Linguistics*, 32:3:417–438.
- H. Maclay and Ch. E. Osgood. 1959. Hesitation phenomena in spontaneous english speech. *Word*, 15:19–44.
- S. Rosset and S. Petel. 2006. The Ritel Corpus - An annotated Human-Machine open-domain question answering spoken dialog corpus. In *LREC'06*, pages 1640–1643, Genoa, May.
- S. Rosset, O. Galibert, G. Illouz, and A. Max. 2006. Integrating spoken dialog and question answering: the ritel project. In *in InterSpeech'06*, Pittsburgh, USA.
- S. Rosset, O. Galibert, G. Bernard, E. Bilinski, and G. Adda. 2008. The limsi participation to the qast track. In *Working Notes of CLEF 2008 Workshop*, Aarhus, Denmark, September.
- D. Schiffrin. 1987. *Discourse markers*. Cambridge University Press, Cambridge.
- E. Shriberg. 1994. *Preliminaries to a theory of speech disfluencies*. Ph.D. thesis, University of California, Berkeley, United States.
- E. Shriberg. 2001. To “errrr” is human: Ecology and acoustics of speech disfluencies. *Journal of International Phonetic Association*, 31:1:153–169.
- M. Swerts. 1998. Filled pauses as markers of discourse structure. *Journal of Pragmatics*, 30:485–496.
- D. Toney, S. Rosset, A. Max, O. Galibert, and E. Bilinski. 2008. An Evaluation of Spoken and Textual Interaction in the RITEL Interactive Question Answering System. In *Proceedings of the 6th International Language Resources and Evaluation (LREC'08)*, Morocco.
- D. R. Traum and E. A. Hinkelman. 1992. Conversation acts in task-oriented spoken dialogue. *Computational Intelligence*, 8:575–599.
- B. van Schooten, S. Rosset, O. Galibert, A. Max, R. op den Akker, and G. Illouz. 2007. Handling speech input in the ritel qa dialogue system. In *in InterSpeech'07*, Anvers, Belgique.
- I. Vasilescu, M. Adda-Decker, and R. Nemoto. 2009. Caractéristiques acoustiques et prosodiques des hésitations vocaliques dans trois langues. *Traitement automatique des langues*, 49:3:199–298.
- N. Ward and W. Tsukahara. 2003. A study in responsiveness in spoken dialog. *International Journal of Human-Computer Studies*, 59(6):959–981.
- N. Ward. 2006a. Evaluating real-time responsiveness in dialog. In *in InterSpeech'06*, Pittsburgh, USA.
- N. Ward. 2006b. Non-lexical conversational sounds in american english. *Pragmatics & Cognition*, 14:1:129–182.
- V. Zue and J. Glass. 2000. Conversational interfaces: Advances and challenges. *Proceedings of the IEEE, Special Issue on Spoken Language Processing*, 88.