Characterization of Conversational Activities in a Corpus of Assistance Requests

François Bouchet

Abstract. Modeling linguistic interaction is a crucial point to provide assistance to ordinary users interacting with computer-based systems and services. A first issue is more particularly the characterization of the linguistic phenomena associated with the Function of Assistance, in order to define an assisting rational agent capable of pertinent reactions to users' requests. In this paper, we present a corpus based on users' requests registered in actual assisting experimentations. First we compare it to similar task-oriented dialogical corpora through a method based on interactional profiles (speech acts oriented analysis) that assesses its specificity. Then we show through an annotation of conversational activities that the collected corpus is heterogeneous (40% of its requests are actually not task-oriented), and use again the interactional profiles approach to go towards an automatic identification of those activities.

1. Introduction

1.1. Assisting Ordinary Users

The number of novice (or ordinary) users[1] of computer applications and services has been quickly increasing for the past decades and is likely to keep growing for some time. An example of such users could be the average Net surfer swapping between websites in order to use sporadically web-services (buying an airplane ticket...) or producing personal web-content (sharing photos, videos...). Because of the complex nature of the new websites and commercial applications, these users will inevitably face difficulties to achieve their objectives. That situation can lead users into a cognitive distress, with a significant negative impact for the application providers.
Natural language (NL) has been shown to be an ideal way to provide assistance to novice users interacting with computer applications or services. First, it appears that it is a modality spontaneously used when a problem arises, particularly (but not exclusively) in the case of novice users, and which closely reflects the cognitive processes of the user (Ericsson & Simon 1993), but the use of multimodality for assistance also allows a clear cognitive separation between the application and the help system (with graphical (Morrell & Park 1993) or dialogical (Amalberti 1996) modalities). But since these situations arise mainly because of the lack of knowledge about the software, it leads to many linguistic difficulties such as: user-specific vocabulary, degraded spelling (for typed requests) or degraded prosody (for oral requests), bad categorizations, *etc.* It thus makes this type of requests really difficult to process and to interpret.

1.2. Characterizing the Function of Assistance

In order to bring an answer to this need of assistance, the DAFT project[2] from LIMSI-CNRS (Sansonnet, et al. 2005) intends to develop Assistant Conversational Agents (ACA)

[1] By ordinary user, we mean a non expert person using a software application.
[2] `http://www.limsi.fr/~jps/research/daft/`

able to analyze Natural Language requests expressed without constraints by ordinary users during their use of applications of increasing complexity (applets, webpages, active websites, text processor…). This choice is also motivated by the additional benefits brought by the use of an embodiment for dialogue system in terms of trust and believability, a phenomenon known as the 'Persona Effect' (Lester, et al. 1997).

The objectives of the data processing sequence of DAFT assistance system is to characterize generic components of the Function of Assistance, and to propose a rational agent engine able to assist ordinary users in the most frequent cases. The global architecture of the system is classically made of:

- a semantic analysis module of NL requests to build formal requests,
- a module of reasoning on the application model that returns a formal request,
- a production module aiming to express the agent's answer in a multimodal way (spoken or written way, action on the application, gestures from an animated virtual character…).

As a preliminary step towards the creation of this NL processing chain, we have chosen a corpus-based approach to study the phenomena we had to model, leading to three questions: is our corpus original (*i.e.* different from existing ones) and how to prove it? What kind of conversational activities does it actually contain? And how is it possible to distinguish them automatically?

In a first part of this article, we describe how that corpus has been built up and give an excerpt of collected requests. In a second part, we introduce the concept of interactional profiles to compare, in terms of speech acts, our corpus to similar ones. Finally, we show in section 4 that our corpus actually contains requests representing different conversational activities and apply again the interactional profile method to compare those subcorpora.

2. Corpus collection and building

2.1. Methodology

Currently, very few public data is actually available concerning Human-Computer dialogue. Moreover, our scope is rather different from classical dialogue systems and actually closer from Natural Languages Interfaces (Androutsopoulos & Aretoulaki 2003), since we're dealing with *isolated requests* rather than with *dialogic sessions*. On top of this, it was crucial for us to control precisely the experimental conditions for the collection of requests which had to deal specifically with the assistance. For all those reasons[3], we have chosen to collect our own specific corpus, which we'll refer to as the Daft corpus in this article.

It has been built up from three different sources (each providing roughly a third of the total corpus):

1. During two years, about 100 human subjects have been faced with several applications assisted by the Daft system (in its 1.0 version): three Java applets (modal and

[3]The validity of that difference hypothesis will be further investigated in section 3.

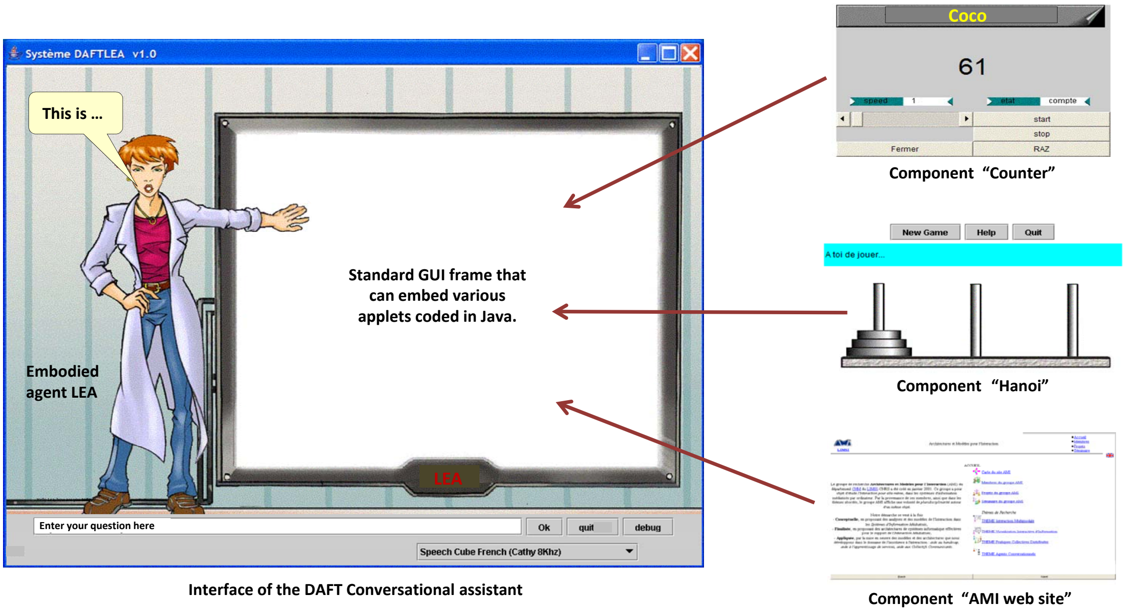

Figure 1: The conversational agent interface into which were embedded different kind of applications (Java applets or dynamic websites)

unmodal, *i.e.* with threads), two websites (one was active, *i.e.* could be dynamically edited by users – *cf.* figure 1 for an example of interface);

2. From two thesauri (Molinsky & Bliss 1994, Atkins & Lewis 1996), we have manually constructed some requests in order to provide a wider linguistic coverage of the assistance vocabulary and idioms;

3. Recently, we have added to the corpus some FAQ extracted from integrated help systems and websites concerning two widely used document creation softwares (LaTeX and Microsoft Word).

2.2. General view of the Daft corpus

Table 1 shows excerpts from the Daft corpus (currently made of 11.000 requests), emphasizing some characteristics:

- more than half of the users' sentences are *not well-formed* (spoken expressions, spelling/syntactic/grammatical mistakes, acronyms from SMS and internet slang...), and some mistakes are not that easy to detect and correct with classical NLP tools;

- requests are *not* stored as part of a dialogue but as isolated sentences. As mentioned previously, it appeared as suggested by (Capobianco & Carbonell 2002) that in the domain of assistance, dialogic interactions are almost always limited to a single conversational turn and hence can be dealt with as *isolated requests*.

N°	**Original request**	**Translated request**
1	**appuies** sur le bouton quitter	**clicks** on the quit button
2	**clickersur** le bouton back	**clickon** the back button
3	bon, reviens à **l apage** d'accueil	ok, come back to **th ehomepage**
4	**a** quoi sert cette fenêtre**,**	what is this window for**,**
5	**c quoi** le GT ACA	**WDYM** by GT ACA
6	le bouton "fermer" et le bouton "quitter" ont exactement le même fonctionnement?	do the "close" button and the "quit" button work exactly the same way?
7	je ne **vosi** aucune page de **demso** !!	I **cna't** see any **demso** page!!
8	j'ai été surpris qu'il manque une fonction d'annulation globale	I was really surprised to see a global cancel function is missing
9	ça serait mieux si on pouvait aller directement au début	it'd be better to be able to go directly at the beginning
10	auf viedersen	**auf viedersen**
11	bon à rien !	you good-for-nothing!
12	Quel genre de musique tu aimes ?	What kind of music do you like?
13	ca marche **:-)**	works for me **:-)**
14	j'aime tes cheveux Léa	I like your hair Lea

Table 1: Original and translated examples from the Daft corpus – mistakes, in bold, have been translated as closely as possible from the original French version

3. Corpora comparison

We have seen in 2.1. that our initial hypothesis for the corpus collection was that our domain of study was specific enough to prevent us from using requests from similar existing corpora. We thus focus in this section on the validation of that hypothesis through an analysis of speech acts in different corpora annotated with different taxonomy.

3.1. Data: corpora to compare

We have chosen for this comparative study three reference corpora of annotated task-oriented dialogues:

- Switchboard (Jurafsky, et al. 1998): 200.000 manually annotated utterances from task-oriented phone talks;
- MapTask (Carletta, et al. 1996): 128 dialogues in which one person has to reproduce a route on a map, following instructions from another person with a similar map;
- Bugzilla (Ripoche 2006): 1.200.000 comments from 128.000 bug reports created during the development of the Mozilla Foundation's suite.

Switchboard and MapTask are coming from oral interactions and hence are naturally richer in words than written corpora (Kelly & Chapanis 1977), but the closeness of activities appeared more important for our comparison than this origin difference. As for the language difference (those corpora being in English whereas ours is in French), it is probably not significant in terms of speech acts. Besides, although some French (oral as well) corpora could also be relevant for this comparison, like Air France (Morel 1989), Ozkan (Ozkan 1994) or even a small French MapTask corpus (Post 2000), they aren't provided with a speech act taxonomy and annotations as it is the case the three ones aforementioned.

3.2. Methodology: interactional profiles and taxonomy mapping

Interactional profile is defined in (Ripoche 2006) as "the distribution of speech acts appearing in a given interactional unit". The interactional unit itself is nothing but a coherent set of speech acts chosen according to the analysis objective: a single utterance, a dialogue, a corpus, *etc.* Once the interactional unit has been chosen, the ratio of each speech act in this unit is calculated, and the profile itself is generally displayed as an histogram in order to have a synthetic view associated with the class of interaction. The main interest of interactional profiles is not as much their intrinsic value as the possibility they offer to allow *comparison* between two different classes of interactions.
This approach is fundamentally close from the model developed in (Twitchell, et al. 2004) but has some noticeable differences though:

- since for more accuracy we prefer to manually annotate the speech acts, our approach is discrete rather than probabilist (Twitchell et al. (2004) allowing elements of an unit to belong to many speech acts with a probability function);

- our approach is conceptually more collective (to study a global behaviour) than individual (study of one person's interactions), although both methods can certainly be used in both contexts;

- we consider the interactional profiles defined as having an absolute value, whereas Twitchell et al. (2004) subtract to it a global average profile of interactions supposed to have a normative value.

Here, the interactional unit chosen is the corpus as a whole, and the speech acts set for the shared taxonomy is made of the five classical searlian speech acts (Searle 1969), which are generic enough to allow comparison. We thus had to map the existing taxonomies used to annotate those corpora into that common one.
In the case of Switchboard, although the original annotation was done along four dimensions, it has appeared that combinations between those dimensions were rather limited, allowing to distinguish a total of 42 main categories in the DAMSL taxonomy (Jurafsky et al. 1998), which are the ones we have been using here. Some speech acts being very specific (for instance Switchboard's "self-talk") are however not trivial to convert even into a very general taxonomy as the one chosen. Similarly, although annotated at different levels, the speech acts from MapTask can be considered in a flatten way for this mapping, as displayed in table 2.

3.3. Results

The interactional profiles of those four corpora are diplayed on figure 2. Because of the impossiblity to have a perfect mapping explained in the previous section, their interpretation needs to be done cautiously. Nevertheless, some very distinct characteristics seem to distinguish the Daft corpus from the three others, namely:

Searle	Assertive	Commissive	Directive	Expressive	Declarative	Unknown
MapTask	clarify explain reply-w	-	align check instruct query-w query-yn	acknowledge ready reply-n reply-y	-	uncodable

Table 2: Speech acts mapping for the MapTask taxonomy

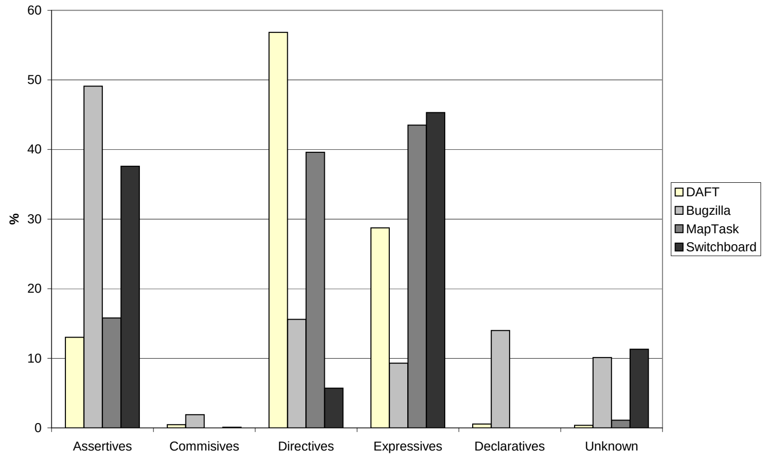

Figure 2: Interactional profiles comparison of four task-oriented dialogue corpora

- a majority of *directives* (57%), explained by the high number of orders or questions to the agent. Although other corpora are also task-oriented, interactions were only between humans, and it seems likely that talking to a computer (even through an ECA) make requests more direct, as users generally don't expect the agent to be able to make complex inferences.

- the number of *assertives* is rather low (13%), users prefering to express their feelings and states of mind (29%) concerning the situation rather than those same facts in a neutral and "objective" way as they do for example in the Bugzilla corpus, since not knowing how to do something is considered as a rather stressing situation.

- very few *commissives* (1%) are observed, which can be easily explained by the relationship user-agent: if the agent is in essence subordinate to the user, the latter rarely feel commited to do whatever the agent can suggest to him (even when the answer is perfectly relevant).

The use of conversational profiles for this comparison clearly helped to demonstrate the differences existing between those corpora. More particularly, it confirmed that the linguistic domain covered by the Function of Assistance to novice users required a dedicated corpus.

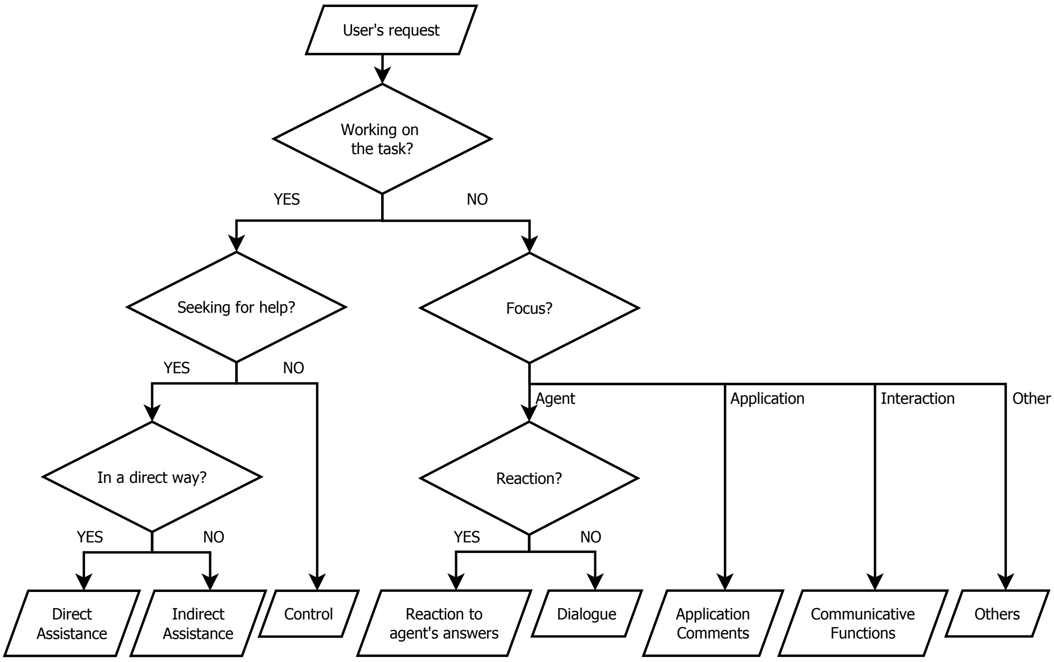

Figure 3: Requests activity classification protocol

4. Conversational activities analysis

During the corpus collection phase, human subjects were informed that they had to do some tasks for which they could ask help (if needed) from an artificial assistant agent embedded in the program to assist them. Nevertheless, subjects were completely free to act and particularly they could type what they want without any constraint, and various behaviours were observed, users having sometimes completely abandoned their original task. Eventually, it appeared that many of the collected sentences were not really linked to the assistance domain (cf. table 1). Hence we got interested in trying to identify and categorize those other conversational activities that were appearing in the corpus.

4.1. Methodology: conversational activity annotation protocol

For this purpose, we have randomly extracted from the actually collected part of the Daft corpus (*i.e.* not the manually built up parts mentioned in 2.1.) two subsets of sentences, each subset having a size equal to the tenth of the total corpus size. The two subsets have been manually annotated by a single annotator, one after another in time. The first annotation process was used to refine the protocol (described below and summarized on figure 3), whereas the second one was strictly following it. Keeping in mind our objective is to study assistance, the first step was to know if the user was seeking help to accomplish tasks through its request, thus defining a first high level granularity distinction between:

1. **task-oriented** activities (ex: sentences[4] 1 – 9): where the user is working on the application to go towards the goal he has been given (independantly from knowing if he actually succeeds in getting closer from that goal).

[4]Sentence numbers always refer to the request examples given in table 1

2. **chat-oriented** activities (ex: sentences 10 – 14): where the user is interacting with the system for a reason that is not directly relevant to accomplish the task.

In the case of task-oriented requests, the user is either working directly on accomplishing its task or trying to get help from the system in order to do so. And in that former case, his request for help appears more or less obvious, thus leading us to distinguish three distinct activities:

1. **control** (sentences 1 – 3): direct *controls*, to make the agent interact himself directly with the application software in which it is embedded.

2. **direct assistance** (sentences 4 – 6): *help requests* explicitly made by the user.

3. **indirect assistance** (sentences 7 – 9): user's judgements concerning the application that would lead a human being to interpret them as a need for assistance; it certainly requires the system to use pragmatics to detect the implicit meaning.

For chat-oriented requests, although they are less relevant to our objective, we have been categorizing them according to the element of focus of the user's in its requests, distinguishing, in the cases where it is focused on the agent itself, replies to a previous agent utterance from a chat interaction started by the user. That makes a total of five different subactivities:

1. **reactions** to an agent's answer: a set of ways to agree or disagree to the agent's answer, marks of incredulity ("I don't think so"), lack of understanding ("You lost me") or insistence ("please answer to me").

2. **communicative functions**: this set is made of forms to start or end the communication with the agent ("hello", "bye", "I don't need your help anymore"...) as well as some phatic acts ("are you there?").

3. **dialogue** with the agent: sentences where the agent becomes user's focus, from orders ("Shut up!") to questions ("do you have a soul?") and from threats ("don't force me to kill you") to compliments ("you look cute").

4. **comments** about the application: comments without any assistance value ("This page looks nice").

5. **others**: a mix of the rest of the chat requests, not easy to classify with more details ("I'm an ordinary user", "I want to do a cognitive gift"...).

4.2. Results

No significant differences have been observed between both subsets annotated, allowing us to generalize the results obtained to the rest of the collected corpus : the figures 4 and 5 show the average distribution of requests from both subsets. Focusing on assistance, we can consider our collected corpus can be divided into four "subcorpora", each corresponding to a particular activity: control, direct and indirect assistance and chat.

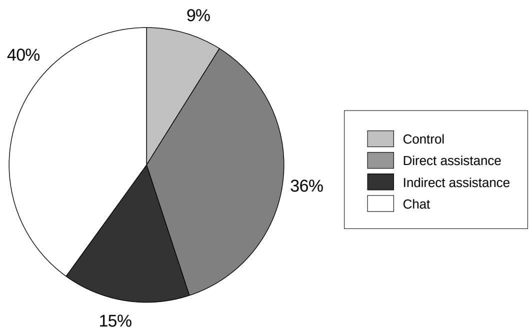

Figure 4: Daft corpus requests distribution of conversational activities

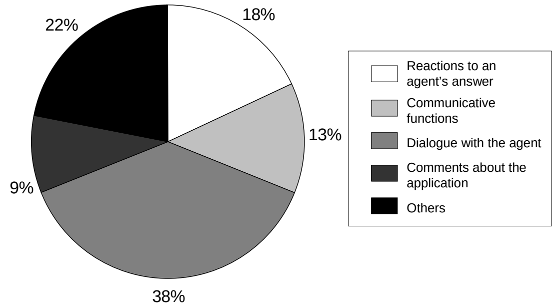

Figure 5: Chat subcorpus detailed distribution of conversational activities

The existence of the control subcorpus demonstrates that the user not only expects the agent to be able to assist him to use the application, but he also wants him to be able to act on this very application. The same way, the relative importance of chat-oriented requests, certainly related to the use of an embodied agent (*cf.* figure 1), shows that the user wants as well an agent able to react to comments not related to the task he is trying to carry out. Nonetheless, to really be able to deal properly with chat requests would require much more advanced dialogue skills: a wide range of vocabulary, personal life facts, an opinion about virtually anything, *etc.* We would thus be losing the methodological cut in complexity intended by focusing on a subdomain of natural language in the case of assistance. Finally, for the conception of our assisting agent, we take only into account control, (direct and indirect) assistance and reactions to an agent's answer activities, since that latter doesn't call into question the choice of dealing with isolated requests as it can easily be treated by only keeping in memory the previous assistance request.

4.3. Subcorpora comparison

In a similar way to what we have done in section 3., we can compare the four different subcorpora identified in section 4. within the Daft corpus by using interactional profiles methodology introduced in 3.2. (without any need for preliminary mapping though). Finding objective difference criteria like different speech acts distribution could indeed be useful for a potential automatic identification of a request activity, which would allow the agent to deal differently with control orders, assistance requests and chat utterances.

The results of that comparison are displayed on figure 6. Not only this analysis confirms the non-homogeneity of the Daft corpus (which average interactional profile is reminded in white), but it reveals a very clear difference, in terms of speech acts, between direct (mainly directives and some expressives) and indirect assistance (mainly assertives and expressives) requests. This result is particularly interesting because classical methods based on vocabulary or linguistic parameters (as described in details in (Bouchet 2007)) fail to discriminate efficiently those two kind of assistance: interactional profiles are hence perfectly complementary.

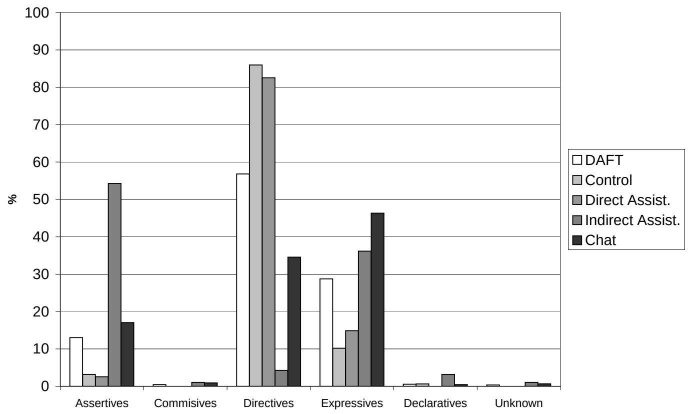

Figure 6: Interactional profiles comparison of Daft corpus' conversational activities

5. Conclusion and outlook

Using interactional profiles, we have shown that the Daft corpus was different from similar corpora in terms of speech acts distribution, certainly linked to the fact it is not human-human but human-computer interaction, thus confirming its necessity to study the Function of Assistance. Through a manual annotation of conversational activities within the Daft corpus, we have identified three assistance-related activities (control, direct and indirect assistance) representing 60% of the requests, the rest of them being chat-oriented. Finally, using again interactional profiles to compare the subcorpora defined by those activities, we managed to distinguish direct from indirect assistance requests.

Logical follow-up of this work shall focus on one side on getting a more accurate automatic identification of conversational activity (as a valuable first step analysis for the assisting agent), and on the other side on the formal modeling of those requests, particularly by taking into account the need for pragmatics in the case of indirect assistance.

Acknowledgements

The author particularly would like to thank anonymous reviewers for the methodological tips and references that have been helpful to improve the overall quality of this article.

References

R. Amalberti (1996). *La conduite des systèmes à risques*. PUF.

I. Androutsopoulos & M. Aretoulaki (2003). 'Natural Language Interaction'. In R. Mitkov (ed.), *The Oxford Handbook of Computational Linguistics*, pp. 629–649. Oxford University Press.

B. T. Atkins & H. M. A. Lewis (1996). *Language in Use*. Harper Collins Publishers.

F. Bouchet (2007). 'Caractérisation d'un Corpus de Requêtes d'Assistance'. In N. Hathout & P. Muller (eds.), *Actes de la 11e Rencontre des Etudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RECITAL 2007)*, vol. 2, pp. 357–366, Toulouse, France. IRIT Press.

A. Capobianco & N. Carbonell (2002). 'Contextual Online Help: a Contribution to the Implementation of Universal Access'. In S. Keates, P. J. Clarkson, & P. Robinson (eds.), *Universal Access and Assistive Technology*, pp. 131–140. Springer, London.

J. Carletta, et al. (1996). 'HCRC dialogue structure coding manual'. Tech. rep., HCRC, University of Edinburgh.

K. A. Ericsson & H. Simon (1993). *Protocol Analysis: Verbal Reports as Data*. MIT Press, Cambridge, MA, USA.

D. Jurafsky, et al. (1998). 'Switchboard Discourse Language Modeling Project Final Report'. Tech. rep., Center for Speech and Language Processing, Johns Hopkins University, Baltimore, MD, USA.

M. J. Kelly & A. Chapanis (1977). 'Limited Vocabulary Natural Language Dialogue'. *International Journal of Man-Machine Studies* **9**(4):479–501.

J. C. Lester, et al. (1997). 'The persona effect: affective impact of animated pedagogical agents'. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 359–366, Atlanta, Georgia, United States. ACM.

S. J. Molinsky & B. Bliss (1994). *Inventory of functions and conversation strategies: The Comprehensive course in functional English*, pp. 177–187. Prentice Hall.

M. Morel (1989). *Analyse linguistique de corpus*. Publications de la Sorbonne Nouvelle, Paris.

R. W. Morrell & D. C. Park (1993). 'The effects of age, illustrations, and task variables on the performance of procedural assembly tasks'. *Psychology and Aging* **8**(3):389–99. PMID: 8216959.

N. Ozkan (1994). *Vers un modèle dynamique du dialogue : analyse de dialogues finalisés dans une perspective communicationnelle*. Ph.D. thesis, INP Grenoble.

B. Post (2000). *Tonal and phrasal structures in French intonation*. Ph.D. thesis, Katholieke Universiteit Nijmegen.

G. Ripoche (2006). *Sur les traces de Bugzilla*. Ph.D. thesis, Université Paris-Sud 11.

J. Sansonnet, et al. (2005). 'Une architecture médiateur pour des agents conversationnels animés'. In *WACA'01 : Actes du Premier Workshop Francophone sur les Agents Conversationnels Animés*, pp. 31–39.

J. R. Searle (1969). *Speech Acts: An essay in the Philosophy of language*. Cambridge University Press, new edn.

D. P. Twitchell, et al. (2004). 'Using Speech Act Theory to Model Conversations for Automated Classification and Retrieval'. In *Proceedings of the 9th International Working Conference on the Language-Action Perspective on Communication Modeling*, pp. 121–129, New Brunswick, NJ. Rutgers University Press.