

INTELLIGENT ASSISTING CONVERSATIONAL AGENTS VIEWED THROUGH NOVICE USERS' REQUESTS

Mao Xuetao

Jean-Paul Sansonnet

François Bouchet

LIMSI-CNRS BP133 F-91403 Orsay cedex France
{mxt,jps,bouchet}@limsi.fr

ABSTRACT

Assisting Conversational Agents are Embodied Conversational Agents dedicated to the Function of Assistance for applications and services to the general public, especially on the Internet. We have developed a web-based framework to experiment with assisting agents regarding the key issue of believability, and where good Natural Language Understanding is a primary concern. Now, we are confronted with the difficult issue of the cost of developing and customizing Natural Language Processing tools (NLP-tools) for each new assisted application. In this paper, we propose an approach which is a tradeoff between complex dialogue systems and naive chatbot systems. We think that our approach is worth considering because it focuses on a concise and well circumscribed linguistic domain: the domain of Assistance Requests, that we captured by registering a corpus, in various contexts with ordinary people placed in front of Assisting Conversational Agents.

KEYWORDS

Assisting Conversational Agents, Eliciting User Requirements, Natural Language Processing.

1. INTRODUCTION

1.1 Context

This work is situated in the domain of the Assisting Conversational Agents (ACA), a subclass of Embodied Conversational Agents (Cassell et al., 1999-2000) which aims at bringing Natural Language & Artificial Intelligence-based assistance for ordinary users interacting with artifacts: software applications, services, smart objects, etc. (Maes, 1994). Here, we define an Assisting Conversational Agent as a software tool able to 1) interact in natural language with ordinary people and 2) use symbolic reasoning about the structure and the functioning of the assisted artifact. Indeed, associating such an assistant agent with a new product has long been considered as a good approach to improve their acceptability (Davis, 1989) and effectiveness (Lester et al., 1997), because natural language brings more naturalness in the interaction and symbolic reasoning brings more believability in the agent. This is even truer in the case of elderly people, who represent the majority of novice computer users, and feel more strongly than young children the complexity of tasks on computers (Czaja & Sharit, 1993). The loss of companions which happens when aging and is a major cause of depression for them, also increases their sensitivity to the Persona Effect, making them a primary target for ECA (cf. the European COMPANIONS project (Mival & Benyon, 2007)).

However, till now this approach has endured many setbacks, the "Clippy Effect" (Randall & Pedersen, 1998) being the most prominent. This phenomenon is consistent with the lack of use of help systems by novice users. This is the "motivation paradox" described in (Carroll & Rosson, 1987) which has led to the recent Contextual Help Systems approach (Capobianco & Carbonell, 2001-2002) in which the system tries to identify more closely the user's needs at the moment he/she requests some assistance, using models of the

users category (Shneiderman, 1992), the current task (Jameson, 2003), etc.; for a review of user modeling, see (Kobsa, 2001). Analyzing these acceptability issues, the bottom line is that one has to face a difficult dilemma to know if one should choose to build:

1) A complex custom dialogue system, like Allen's TRAINS for example (Allen et al., 1995). Such systems work fine (especially when used by corporate people) but they entail a critical cost effectiveness issue, mainly in terms of development duration and manpower linguistic skills. This analysis prompted Allen to promote *genericity* as a major challenge for future dialogue systems (Allen et al., 2001).

2) A naive, chatbot-like system, like Alice¹, Elbot², Jabberwacky³, etc. They are very cost effective and have been proven to be well-accepted by ordinary users, due to the "Eliza effect" (Weizenbaum, 1966) which makes ordinary users credit an agent with rough conversational abilities with much more smartness than it actually has. However they lack the symbolic reasoning capabilities and the fine semantics analysis capabilities required to support the Function of Assistance. This drawback has been analyzed by Wollermann (Wollermann, 2004-2006) over four main chatbots (Alice, EllaZ, Elbot, and Ultra-HAL⁴): a collection of linguistic phenomena were evaluated qualitatively in the chatbot answers to users questions, first on semantics aspects (semantic relations, quantifiers, and anaphora) and secondly on pragmatics with some Grice's maxims (1975). This study has shown that chatbots are failing in all these categories and that a deeper semantic/pragmatic analysis is required for finalized/task-oriented dialogue.

1.2 Proposition

In our work on assisting agents, we are currently exploring an ACA architecture which is a tradeoff between complex dialogue systems and naive chatbot systems. Relying on a bottom-up approach, a basic chatbot is provided with a) an improved Natural Language Processing (NLP) chain and b) reasoning heuristics over symbolic models (task, agent user and dialogue session). This typical conversational NLP-chain is illustrated in figure 2. Now, how can we expect to avoid both the pitfalls of huge dialogue systems and of naive chatbots? Our hope relies on the notion of *linguistic domain*: we think the Function of Assistance circumscribes a concise linguistic domain:

— *Circumscribes*: means that the linguistic domain can be distributionally contrasted against the general language domain and moreover against general dialogue activity (e.g. chat);

— *Concise*: means that the linguistic domain is restricted at the lexical semantics level, i.e. exhibits a small number of lexical semantic classes.

In order to validate this hypothesis, we need to register a corpus of Natural Language assistance requests, in various ACA-assisted applications, and to analyze the collected data a) to analyze the distribution of the conversational activities and b) to exhibit the lexical semantic classes required by the NLP-tools.

Section 2 describes the architecture of the NLP-chain and section 3 presents the collecting and the analysis of the semantic classes in the corpus of requests.

2. ARCHITECTURE

2.1 Objectives of the DIVA toolkit

The main objective of the DIVA toolkit consists in building a framework dedicated to the support of controlled experimentations upon ordinary people interacting with artifacts assisted by ACAs (e.g. to collect a corpus of natural language assistance requests, to register users' reactions, etc.). From this point of view, the Internet offers many advantages to reach ordinary users, which has led us to develop a web-based toolkit called DIVA (DOM Integrated Virtual Agents), supporting virtual characters completely integrated with the DOM (Document Object Model) tree structure of the webpages. The two main objectives of this toolkit are:

¹ Alicebot: <http://alicebot.blogspot.com/>

² Elbot, by Artificial Solutions: <http://www.elbot.com/>

³ Jabberwacky: <http://www.jabberwacky.com/>

⁴ Hal – Zabaware: <http://zabaware.com/>

1) An open programming framework making it easy and quick to develop and deploy experimental ACA in web-based applications; and 2) To take advantage of the new rich-client web 2.0 technologies to offer a full control of the interaction with the virtual characters.

The web architecture of DIVA is composed of two layers: 1) a symbolic-server layer dedicated to database resources management and symbolic computing, 2) a rich-client layer supporting: the specific application/service webpage; the animation of the graphic characters; the processing of the textual natural language interaction. Two examples of DIVA webpages are shown in figure 1, more examples are available on the DIVA homepage⁵.

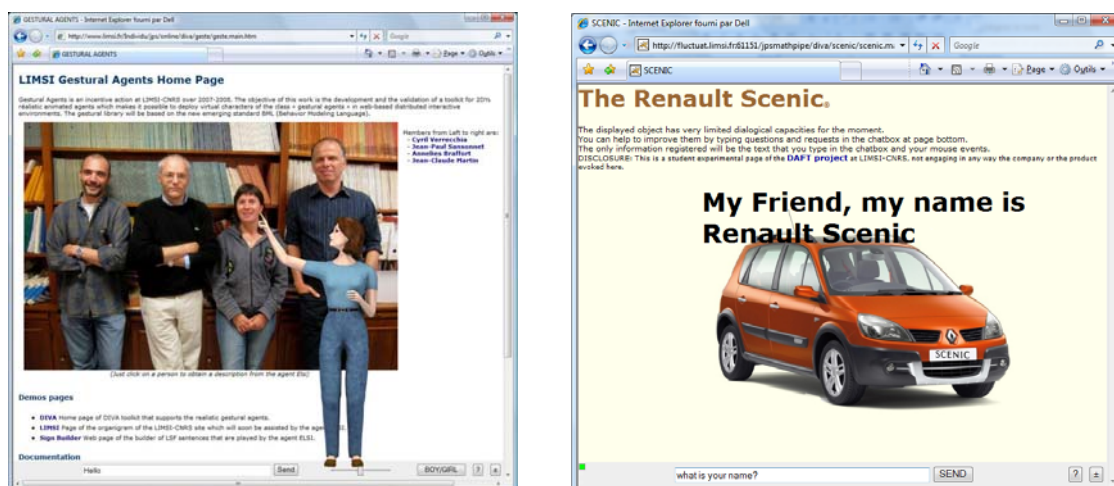


Figure 1: Two examples of DIVA-based Webpages: left: agent ELSI is performing a deictic gesture on a DOM object of the page⁶; right: people can chat with agent 'Scenic' of a well-known car corp.

2.2 NLP-Chain of the DIVA toolkit

The Natural Language Processing chain (NLP-chain) of the DIVA toolkit is detailed in figure 2. Like in most chatbots, the DIVA NLP-chain is based on pattern matching rules (in RegExp language) but it has a more sophisticated architecture, organized in two main phases with sub-phases:

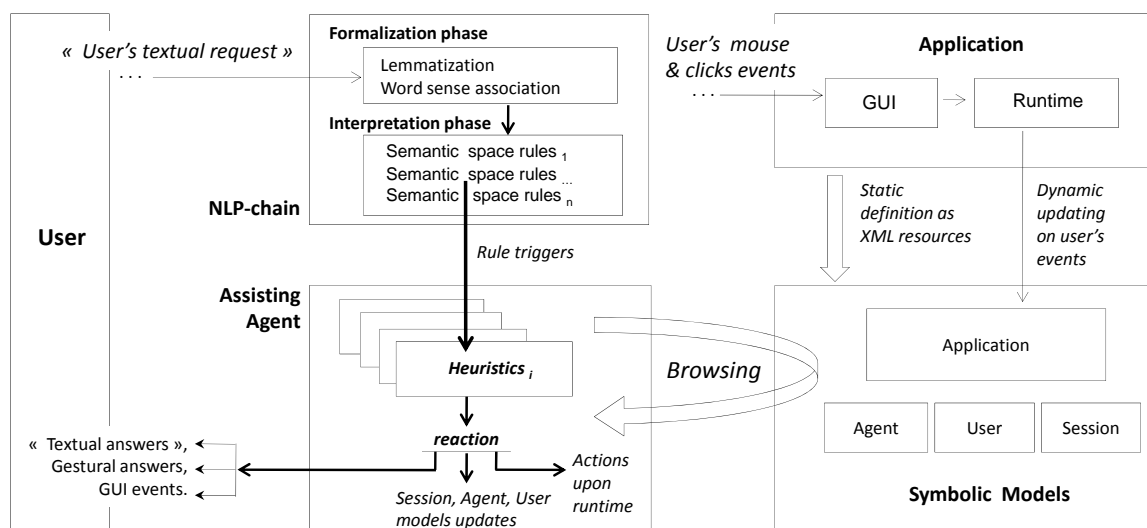


Figure 2: General structure of request handling in the DIVA assisting agent.

⁵ DIVA URL: <http://www.limsi.fr/Individu/jps/online/diva/divahome/index.html>

⁶ Gestural Agents URL: <http://www.limsi.fr/Individu/jps/online/diva/geste/geste.main.htm>

1) *The formalization phase*: it is based on two sets of filtering rules applied in sequential order:

- Syntactical level: a typical string pre-processing is followed by a lemmatization phase;
- Word-sense association level: lemmas are then transformed into semantic classes or ‘synsets’ as in Wordnet (Fellbaum, 1998).

At the end of the formalization phase, the request is transformed into an intermediate formal form, called the Formal Request Form (FRF).

2) *The interpretation phase*: it is based on a set of rules of the form <pattern → reaction> where patterns are applied to FRF expressions and reactions are procedural heuristics defining the behavior of the agent in response to the user’s requests.

Here are two examples of users’ requests translated into FRF:

REQ₁ = “If I want to buy such a car, what can I do?”

FRF₁ = < QUEST IF THEUSER TOWANT TOOBTAIN such a car WHAT TOCAN THEUSER TODO >

The filtering process has extracted 9 synsets (uppercase symbols — their gloss is given in table 4) from REQ₁ that are put in FRF₁. Some lemmas have no associated synsets because they are not in the generic ontology (e.g. ‘car’).

REQ₂ = “Adopt a less provocative attitude, please.”

FRF₂ = < TOTAKE a LESSTHAN ISUNPLEASANT THEBELIEF TOSAYPLEASE >

The interpreting phase is organized into several layers, called ‘semantic spaces’ or in short ‘spaces’. Most spaces are dedicated to a generic conversational domain, making it easy to share and reuse them from an experiment to another. Each semantic space contains a set of rules that defines a behavior of the agent. For example, assume that the user gives his/her name to the agent:

“My name is Jane” → < USERNAME BE jane >

Code: Interpretation rule handling the name of the user

```
<rule id="on-username" pat="&lt; USERNAME BE (\w+) &gt;" >
  <do>
    THETOPIC.x = TALK_capitalizefirst(TALK_getmatch(1));
    if (THETOPIC.x == THEUSER.name) { // one of the strings in [] is chosen randomly

      TALK_say(['I knew it already', 'You said it !']);
      TALK_decrease( THEAGENT.coop, 0.1); // decreases by 10%
      TALK_exit(); // prevents tag <say> from being launched
    } else { THEUSER.name = THETOPIC.x; TALK_increase( THEAGENT.coop, 0.5); }
  </do>
  <say> // one of the tags <p> is chosen randomly
    <p>From now, I will call you _THETOPIC.name_.</p>
    <p>Ok, your name is _THETOPIC.name_...</p>
    <p>OK you are _THETOPIC.name_.</p>
  </say>
</rule>
```

The <say> tag makes use of the meta-variable `_THETOPIC.name_` thus producing for example “*From now on I will call you Jane.*”

We can see from this example that to be able to build a reaction, the agent requires a *knowledge base* registering the relevant assistance information about the application, but also about the agent and the user (e.g. to store the user’s age in the above example). In DIVA, the symbolic information about the assisted application is stored in its so-called *topic* XML-file. For example, here is a topic file (abbreviated version) of a well-known car, personified as a DIVA agent, as shown in figure 1 – right:

Code: Excerpt from the topic file of the Scenic agent

```
<?xml version="1.0" encoding="utf-8"?>
<topic id="TOPICSCENIC">
  <ObjClass>artefact</ObjClass>
  <objSubClass>car</objSubClass>
  <ObjName>Renault Scenic</ObjName>
  <objWeight unit="kg">2205</objWeight>
  <objPrice unit="€" val="range">[2600.00-200, 2600+200]</objPrice>
  <objSpeed unit="kms">40</objSpeed>
</topic>
```

3. CORPUS ANALYSIS

3.1 Corpus objectives and collection

For some time now, we have been able to make several experiments with ordinary people interacting in a textual manner in the specific context of Web-based conversational assisting agents (see table 1):

- AMI is an experimental interactive website (describing a research team) that can be edited in a dialogical way with agent Lea;
- MARCO, an agent providing information on the active website of the French Research group on Animated Conversational Agents GT ACA (<http://www.limsi.fr/aca/>);
- TICKET is a webpage about a “book ticket service” where the users can dialogue with the agent Elsi;
- SCENIC is a webpage giving interactive information about a popular French car. In this experiment, the agent is not a virtual character but a picture of the car itself (cf. Figure 1— right);
- MAO-XUETAO is a webpage giving interactive information about an ordinary Chinese PhD student. In this experiment, the agent is personified by a picture of the person;
- MAO-ZEDONG is a webpage giving interactive information about a famous person. The agent is also personified by a picture of the person;

Table 1. Description of the context for the collecting of six corpora in various assisting situations.

Corpus name	Language	Subjects (adult M/F)	Utterances	Platform
AMI	French	15 M/F	500 (with additional small experiments)	Daft
MARCO	French	40 M/F*	275	WebLea
TICKET	English	4 M / 2 F	90	DIVA
SCENIC	English	4 M / 6 F	96 (company) +125 (car)	DIVA
MAO-XUETAO	English	4 M / 5 F	282	DIVA
MAO-ZEDONG	English	1 M	282**	DIVA

* MARCO was anonymously collected among the GT ACA members (adult M/F colleagues) with about 40 participants.

** This data was obtained by manual transposition of the MAO-XUETAO utterances into the MAO-ZEDONG context.

AMI and MARCO corpora were registered with French speaking subjects while other experiments were registered with English speaking subjects (in order to provide at least two languages as the DIVA textual input modality). Three software platforms, successively developed by our team, have been used: DAFT a Java applet-based 2D-Cartoon ACA; WEBLEA a JavaScript-based 2D-Cartoon ACA and DIVA a JavaScript-based 3D-realistic ACA (WEBLEA and DIVA are available online). After the translation of the French corpora into English and the dropping of repeated utterances we got the final global corpus with 1612 items. Table 2 shows selected excerpts of the collected data.

More than half of the users’ requests are not syntactically well-formed or exhibit idiosyncratic forms (this is emphasized in bold in table 2): for example, expressions from the spoken language, spelling, syntactic or grammatical mistakes, acronyms from SMS and internet slang etc. Some of those forms are not easy to detect and to fix with classical NLP tools, but it is possible to analyze them correctly by handling those specificities.

The requests are collected from the subjects and stored as a list of utterances in order to be processed by the NLP tools (sentence analysis and building of statistic resources files for POS or WSD resolution).

The lexical distribution of the corpus is statistically different from generalist textual data, again making it difficult to use classical NLP-tools. Actually we have made comparative studies with two linguistic domains:

- 1) Narrative texts taken from newspaper “Le Monde” as available in the Multitag corpus (Paroubek, 2000)
- 2) Three dialogical corpora: Switchboard (Jurafsky, 1997), MapTask (Thompson et al., 1993) and the bugs reports management traces of Bugzilla (<https://bugzilla.mozilla.org>) analyzed by (Bouchet, 2006).

These two studies (Bouchet, 2007) showed that the distribution of the corpus or requests is specific on several variables: phrase length, lemma lexicon, lexical semantic classes, speech acts and speech act classes.

Table 2: Selected excerpt from the collected data (Daft corpus).

N°	Original collected request (in French)	Translation in English (including mistakes in bold)
1	cliques le bouton quitter	clicks the quit button
2	cliquesur le bouton retour	clickon the back button
3	ok, reviens à l apage d'accueil	ok, come back to th ehomepage
4	donne moi un plan du site	give me a map of the website
5	à quoi sert cette fenêtre,	what is this window for,
6	c koi le GT ACA	WDYM by GT ACA
7	est-ce que le bouton "fermer" et le bouton "quitter" fonctionnent exactement pareil ?	do the "close" button and the "quit" button work exactly the same way?
8	j'ai une question à poser à un des membres, comment je peux le joindre?	I have a question to ask to one of the members, how can I contact him?
9	quand est la prochaine reunion ?	when is the next meeting?
10	où peux t-on trouver le programme de la conférence?	where can-we find the conference schedule?
11	existe t-il une version condensée de l'aide	is there a shorten version of the help
12	je ne vosi aucune page de demso !!	I cna't see any demso page!!
13	le lien me semble cassée	the link seems to me to be broken
14	j'ai été vraiment surpris de constater qu'il manque une fonction d'annulation globale	I was really surprised to see there's no global cancel function
15	ça serait quand même mieux si on pouvait aller directement au début	it'd be better to be able to go directly at the beginning
16	auf viedersen	auf viedersen
17	espèce de bon à rien !	you good for nothing!
18	Quel genre de musique tu aimes ?	What kind of music do you like?
19	tu t'habilles tous les jours de la me me facon ?	you're dressing the same way everyday?
20	ca marche :-)	works for me :-)

3.2 Analysis of the corpus

In this section, we develop the analysis of the collected corpus on two specific key-points: a) the classification of the conversational activities through the distribution of the speech acts; b) the manual exhibition of the lexical semantic classes from the distribution of the lemma.

3.2.1 Characterization of the conversational activities

During the corpus collection phase, the subjects were requested to do some tasks for which they could ask help (if needed) from an artificial assistant agent embedded in the program to assist them. Subjects were completely free to act and particularly they could type whatever they wanted without any constraint. Consequently, various behaviors have been observed, with users sometimes completely abandoning their original task, and it eventually appeared that many of the collected sentences were not really linked to the assistance domain itself. Hence we got interested in trying to identify and categorize the conversational activities that are actually appearing in the corpus. This allowed us to distinguish four main classes of conversational activities with unequal distribution as shown in table 3. A more detailed analysis of the Daft corpus can be found in (Bouchet, 2007).

Table 3. Description of the conversational activities.

Activity	% in corpus	Description	Lines in Table 2
Control	9%	Utterances prompting the agent itself to operate upon the application software in which it is embedded.	1-4
Direct Assistance	36%	Utterances expressing directly a need for help.	5-11
Indirect Assistance	15%	Utterances containing user's judgments concerning the application that are actually implicitly expressing the fact that the user is actually in need of assistance (can be elicited by pragmatic reasoning).	12-15
Chat	40%	Utterances related to other conversational activities. They are often oriented towards the "agent as a person" instead of the application.	16-20

As we are mainly interested in control and assisting activities, one can view chat activity as mere 'noise'. Actually this class cannot easily be discarded because a) it represent 40% of the corpus and b) control and assisting activities are in fact deeply embedded within chat activity. Indeed, according to the "Eliza effect" (Chalmers, 1992) the personification of the assisting agent by a virtual character prompts the users to enter into affective chat activity, which could be seen as some sort of assistance, not regarding the assisted application, but a more psychological support one might be seeking from a companion agent.

3.2.2 Extraction of the lexical semantic classes

As shown in figure 2, we use an intermediate form between the syntactic layer and the interpretative layer. This form is the result of the translation of the word lemmas into their associated semantic classes, also called semantic keys (or in short *keys*): we carried out a qualitative analysis upon this corpus in order to exhibit the occurring lemmas and to group them as synsets (like in Wordnet), i.e. semantic classes represented by a unique keyword. The total number of semantic exhibited keys is 436, divided into six main classes:

CLASS NAME	KEY NUMBER	SYMBOL PREFIX
NAMES LIST	132	THE-
CATEGORIES LIST	20	AKO-
VERBS LIST	115	TO-
ADJECTIVES LIST	60	IS-
LOCATIONS LIST	23	none
GRAMMATICALS & SPEECH ACTS LIST	86	none

Table 4: Excerpt from the 436 semantic keys of DIVA. (in **bold-red**: keys used in examples of section 2.2)

Keys	Gloss (shortened)	Keys	Gloss (shortened)
TOWORK	Denotes the general activity of achieving some work	ISHONEST	Denotes the quality of somebody who is honest/sincere
TOKNOW	Denotes the mental action of knowing something	ISFEMALE	Denotes the quality of a person with gender: female
TOTAKE	Denotes the action of taking/receiving something	ISUNPLEASANT	Denotes the quality of something that is unpleasant
TODO	Denotes the action of doing something	ISMANDATORY	Denotes the quality of something that is legally/physically mandatory/indispensable
TOCAN	Denotes the abstract action of having the general capacity or right of doing something	WHAT	Denotes the grammatical WH-pronoun: what
TOSAYPLEASE	Denotes the expression of saying please to somebody	WHY	Denotes the grammatical relation: why
TOSPEAK	Denotes the action of speaking	QUANTSMALL	Denotes the quantity: a small quantity/ amount
TOWANT	Denotes the mental action of desiring/wanting something or a state of affairs to happen	NEG	Denotes the grammatical relation: negation
TOSAYHELLO	Denotes the expression of greeting somebody	IF	Denotes the grammatical relation: if
TOOBTAIN	Denotes the general action of obtaining/acquiring something or some information	QUEST	Denotes the grammatical relation: question
THEAVATAR	Denotes the graphical/dialogical assisting character of the application	UNDEFPRON	Denotes the grammatical pronoun: one
THEHELP	Denotes the service/help provided by somebody	LESSTHAN	Denotes the quality of something that is less (according to some quality) than another thing
THEMAXIMUM	Denotes the maximum value that a variable can take	IT	Denotes the grammatical pronoun: it
THEUSER	Denotes the user at first person: I, me, myself	TOBE	Denotes the grammatical auxiliary verb: to be
THEBELIEF	Denotes the name of the mental action of believing/supposing something		

4. CONCLUSION AND PERSPECTIVES

A first objective was to provide an open, web-based framework for experimentations on Assisting Conversational Agents: the DIVA toolkit is now operational in version 1.0. It is downloadable from the DIVA home page and can be freely used for research and teaching purposes. As discussed in section 1.2 the architecture has proved to meet the goal of cost-effectiveness (see the website DIVA exhibiting various assisted applications).

Also, using a first version of the toolkit, we have been able to collect a corpus of assisting requests, making it possible to study with some accuracy the distribution of the conversational activities and the lexical semantics extension of the Function of Assistance as a linguistic domain in its own right, especially in the context of small web-based applications. The small number of exhibited semantic classes shows that the related linguistic domain is a reality, it is characterizable and above all it is tractable for further processing by a NLP-chain.

Presently, a first version of the complete NLP-chain is operational, sufficient to build ongoing experiments on assistance with human subjects. However, the FRF remains a ‘flat’ sequence of semantic keys that will show its limits when we will develop larger web-applications, entailing a broader semantic domain (even several sub-domains). Therefore, the next version of the NLP-chain will rely on a more structured formal request language, currently under development.

REFERENCES

- Allen, J.F. et al., 1995. The TRAINS project: a case study in building a conversational planning agent. *Journal of Experimental & Theoretical Artificial Intelligence*, 7(1), 7-48.
- Allen, J.F. et al., 2001. Towards Conversational Human-Computer Interaction. *AI Magazine*, 22(4), 27-37.
- Bouchet, F., Sansonnet, J-P. Caractérisation de Requêtes d'Assistance à partir de corpus. Conference MFI'07, 2007
- Bouchet, F., Sansonnet, J-P. Etude d'un corpus de requêtes en langue naturelle pour des agents assistants. WACA 2006 : Actes du Second Workshop Francophone sur les Agents Conversationnels Animés, Toulouse, 2006
- Capobianco, A. & Carbonell, N., 2001. Contextual online help: elicitation of human experts' strategies. In *Proceedings of HCI'01*. New Orleans, pp. 824-828.
- Capobianco, A. & Carbonell, N., 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*. London: Springer, pp. 131-140.
- Carroll, J.M. & Rosson, M.B., 1987. Paradox of the active user. In *Interfacing thought: cognitive aspects of human-computer interaction*. MIT Press, pp. 80-111.
- Cassell, J. et al., 1999. Embodiment in conversational interfaces: Rea. In *CHI'99: Proceedings of the SIGCHI conf. on Human factors in comp. syst.* New York, NY, USA: ACM Press, pp. 520-527.
- Cassell, J. et al. eds., 2000. *Embodied Conversational Agents*, MIT Press.
- Chalmers, D.J., French, R.M. & Hofstadter, D.R., 1992. High-level perception, representation, and analogy: A critique of artificial intelligence methodology. *Journal of Experimental & Theoretical Artificial Intelligence*, 4(3), 185-211.
- Czaja, S.J. & Sharit, J., 1993. Age differences in the performance of computer-based work. *Psychology and Aging*, 8(1), 59-67.
- Davis, F.D., 1989. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13(3), 319-340.
- Fellbaum, C., 1998. *WordNet: An Electronic Lexical Database*, Cambridge, MA: MIT Press.
- Grice, H.P., 1975. Logic and Conversation. *Syntax and semantics*, 3, 41-58.
- Jameson, A., 2003. Adaptive interfaces and agents. In *The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications*. Lawrence Erlbaum Associates, Inc., pp. 305-330.
- Jurafsky, D. et al., 1998. *Switchboard Discourse Language Modeling Project Final Report*, Baltimore, MD, USA: Center for Speech and Language Processing, Johns Hopkins University.
- Kobsa, A., 2001. Generic User Modeling Systems. *User Modeling and User-Adapted Interaction*, 11(1), 49-63.
- Lester, J.C. et al., 1997. The persona effect: affective impact of animated pedagogical agents. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. Atlanta, Georgia, United States: ACM, pp. 359-366.
- Maes, P., 1994. Agents that reduce work and information overload. *Commun. ACM*, 37(7), 30-40.
- Mival, O. & Benyon, D., 2007. Introducing the COMPANIONS project: Intelligent, persistent, personalised multimodal interfaces to the internet. In C. Sas & T. Ormerod, eds. *Proc. of the 21st British HCI Group Annual Conference*.
- Paroubek, P., 2000. Language Resources as by-Product of Evaluation: The MULTITAG Example. In *LREC'2000: Proceedings of 2nd International Conference on Language Resources & Evaluation*. pp. 151-154.
- Randall, N. & Pedersen, I., 1998. Who exactly is trying to help us? The ethos of help systems in popular computer applications. In *Proceedings of the 16th annual international conference on Computer documentation*. Quebec, Quebec, Canada: ACM, pp. 63-69.
- Shneiderman, B., 1992. *Designing the user interface (2nd ed.): strategies for effective human-computer interaction*, Addison-Wesley Longman Publishing Co., Inc.
- Thompson, H.S. et al., 1993. The HCRC Map Task corpus: natural dialogue for speech recognition. In *Proceedings of the workshop on Human Language Technology*. Princeton, New Jersey: Assoc. for Computational Linguistics, pp. 25-30.
- Weizenbaum, J., 1966. ELIZA: a computer program for the study of natural language communication between man and machine. *Commun. ACM*, 9(1), 36-45.
- Wollermann, C., 2004. *Evaluierung der linguistischen Fähigkeiten von Chatbots*. Masters thesis. Rheinische-Friedrich-Wilhelms Universität Bonn.
- Wollermann, C., 2006. Position paper. In *Proceedings of Young Researchers' Roundtable on Spoken Dialogue Systems*. Pittsburgh, PA, pp. 75-76.