COOPML: Towards Annotating Cooperative Discourse

Farah Benamara, Véronique Moriceau, Patrick Saint-Dizier
IRIT
118 route de Narbonne
31062 Toulouse cedex France
`benamara, moriceau, stdizier@irit.fr`

Abstract

In this paper, we present a preliminary version of COOPML, a language designed for annotating cooperative discourse. We investigate the different linguistic marks that identify and characterize the different forms of cooperativity found in written texts from FAQs, Forums and emails.

1 What are cooperative responses and why annotate them ?

Grice (Grice, 1975) proposed a number of maxims that describe various ways in which speakers are engaged in a cooperative conversation. Human conversations are governed by implicit rules, used and understood by all conversants. The contents of a response can be just direct w.r.t. the question literal contents, but it can also go beyond what is normally expected, in a relevant way, in order to meet the questioner's expectations. Such a response is said to be cooperative.

Following these maxims and related works, e.g. (Searle, 1975), in the early 1990s, a number of forms of cooperative responses were identified. Most of the efforts in these studies and systems focussed on the foundations and on the implementation of reasoning procedures (Gal, 1988), (Minock et ali., 1996), while little attention was paid to question analysis and NL response generation. An overview of these systems can be found in (Gasterland et al., 1994) and in (Webber et ali., 2002), based on works by (Hendrix et ali., 1978), (Kaplan, 1982), (Mays et ali., 1982), among others. These systems include e.g. the identification of false presuppositions and various types of misunderstandings found in questions. They also include reasoning schemas based e.g. on constant relaxation to provide approximate or alternative, but relevant, answers when the direct question has no response. Intensional reasoning schemas can also be used to generalize over lists of basic responses or to construct summaries.

The framework of Advanced Reasoning for

Question Answering (QA) systems, as described in a recent road map, raises new challenges since answers can no longer be only directly extracted from texts (as in TREC) or databases, but requires the use of a domain knowledge base, including a conceptual ontology, and dedicated inference mechanisms. Such a perspective, obviously, reinforces and gives a whole new insight to cooperative answering. For example, if one asks [1]:

Q4: Where is the Borme les Mimosas cinema ?
if there are no cinema in Borme les Mimosas, it can be responded:
R4: There is none in Borme, the closests are in Londe (8kms) and in Hyeres (20kms),
where close-by alternatives are proposed, involving relaxing Borme, identified as a village, into close-by villages or towns that respond to the question, evaluating proximity, and finally sorting the responses, e.g. by increasing distance from Borme. This simple example shows that, if a direct response cannot be found, several forms of knowledge, reasoning schemas and strategies need to be used. This is one of the major challenges of advanced QA. Another challenge, not yet addressed, is the generation of the response in natural language.

Our first aim is to study, via corpus annotations, how humans deploy cooperative behaviours and procedures, by what means, and what is the form of the responses provided. Our second aim is to construct a linguistically and cognitively adequate formal model that integrates language, knowledge and inference aspects involved in cooperative responses. Our assumption is then that an automatic cooperative QA system, although much more stereotyped than any natural system, could be induced from natural productions without loosing too much of the cooperative contents produced by humans.

From that point of view, the results presented in this paper establish a base for investigating cooperativity empirically and not only in an abstract and

---

[1] Our corpora are in French, but, whenever possible we only give here English glosses for space reasons

introspective way. Our goal is to get a kind of empirical testing and then model for cooperative answering, to get clearer ideas on the structure of cooperative discourse, the reasoning processes involved, the types of knowledge involved and the NL expression modes.

2 Related work

Discourse annotation is probably one of the most challenging domains that involves almost all aspects of language, from morphology to pragmatics. It is of much importance in a number of areas, besides QA, such as MT or dialogue. A number of discourse annotation projects (e.g. PALinkA (Orasan, 2003), MULI (Baumann et ali., 2004), DiET (Netter et ali. 1998), MATE (Dybkjaer et ali., 2000)) mainly deal with reference annotations (be they pronominal, temporal or spatial), which is clearly a major problem in discourse. Discourse connectives and their related anaphoric links and discourse units are analyzed in-depth in PDTB (Miltasakaki et ali. 2004), a system now widely used in a number of NL applications. RST discourse structures are also identified in the Treebank corpora.

All these projects show the difficulty to annotate discourse, the subjectivity of the criteria for both the bracketing and the annotations. Annotation tasks are in general labor-intensive, but results in terms of discourse understanding are rewarding. Customisation to specific domains or forms of discourse and the definition of test-suites are still open problems, as outlined in PDTB and MATE.

Our contribution is more on the pragmatic side of discourse, where there is little work done, probably because of the complexity of the notions involved and the difficulty to interpret them. Let us note (Strenston, 1994) that investigates complex pragmatic functions such as performatives and illocutionary force. Our contribution is obviously inspired by abstract and generic categorizations in pragmatics, but it is more concrete in the sense that it aims at identifying precise cooperative functions used in everyday life in large-public applications. In a first stage, we restrict ourselves to written QA pairs such as FAQ, Forums and email messages, which are quite well representative of short cooperative discourses (see 3.1).

3 A typology of cooperative functions

The typology below clearly needs further testing, stabilization and confirmation by annotators. However, it settles the main lines of cooperative discourse structure.

3.1 Typology of corpora

To carry out our study and subsequent evaluations, we considered three typical sources of cooperative discourses: Frequently Asked Questions (FAQ), Forums and email question-answer pairs (EQAP), these latter obtained by sending ourselves emails to relevant services (e.g. for tourism: tourist offices, airlines, hotels). The initial study was carried out on 350 question-answer pairs. Note that in the tourism domain, FAQ are rather specific: they are not ready-made, prototypical questions. They are rather unstructured sets of questions produced e.g. via email by standard users. From that point of view, they are of much interest to us.

We have about 50% pairs coming from FAQ, 25% from Forums and 25% from EQAP. The domains considered are basically large-public applications: tourism (60%, our implementations being based on this application domain), health (22%), sport, shopping and education. In all these corpora, no user model is assumed, and there is no dialogue: QA pairs are isolated, with no context. This is basically the type of communication encountered when querying the Web. Our corpus is only composed of written texts, but these are rather informal, and quite close in style to spoken QA pairs.

FAQ, Forum and EQAP cooperative responses share several similarities, but have also some differences. Forums have in general longer responses (up to half a page), whereas FAQ and EQAP are rather short (from 2 to 12 lines, in general). FAQ and Forums deal with quite general questions while EQAP are more personal. EQAP provided us with a very rich material since they allowed us to get responses to queries in which we have deliberately introduced various well identified errors and misconceptions. In order to have a better analysis of how humans react, we sent those questions to different, closely related organizations (e.g. sending the same ill-formed questions to several airlines). FAQ, Forums and EQAP also contain several forms of advertising, and metalinguistic parameters outlining e.g. their commercial dimensions.

From the analysis of 350 of QA pairs, taking into account the formal pragmatics and artificial intelligence perspectives, we have identified the typology presented below, which defines the first version of COOPML.

3.2 Cooperative discourse functions

We structure cooperative responses in terms of cooperative functions, which are realized in responses by means of meaningful units (MU). An MU is the smallest unit we consider at this level; it conveys a

minimal, but comprehensive and coherent fragment of information. In a response, MUs are connected by means of transition units (TU), which are introductory or inserted between meaningful units. TUs define the articulations of the cooperative discourse.

In a cooperative discourse, we distinguish three types of MU: direct responses (DR), cooperative know-how (CSF) and units with a marginal usefulness (B) such as commentaries (BC), paraphrases (BP), advertising, useless explanations w.r.t. to the question. These may have a metalinguistic force (insistence, customer safety, etc) that we will not examine in this paper. DR are not cooperative by themselves, but they are studied here because they introduce cooperative statements. Let us now present a preliminary typology for DR and CSF, between parentheses are abbreviations used as XML labels.

Direct responses (DR): are MUs corresponding to statements whose contents can be directly elaborated from texts, web pages, databases, etc., possibly via deduction, but not involving any reformulation of the original query. DR include the following main categories:

- Simple responses (DS): consisting of yes/no forms, modals, figures, propositions in either affirmative or negative form, that directly respond the question.

- Definitions, Descriptions (DD): usually text fragments defining or describing a concept, in response to questions e.g. of the form *what is 'concept'?*.

- Procedures (DP): that describe how to realize something.

- Causes, Consequences, Goals (DCC): that usually respond to questions in *Why/ How?*.

- Comparisons and Evaluations (DC): that respond to questions asking for comparisons or evaluations.

This classification is closely related to a typology of questions defined in (Lehnert, 1978).

Responses involving Cooperative Know-how (CSF) are responses that go beyond direct answers in order to help the user when the question has no direct solution or when the question contains a misconception of some sort. These responses reflect various forms of know-how deployed by humans. We decompose them into two main classes: **Response Elaboration** (ER) and **Additional Information** (CR). The first class includes response units that propose alternative responses to the question whereas the latter contains a variety of complements of information, which are useful but not absolutely necessary. ER are in a large part inspired from specific research in Artificial Intelligence such as constraint relaxation and intensional calculus.

Response elaboration (ER) includes the following MUs:

- Corrective responses (CC): that explain why a question has no response when it contains a misconception or a false presupposition (formally, a domain integrity constraint or a factual knowledge violation, respectively), For example: Q5: *a chalet in Corsica for 15 persons?* has no solution, a possible response is: R5a: *Chalets can accomodate a maximum of 10 persons in Corsica.*

- Responses by extension (CSFR): propose alternative solutions by relaxing a constraint in the original question. There are several forms of relaxations, reported in (Benamara et al. 2004a), which are more subtle than those developed in artificial intelligence. For example, we observed relaxation on cardinality, on sister concepts or on remote concepts with similar prominent properties, not studied in AI, where relaxation operates most of the time on the basis of ancestors.

  Response R5a above can then be followed by CSFRs of various forms such as: R5b: *we can offer (1) two-close-by chalets for a total of 15 persons, or (2) another type of accomodation in Corsica: hotel or pension for 15 persons*. Case (1) is a relaxation on cardinality (duplication of the resource) while (2) is a relaxation that refers to sisters of the concept chalet.

- Intensional responses (CSFRI): tend to abstract over possibly long enumerations of extensional responses in order to provide a response at the best level of abstraction, which is not necessarily the highest. For example, Q6: *How can I get to Geneva airport ?* has the following response: R6a: *Taxis, most buses and all trains go to Geneva airport*. This level is prefered to the more general but less informative response R6b: *Most public transportations go to Geneva airport*.

- Indirect responses (CSFI): provide responses which are not direct w.r.t. the question (but which may have a direct response), e.g.: *is your camping close to the highway?*, can be

indirectly, but cooperatively answered:
yes, but that highway is quiet at night.. A direct response would have said, e.g.: *yes, we are only 50 meters far from the highway*, meaning that the camping is of an easy access.

- Hypothetical responses (CSFH): include responses based on an hypothesis. Such responses are often related to incomplete questions, or questions which can only be partly be answered for various reasons such as lack of information, or vague information w.r.t the question focus. In this case, we have a QA pair of the form: Q7: *Can I get discounts on train tickets ?* R7: *You can get a discount if you are less than 18 years old or more than 65, or if you are travelling during week-ends.*

- Clustered, case or comparative responses (CSFC): which answer various forms of questions e.g. with vague terms (e.g. *expensive, far from the beach*). For example, to Q8: *is the hotel Royal expensive?* it is answered: R8: *for its category (3*) it is expensive, you can find 4* hotels at the same rate.*

The most frequent forms of responses are CSFR, CSFI, CSFC, CSFRI; the two others (CC and CSFH) are mainly found in email QA.

Additional Information units (CR) contain the following cases:

- precisions of various forms, that deepen the response (AF): this 'segment' or 'continuum' of forms ranges from minor precisions and generalizations to elaborated comments, as in Q9: *Where can I buy a hiking trail map of Mount Pilat ?* which has the response R9 that starts by an AF: R9: *The parc published a 1:50 000 map with itineraries,... this map can be bought at bookshops....*

- restrictions (AR): restrict the scope of a response, e.g. by means of conditions: Q10: *Do you refund tickets in case of a strike ?* R10: *yes, a financial compensation is possible provided that the railway union agrees....*

- warnings (AA): warn the questioner about possible problems, annoyances, dangers, etc. They may also underline the temporal versatility of the information, as it is often the case for touristic resources (for example, hotel or flight availability),

- justifications (AJ): justify a negative, unexpected or partial response: Q11: *Can I be refunded if I loose my rail pass ?*, R11: *No, the rail pass fare does not include any insurance against loss or robbery.*

- concessives (AC): introduce the possibility of e.g. exceptions or specific treatments: *Children below 12 are not allowed to travel unaccompanied, however if a passenger is willing to take care about him....*

- suggestions - alternatives - counter-proposals (AS): this continuum of possibilities includes the proposition of alternatives, more or less marked, when the query has no answer, in particular via the above ER. Q12: *Can I pay the hotel with a credit card?*, R12: *yes, but it is preferable to have cash with you: you'll get a much better exchange rate and no commission.*

The different MU have been designed with no overlap, it is however clear that there may have some forms of continuums between them. For example, CSFR, although more restricted, may be viewed as an AS, since an alternative, via relaxation, is proposed. We then would give preference to the CSF group over the CR, because they are more precise.

A response does not involve more, in general, than 3 to 4 meaningful units. Most are linearly organized, but some are also embedded. At the form level, response units of CSF (ER and CR) have in general one or a combination of the following forms: adverb or modal (RON), proposition (RP), enumeration (RE), sorted response (via e.g. scalar implicature) (RT), conditionals (RC) or case structure (RSC). These forms may have some overlap, e.g. RE and RT.

3.3 Annotating Cooperative Discourse: a few illustrations

Fig. 1 (next page) presents three examples annotated with COOPML.

3.4 Identifying cooperative response units

The question that arises at this stage is the existence of linguistic markers that allow for the identification of these response units. Besides these markers, there are also constraints on the organization of the cooperative discourse in meaningful units. These are essentially co-occurrence, incompatibility and precedence constraints. Finally, it is possible to elaborate heuristics that give indications on the most frequent combinations to improve MU automatic identification.

In the following subsections we first present a typology for MU delimitation, then we explain how direct responses (DS) are identified, mainly, via the

Figure 1: Discourse annotation

domain ontology whose structure and contents is presented. We end the section by the linguistic marks that identify a number of additional information units (CR).

3.4.1 Typology of MU delimitators

Identifying meaningful response units consists in two tasks: exploring linguistic criteria associated with each form of cooperative response unit and finding the boundaries of each unit. Cooperative discourse being in general quite straightforward, it turns out that most units are well delimited naturally: about 70% of the units are single, complete sentences, ending by a dot. The others are either delimited by transition units TU such as connectors (about 20%) or by specific signs (e.g. end of enumerations, punctuation marks). Delimiting units is therefore in our perspective quite simple (it may not be so in e.g. oral QA or dialogues).

3.4.2 Identification of direct responses (DS) via the domain ontology

The identification (and the production) of a number of cooperative functions (e.g. relaxation, intensional responses, direct responses) rely heavily on ontological knowledge.

Let us present first the characteristics of the ontology required in our approach. It is basically a conceptual ontology where nodes are associated with concept lexicalizations and essential properties. Each node is represented by the predicate :

```
onto-node(concept, lex, properties)
```

where *concept* has *properties* and lexicalisations *lex*. Most lexicalisations are entries in the lexicon (except for paraphrases), where morphological and grammatical aspects are described. For example, for *hotel*, we have (coded in Prolog):

```
onto-node(hotel,
[[hotel], [residence, hoteliere]],
```

```
[night-rate, nb-of-rooms,
facilities]).
```

There are several well-designed public domain ontologies on the net. Our ontology is a synthesis of two existing French ontologies, that we customized: TourinFrance (www.tourinfrance.net) and the bilingual (French and English) thesaurus of tourism and leisure activities (www.iztzg.hr/indokibiblioteka/THESAUR.PDF) which includes 2800 French terms. We manually integrated these ontologies in WEBCOOP (Benamara et al. 2004a) by removing concepts that are either too specific (i.e. too low level), like some basic aspects of ecology or rarely considered, as e.g. the economy of tourism. We also removed quite surprising classifications such as *sanatorium* under *tourist accommodation*. We finally reorganized some concept hierarchies, so that they 'look' more intuitive for a large public. Finally, we found that some hierarchies are a little bit odd, for example, we found at the same level *accommodation capacity and holiday accommodation* whereas, in our case, we consider that *capacity* is a property of the concept *tourist accommodation*.

We have, at the moment, 1000 concepts in our tourism ontology which describe accommodation and transportation and a few other satellite elements (geography, health, immigration). Besides the traditional 'isa' relation, we also coded the 'part-of' relation. Synonymy is encoded via the list of lexicalizations.

Direct responses (DS) are essentially characterized by introductory markers like *yes/no/this is possible* and by the use of similar terms as those given in the question (55% of the cases) or by various lexicalizations of the question terms, studied in depth in (Benamara et al, 2004b). An obvious situation is when the response contains a subtype of the ques-

tion focus: *opening hours of the hotel → l'hotel vous accueille 24h sur 24 (approx. hotel welcomes you round the clock)*. In terms of portability to other domains than tourism, note that the various terms used can be identified via the ontology: synonyms, sisters, subtypes.

3.4.3 Linguistic marks

In this section, for space reasons, we explore only three typical CR: justifications (AJ), restrictions (AR) and warnings (AA). These MUs are characterized by markers which are general terms, domain independent for most of them. The study of these marks for French reveals that there is little marker overlap between units. Markers have been defined in a first stage from corpus analysis and then generalized to similar terms in order to have a larger basis for evaluation. We also used, to a limited extend, a bootstrapping technique to get more data (Ravinchandran and Hovy 2002), a method that starts by an unambiguous set of anchors (often arguments of a relational term) for a target sense. Searching text fragments on the Web based on these anchors then produces a number of ways of relating these anchors.

Let us now characterize linguistic markers for each of these categories:

Restrictions (AR) are an important unit in cooperative discourse. There is a quite large literature in linguistics about the expression of restrictions. In cooperative discourse, the expression of restrictions is realized quite straightforwardly by a small number of classes of terms:

(a) restrictive locutions: *sous réserve que, à l'exception de, il n'est pas autorisé de, toutefois, etc.* (provided that),

(b) the negative form *ne ... que* that is typical of restrictions, is very frequently used

(c) restrictive modals: *doit obligatoirement, impérativement, nécessairement* (must obligatorily),

(d) quantification with a restrictive interpretation: *seul, pas tous, au maximum* (only, not all).

Justifications (AJ) is also an important meaningful unit, it has however a little bit fuzzy scope. Marks are not very clearcut. Among them, we have:
(a) marks expressing causality, mainly connectors such as: *car, parce que, en raison de*,

(b) marks expressing, via other forms of negation than in AR, the impossibility to give a positive response, or marks 'justifying' the response: *il n'y a pas, il n'existe pas, en effet* (because, there is no, indeed).

Warnings (AA) can quite clearly be identified by means of:

(a) verbal expressions: *sachez que, veuillez à ne pas, mieux vaut éviter, n'oubliez pas, attention à, etc.* (note that, do not forget, etc.),

(b) expressions or temporal morphological marks that indicate that data is sensitive to time and may be true only at some point: *mise à jour, changements fréquents, etc.* (frequent updates),

(c) a few other expressions such as: *il n'existe pas, mais (but) ... + comparative form*.

Except for the identification of DS, which require quite a lot of ontological resources, marks identified for the other MU studied here are quite general. Portability of these marks to other domains and possibly to other languages should be a reasonably feasible challenge.

The response elaboration part (ER) is more constrained in terms of marks, because of the logical procedures that are related to. For example, the CSFR, dealing with constraint relaxation, involves the use of sister, daughter and sometimes parent nodes of the focus, and often proposes at least 2 choices. It is in general associated with a negative direct response, or an explanation why no response can be found. It also also contains some fixed marks that indicate a change of concept, such as *another type of*. This is easily visible in the pair Q2-R2 (section 3.3) with the mark: *the closests*.

3.4.4 Constraints between units

A few constraints or preferences can be formulated on the organization of meaningful units, these may be somewhat flexible, because cooperative discourse may have a wide range of forms:

(a) coocurrence: any DR can co-occur with an AS, AF, AR, AA or AJ,

(b) precedence: any DR precedes any (unmarked) AA, AR, AC, ACP, B, or any sequence DS-BP. Any CC precedes any CSFR, CSFH or CSFRI,

(c) incompatibility: DS + DP, CSFR + CSFI, CSFC + CSFH. Furthermore CR cannot appear alone.

Frequent pairs are quite numerous, here are the most typical ones: DS + P, DS + AR, CC + CSFR or CSFH or CSFRI, DS + AJ, DS(negative) + AJ + AS, DS + AF, DS(negative) + CSFR. These can be considered in priority in case of ambiguities.

3.5 Evaluation by annotators

At this stage, it is necessary to have evaluated by human annotators how clear, well-delimited and easy to use this classification is. We do not have yet precise results, but it is clear that judgments may vary from one annotator to another. This is not only due to the generic character of our definitions, but also to the existence of continuums between categories,

and to the interpretation of responses that may vary depending on context, profile and culture of annotators.

An experiment carried out on three independent subjects (annotation task followed by a discussion of the results) reveals that there is a clear consensus of 80% on the annotations we did ourselves. The other 20% reflect interpretation variations, in general highly contextual. These 20% are almost the same cases for the three subjects. In particular, at the level of additional information (CR), we observed some differences in judgement in particular between restrictions (AR) and warnings (AA), and a few others between CSFH and CSFC whose differences may sometimes be only superficial (presentation of the arguments of the response).

3.6 Evaluation of prototype: a first experiment

We can now evaluate the accuracy of the linguistic marks given above. For that purpose, we designed a programme in Prolog (for fast prototyping) that uses: (1) the domain lexicon and ontology, to have access e.g. to term lexicalizations and morphology, and (2) a set of 'local' grammars that implement the different marks. Since these marks involve lexical and morphological variations, negation, and some long-distance dependencies, grammars are a good solution.

Tests were carried out on a new corpus, essentially from airlines FAQ. 134 QA pairs have been selected from this corpus containing some form of cooperativity. The annotation of this corpus is automatic, while the evaluation of the results is manual and is carried out in parallel by both ourselves and by an external professional evaluator. These 134 QA pairs contain a total of 237 MU, therefore an average of 1.76 MU per response. Most responses have 2 MU, the maximum observed being 4. Surprisingly, out of the 134 pairs, only 108 contain direct responses followed by various CSF, the other 16 only contain cooperative know-how responses (CSF), without any direct response part.

Evaluation results, although carried out on a relatively small set of QA pairs, give good indications on the accuracy of the linguistic marks, and also on the typology of the different MU. We consider here the MU: DS, AJ, AR, AA, as characterized above:

Unit	A	B	C	Total	correct annotation
DS	102	6	0	108	88%
AJ	27	6	3	36	75%
AR	36	4	2	42	86%
AA	24	0	0	24	100%

A: number of MU annotated correctly for that category, B: MU not annotated (no decision made), C: incorrect annotation.

MU boundaries have been correctly identified in 88% of the cases, they are mostly related to punctuation marks.

There are obviously a few delicate cases where annotation is difficult if not impossible. First, we observed a few discontinuities: an MU can be fragmented. In that case, it is necessary to add an index to the tag so that the different fragments can be unambiguously related, as in:
Q: What is the deadline for an internet reservation?
R: $< DR\ index = 1 >$ In the case of an electronic ticket, you can reserve up to 24h prior to departure $< /DR > . < B >$ You just need to show up at the registration desk $< /B > . < DR\ index = 1 >$ In the case of a traditional ticket ... $< /DR >$.
The index=1 allows to tie the two fragments of the enumeration.

In a number of cases the direct response part is rather indirect, making its identification via the means presented above quite delicate:
Q: I forgot to note my reservation number, how can I get it?
R: A confirmation email has been sent to you as soon as the reservation has been finalized.... To identify this portion of the response as a DR, it is necessary to infer that the email is a potential container for a reservation number.

4 Conclusion and Perspectives

We reported in this paper a preliminary version, for testing, of COOPML, a language designed to annotate the different facets of cooperative discourse. Our approach, still preliminary, can be viewed as a base to investigate the different forms of cooperativity on an empirical basis. This work is of much interest to define the formal structure of a cooperative discourse. It can be used in discourse parsing as well as generation, where it needs to be paired with other structures such as rhethorical structures. It is so far limited to written forms. We believe the same global structure, with minor adaptations and additional marks, is valid for dialogues and oral communication, but this remains to be investigated. The main application area where our work is of interest is probably advanced Question-Answering systems.

Besides cooperative discourse annotation, we have investigated the different forms lexicalization takes between the question and the different parts of the response, the direct response (DR), the response elaboration (ER) and the additional information (CR). These are subtle realizations of much

interest for natural language generation. These elements are reported in (Benamara and Saint-Dizier, 2004b).

COOPML will be extended and stabilized in the near future along the following dimensions:

- analyze the linguistic marks associated with the MU not investigated here, and possible correlations or conflicts between MU,

- analyze its customisation to various application domains: since quite a lot of ontological and lexical knowledge is involved, in particular to identify DS, this needs some elaboration,

- investigate portability to other languages, in particular investigate the cost related to linguistic resources development,

- develop a robust annotator, for each of the levels identified, and make it available on a standard platform,

- investigate knowledge annotation. This point is quite innovative and of much interest because of the heavy knowledge load involved in the production of cooperative responses.

References

Baumann, S., Brinckmann, C., Hansen-Schirra, S., Kruijff, G., The MULI Project : Annotation and Analysis of Information Structure in German and English., LREC, 2004.

Benamara, F., Saint-Dizier, P., *Dynamic Generation of Cooperative NL responses in WEBCOOP*, 9th EWNLG, Budapest, 2003.

Benamara. F, and Saint Dizier. P, *Advanced Relaxation for Cooperative Question Answering*, in: *New Directions in Question Answering*, To appear in Mark T. Maybury, (ed), AAAI/MIT Press, 2004 (a).

Benamara. F, and Saint Dizier. P, *Lexicalisation Strategies in Cooperative Question-Answering Systems* in Proc. Coling'04, Geneva, 2004 (b).

Dybkjaer, L., Bernsen, N.O., The MATE Workbench. A Tool in Support of Spoken Dialogue Annotation and Information Extraction, In B. Yuan, T. Huang, X. Tank (Eds.): Proceedings of ICSLP'2000', Beijing,", 2000.

Gal, A., *Cooperative Responses in Deductive Databases*, PhD Thesis, Univ. of Maryland, 1988.

Gaasterland, T., Godfrey, P., Minker, J., *An Overview of Cooperative Answering*, Papers in non-standard queries and non-standard answers, Clarendon Press, Oxford, 1994.

Grice, H., *Logic and Conversation*, in Cole and Morgan (eds), Syntax and Semantics, Academic Press, 1975.

Hendrix, G., Sacerdoti, E., Sagalowicz, D., Slocum, J., Developing a Natural Language Interface to Complex Data, *ACM transactions on database systems*, 3(2), 1978.

Kaplan, J., Cooperative Responses from a Portable Natural Language Query System, in M. Brady and R. Berwick (ed), *Computational Models of Discourse*, 167-208, MIT Press, 1982.

Lehnert, W., *The Process of Question Answering: a Computer Simulation of Cognition*, Lawrence Erlbaum, 1978.

Mays, E., Joshi, A., Webber, B., Taking the Initiative in Natural Language Database Interactions: Monitoring as Response, *EACL'82*, Orsay, France, 1982.

Miltsakaki, E., Prasad, R., Joshi, A., Webber, B., The Penn Discourse Treebank, LREC, 2004.

Minock M, Chu W, Yang H, Chiang K, Chow, G and Larson, C, *CoBase: A Scalable and Extensible Cooperative Information System*. Journal of Intelligent Information Systems, volume 6, number 2/3,pp : 223-259, 1996.

Netter, K., Armstrong, S., Kiss, T., Klein, J., DiET - Diagnostic and Evaluation Tools for Natural Language Applications,, Proceedings of 1st LREC, Granada.", 1998.

Orasan, C., *PALink: A Highly Customisable Tool for Discourse Annotation*, Paper from the SIGdial Workshop, 2003.

Ravinchandran, D., Hovy, E., Learning Surface Text Patterns for a Question Answering System, *ACL 2002*, Philadelphia.

Reiter, R., Dale, R., *Building Applied Natural Language Generation Systems*, Journal of Natural Language Engineering, volume 3, number 1, pp:57-87, 1997.

Searle, J., *Indirect Speech Acts*, in Cole and Morgan (eds), Syntax and Semantics III, Academic Press, 1975.

Strenston, J., Introduction to Spoken Dialog, Longman, 1994.

Webber, B., Gardent, C., Bos, J., *Position Statement: Inference in Question-Abswering*, LREC proceedings, 2002.