

Semantic Representation of Non-Sentential Utterances in Dialog

Silvie Cinková

Charles University in Prague
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics
Malostranské náměstí 25
CZ-118 00 Praha 1
cinkova@ufal.mff.cuni.cz

Abstract

Being confronted with spontaneous speech, our current annotation scheme requires alterations that would reflect the abundant use of non-sentential fragments with clausal meaning tightly connected to their context, which do not systematically occur in written texts. The purpose of this paper is to list the common patterns of non-sentential fragments and their contexts and to find a smooth resolution of their semantic annotation.

1 Introduction

Spontaneous speech, even assuming a perfect ASR, is hard to parse because of the enormous occurrence of disfluencies and syntactic deviations. Some disfluencies can be regarded as speaker's errors, which are being corrected or remain uncorrected during the speaker's turn. Such disfluencies are e.g.:

- stammering (We w-went there to-together)
- restart with or without an interregnum (John no sorry Jane was there, too)
- repetitions (So you like you like drinking)
- hesitation sounds, long silence, fillers, filler phrases, etc. (EH so ... you kinda like you know HMM drinking)

In NLP, such disfluencies can be removed before any syntactic or semantic processing since

they cause confusion without adding any semantic information. In machine-learning tasks, disfluency is sought to be automatically removed by learning from disfluency-marked corpora or corpora of text edits (Hajič et al., 2008; Fitzgerald and Jelinek, 2008) to smooth the input text into written-language standard before parsing.

On the other hand, there is another sort of disfluencies, which do not disturb the course of the dialog, namely contextual ellipsis: even though most people remember being taught at school to answer questions with a complete sentence, not even educated speakers performing a sophisticated dialog always do so, and yet they do not sound incorrect. Clearly, an extensive use of ellipsis is an inherent feature of verbal interaction between speakers, which is usually smoothly perceived by the listener and thus all right in its place.

Such “fragmentary utterances that do not have the form of a full sentence according to most traditional grammars, but that nevertheless convey a complete clausal meaning” are called **non-sentential utterances (NSUs)**¹. A consistent reconstruction of their clausal meaning is inevitable for any semantic representation of dialogs. The present paper describes a tentative semantic representation of NSUs in the Functional Generative Description (FGD) framework (Sgall et al., 1986).

¹ The term NSU as well as its definition comes from Fernández et al., 2007.

2 NSUs in PhotoPal Dialogs

2.1 NSU taxonomy

Fernández et al. (2007) introduce a taxonomy of NSUs based on the dialog transcripts from BNC (Burnard, 2000). They stress that NSUs are not limited to question-answer pairs but can appear as responses to any preceding utterance. Our observations confirm this. NSUs are highly ambiguous without context. Consider the following example:

A: I left it on the table.
B: On the table.
I confirm/I understand what you say: you left it on the table.

A: Where did you leave it?
B: On the table.
I answer your question: I left it on the table.

A: I think I put it er...
B: On the table.
I know in advance what you want to say or what you would want to say if you knew that.

A: Should I put it back on the shelf?
B: On the table.
No, don't put it back on the shelf, but put it on the table instead.

If reconstructed into a complete sentence, the NSU would get different shapes in the respective contexts (see the paraphrases in italics).

The NSU taxonomy proposed by Fernández et al. (2007) divides the NSUs into 15 classes:

- Clarification Ellipsis (Two people [*did you say were there?*])
- Check Question ([...]Okay?)
- Reprise Sluice (What[*did you say*]?)
- Direct Sluice (What?/Who?/When?)
- Short Answer [to wh-question] (My Aunty Peggy.)
- Plain Affirmative Answer / Rejection (Yes. / No.)

- Repeated Affirmative Answer (Very loud, yes.)
- Helpful Rejection (No, Billy.)
- Plain Acknowledgement (Mhm.)
- Repeated Acknowledgement (part of the preceding segment repeated)
- Propositional and Factual Modifiers (Probably not. / Oh, great!)
- Bare Modifier Phrase (adjuncts modifying a contextual utterance)
- Conjunct (fragments introduced by conjunctions)
- Filler (fragments filling a gap left by a previous unfinished utterance)

2.2 PhotoPal Dialog Corpora

Our goal is semantically annotated spoken conversations between two speakers over a family album. One English corpus (NAP) and one Czech corpus have been built within the Companions project (www.companions-project.org) as gold-standard data for a machine-learning based dialog system (“PhotoPal”) that should be able to handle a natural-like conversation with a human user, helping to sort the user’s photographs and encouraging the user to reminisce. The PhotoPal is supposed to keep track of the mentioned entities as well as to make some inferences.

The NAP corpus (Bradley et al., 2008) comprises about 200k tokens of literal manual transcriptions of audio recordings, which are inter-linked with a multiple disfluency annotation (Cinková et al., 2008). The Czech PhotoPal corpus is still growing (Hajič et al., 2009), comprising about 200k tokens at the moment (including double annotation).

To ease the understanding, all authentic corpus examples will be taken from the English NAP corpus. However, most examples in this paper are taken from Fernández et al. (2007) and modified when needed to illustrate a contrast.

3 Semantic representation of NAP NSUs

3.1 Functional Generative Description

The Functional Generative Description (FGD) is a stratified formal language description based on the structuralist tradition, developed since the

1960's. The unique contribution of FGD is the so-called tectogrammatical representation (TR). It is being implemented in a family of semantically annotated treebanks.

3.2 Tectogrammatical Representation

Being conceived as an underlying syntactic representation, the TR captures the linguistic meaning of the sentence, which is its basic description unit. In the TR annotation, each sentence is represented as a projective dependency tree with nodes and edges. The attribute values include references to the analytical (surface-syntax) layer. Only content words are represented by nodes. Function words are represented as attribute values. Each node has a semantic label ("functor"), which renders the semantic relation of the given node to its parent node. The TR annotation captures the following aspects of text:

- syntactic and semantic dependencies
- argument structure (data interlinked with a lexicon)
- information structure (topic-focus articulation)
- grammatical and contextual coreference
- ellipsis restoration.

Fig. 1 shows a sentence with restored ellipsis. The elided predicate in the second conjunct was copied from the first conjunct predicate (copied and generated nodes have square shape).

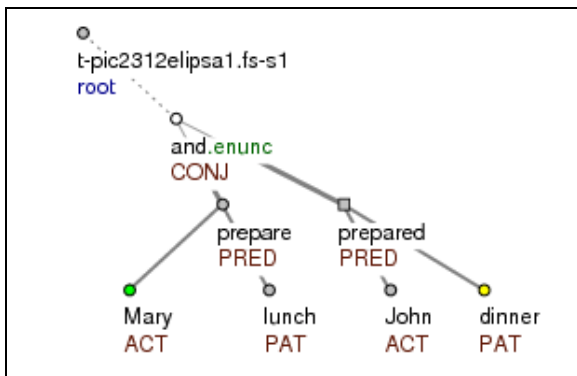


Fig.1 *Mary prepared the lunch, and John [prepared] the dinner.*

3.3 Ellipsis Restoration and Contextual Coreference

Assumingly, any tectogrammatical representation of NSUs is about the most appropriate resolution of contextual ellipsis and coreference. TR distinguishes two types of ellipsis:

- contextual ellipsis, i.e. ellipsis occurring when the lexical content of the omitted element is clear from the context and easily recoverable. The speaker omitted this element, since he considered its repetition unnecessary.
- grammatical ellipsis, i.e. such ellipsis that occurs when the elided element cannot appear on the surface for grammatical reasons but is cognitively present in the meaning of the utterance (e.g. the unexpressed subject of controlled infinitives).

Every occurrence of a given verb must correspond to the appropriate lexicon frame. Any obligatory arguments missing must be filled in as node substitutes even if the node could be copied from the context. The substitutes have special lemmas according to their function.

Fig. 2 illustrates a contextual ellipsis of a dependent node. The tree represents the answer: He has [wrapped the book] to the question: Has the shop assistant wrapped the book? In fact, the tree renders the sentence He has. To complete the argument structure frame of the verb *wrap*, the node *book* with the Patient semantic label is inserted into the frame in form of a node with the t-lemma substitute for personal pronoun (*#PersPron*, square node) exactly in the same way as the expressed *he*. The node-constituting lexical verb *wrap* is copied from the previous sentence as a square node while *has* becomes its attribute value, since it is an auxiliary verb. The subject *He* is only converted into the *#PersPron* substitute (with appropriate values inside).

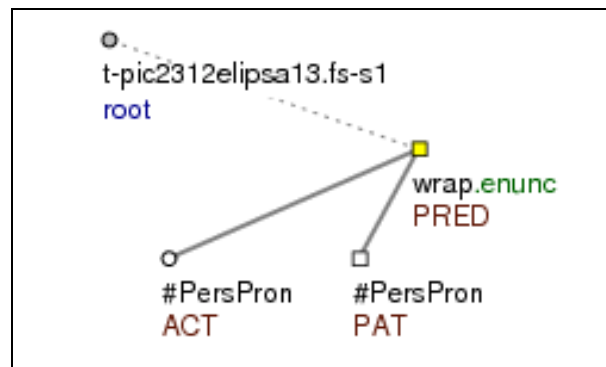


Fig. 2 *He has [wrapped the book].*

In the complete TR annotation, a contextual-coreference arrow would lead from the

#PersPron nodes to their antecedent nodes in the previous sentence (to assistant and book, respectively).

3.4 Basic Principles of NSU Representation in TR

The effort to reconstruct the clausal meaning of non-sentential utterances was motivated by the following basic assumptions:

- The text contains utterance-response pairs.
- NSU is the response to an utterance U^2 .
- The utterance U has a finite-verb predicate UPred with or without modifiers (arguments and adjuncts) UMods, which can be assigned functors.
- Even UPred can be an elided predicate.
- All NSUs (except interjections but incl. plain yes and no) contain an implicit (elided) predicate NSUPred. NSUPred is either identical with UPred, or it is an unknown verb, but we can imagine how it relates NSU and U .
- NSU can be attached to a finite clause.
- NSU inherits UPred along with all UMods.
- When there is a semantic conflict, NSUMods overrule the inherited implicit UMods in NSU (repetition is also regarded as conflict).
- NSUMod overrules UMod in the highest position possible in the subtree.

3.5 TR Representation Elements for NSU

This annotation introduced a new category into the annotation scheme. We called the category response_type and designed it in the same way as the coreference annotation. It is visualized as arrows of various colors pointing from NSUMod to UMod. Each type is indicated by a different color.

The utterance-response pair consists of two parts: the antecedent utterance U and the response NSU. The finite verb predicate UPred is typically the effective root of U , which has the functor PRED, but not necessarily. On the other hand, the elided predicate of NSU, called NSU-

Pred, is the effective root of NSU and has the functor PRED. Fig. 3 describes U in more detail.

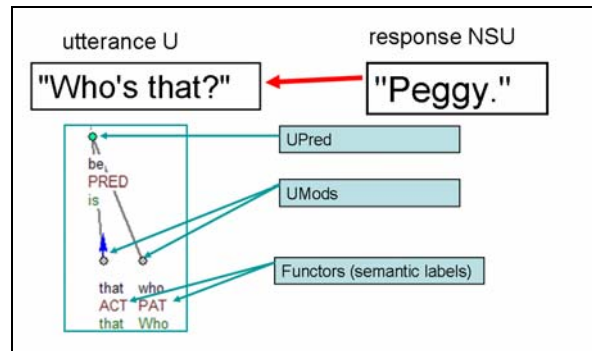


Fig 3. Utterance-response pair.

Whenever the clausal meaning of NSU can be reconstructed by using the copy of UPred as predicate, the t-lemma substitute for NSUPred is #VerbPron, which is normally also used for the pro-form do (dummy-do). NSUPred is always linked to UPred by a contextual-coreference arrow. When the clausal meaning of NSU cannot be directly reconstructed by using the copy of UPred as the predicate, NSUPred is rendered as the coreference-less t-lemma substitute #EmpVerb, which is normally used for cases of grammatical ellipsis of the predicate. #EmpVerb has no obligatory arguments and inherits no modifiers from anywhere. An NSUPred that has coreference inherits all modifiers from UPred, but these are not explicitly copied to NSUPred. NSUPred's own arguments are regarded as added to the inherited modifiers. Hence the NSU "Peggy." does not have to be explicitly reconstructed as "That is Peggy." (the left figure in Fig.4), but just with the coreferential predicate (the right figure).

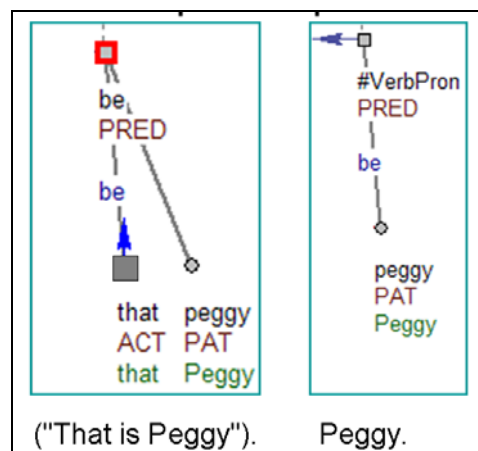


Fig. 4 Response NSU: Full explanatory reconstruction (left) and the actual annotation resolution (right).

² NSU is regarded as a response even if U is a statement and NSU a question.

Obviously, NSUMods can be in a semantic conflict with the inherited UMods. These cases are marked by several types of arrows leading from the given NSUMod to the conflicting UMod in the antecedent utterance U. We distinguish four types of semantic conflict between NSUMod and UMod:

- overruling
- rephrasing
- wh-path
- other

3.6 Overruling

Overruling is the most typical semantic conflict where an NSUMod gives exactly the same type of information, but relating to a different entity in the real world. If NSU is to be expressed as a clause that uses the predicate of U, the conflicting UMod is erased (or prevented from inheriting) by the explicitly present NSUMod. E.g. in the following utterance-response pair:

U: I'm in a little place called Hellenthorpe.
 NSU: Ellenthorpe.
 NSU-paraphrase: *You are in a little place called ~~Hellen-~~
~~thorpe~~ Ellenthorpe.*

Even the explicit repetition is regarded as overruling:

U: There were just two people in the class.
 NSU: Two people?.
 NSU-paraphrase: *Were there just ~~two people~~ two people in the class?*

In the tree representation, the crossed text would be visible only in the tree of U, and an overruling-reference arrow would point at them from the relevant NSUMod. This conception prevents doubling the same modifier in NSU.

3.7 Rephrasing

When an NSUMod is rephrasing an UMod, then UMod and NSUMod refer to the same entity in the real world, or one refers to the entire entity whereas the other one refers only to its part, etc., using a different wording. The NSUMod-UMod relation marked as rephrasing is meant to be-

come the starting material for bridging anaphora research. Example:

U: There were just two people in the class.
 NSU: Just two students?
 NSU-paraphrase: *Were there just ~~two people~~ two students in the class?*

It is also applied when the context is unambiguous for the speakers but ambiguous for the annotator, who lacks their background knowledge of the given situation. In the following example the annotator may not know whether this part or just the end of this part should come up, because he does not see the speakers pointing at the crane, but it is rather evident that it is not a completely different part of the crane but something at the end of it:

U1: You lift the crane, so this part comes up.
 NSU1/U2: The end?
 NSU1/U2-paraphrase1: *Do you mean the end comes up?*
 NSU1/U2-paraphrase2: *Do you mean the end of this part comes up?*
 NSU2/U3: Just this.
 NSU3: Okay.

The category “Other” (see below) is though strongly preferred in ambiguous cases.

3.8 Wh-path³

The wh-path relation is the relation between the modifier that is focused by a wh-word in an U that is a direct or indirect question and a NSUMod that makes a good answer.

Overruling as well as rephrasing assume that the conflicting modifiers have the same functor. The wh-path category is different from the others in that it allows setting in conflict a UMod with an NSUMod with different semantic labels (functors). Our tentative annotation suggests that regular patterns will occur; e.g. with the question about direction/location. When asking where, speakers often get replies that would actually match questions with whom (functor ACMP) or with which intention (functor INTT,

³ The term was found in Hajičová (1995) and reused by placing it in context with other response types.

e.g., go shopping), and yet they are perceived as good answers.

The relation between an utterance U which is a statement and an NSU which is a sluice is not wh-path but overruling. Cf.:

U: Where would you like to go tomorrow?
 NSU: Downtown with Mary, to do some shopping. (wh-path)

U: I would like to go downtown with Mary tomorrow.
 NSU: Where? (overruling)

Sluices are not regarded as ambiguous in the sense whether referring to the same entity as the corresponding wh-word or not. They are not eligible for the relation “other” (see next section).

3.9 Other

“Other” is meant for inherently ambiguous cases of conflicting UMod and NSUMod where it is impossible to decide whether NSUMod is rephrasing or overruling UMod. Textual ambiguity arises when NSU is a question that does not find a proper answer in the context:

U1: He’s got the best room.
 NSU1/U2: Room 128?
 NSU1/U2-paraphrase: Has he got the best room Room 128?
 U3: I don’t know which number.

3.10 TR-Conditioned Criteria for NSU types

The original idea of the tectogrammatical representation of NSU was to adopt the taxonomy proposed by Fernández et al. (2007). However, the rules of TR made some classes collapse as they yielded identical tectogrammatical tree structures. The main criteria for tectogrammatical representation of NSU were the following:

Is the NSU a phrase or just an interjection? (Cf. Fig. 5 and 6)

- If it is a content word or a phrase, it should be reconstructed into a clause by adding a predicate.
- If it is an interjection except yes and no (and their colloquial variants), no predicate is added.
- If it is yes/no (and variants), a predicate should be added.

- If the interjection acts as a backchannel, yes and no make no exception.

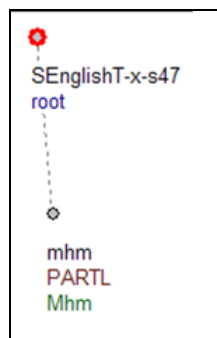


Fig. 5 Interjection

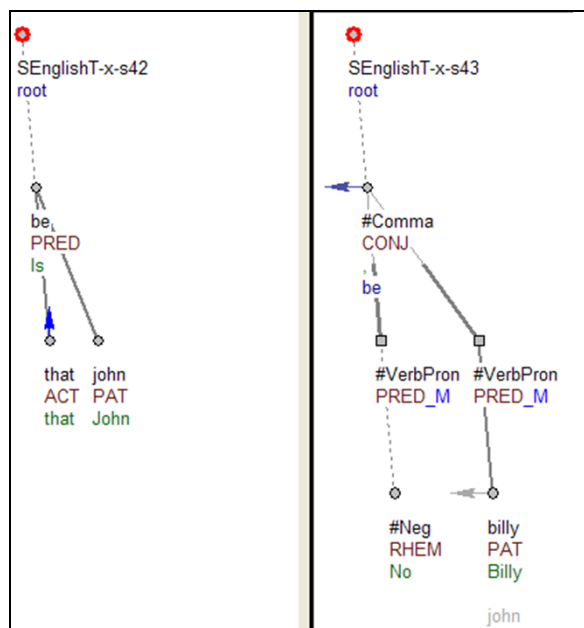


Fig. 6 Is this John? No, Billy [This is not John, this is Billy.]

Can we copy UPred to make NSU a clause?

- If we can, NSUPred has the t-lemma substitute #VerbPron and a coreferential arrow points from NSUPred to UPred.
- If we cannot, NSUPred has the t-lemma #EmpVerb with no coreferential arrow. No response type arrows point from NSUMods to UMods. In specific cases the coreference to UPred leads from elsewhere (Fig.7).

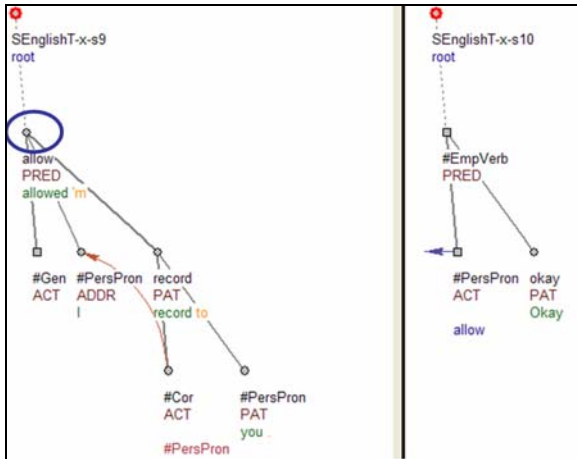


Fig. 7 Check question/Evaluative response related to text:
 U: I am allowed to record you.
 NSU (same speaker): Okay?
 NSU-paraphrase: Is it (that I'm allowed to record you) okay?
 or
 U: I am allowed to record you.
 NSU (turn switch): Okay.
 NSU-paraphrase: It <is> okay that you are allowed to record me.

3.11 More Examples of U-NSU relation resolution

Fernández et al. (2007) distinguish two types of sluice: the direct and the reprise sluice. In TR, each has a different semantic representation. The direct sluice has the coreferential predicate while the reprise sluice, which can be paraphrased as *What did you mean by saying this?*, has the empty-verb predicate and the wh-word gets the functor EFF, which is normally assigned to what is being said in the argument structure pattern of verbs of saying (Fig. 8).

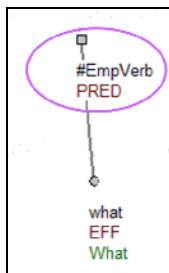


Fig. 8 Reprise sluice

Fig. 9 shows a sentence with wh-path linking modifiers with different functors.

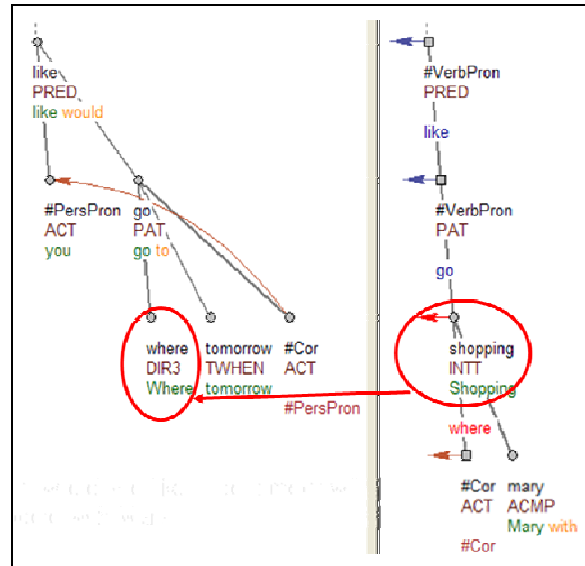


Fig. 9 Wh-path linking Mods with different functors

U: Where would you like to go tomorrow?
 NSU: Shopping with Mary.
 NSU-paraphrase: Tomorrow I would like to go shopping with Mary.

Choice questions (Fig.10) represent an interesting example in which one NSUMod can enter different relations to different UMods. The NSUMod *beer* overrules the coordinated UMod *Coke* or *Pepsi*, and at the same time it is connected with the wh-question *Which do you like to drink?* by wh-path.

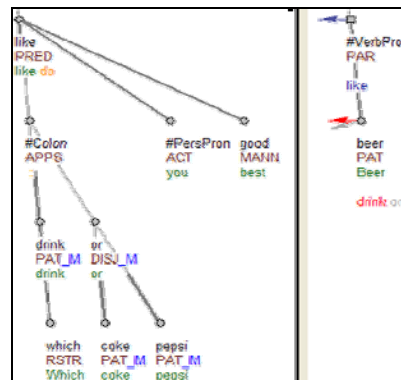


Fig. 10 Choice question.

U: Which do you like to drink: Coke or Pepsi?
 NSU: Beer.
 NSU-paraphrase: I like to drink beer.

Seeing the many rephrasing cases in the data, which are supposed to be subject to further anaphora annotation (bridging etc.), we had to ask the question whether the boundary between *response_type* and *coreference* can be reliably determined. We found good evidence in the made-up but not unlikely example below (Fig.

11). In this context, him will be coreferential with Paul and her will be coreferential with Mary. On the other hand, him will overrule Mary and her will overrule Paul (only the relations of him are marked in the figure).

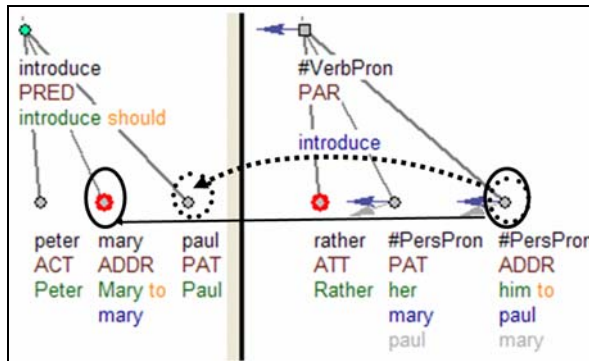


Fig. 11 Coreference vs. response type

3.12 Current and Future Work

The proposed enhancement of the annotation scheme has been tested on a corpus of approx. 200 NSUs with context manually extracted from the NAP transcripts as well as on example sentences from Fernández et al. (2007) and many sentences obtained by their modification performed in order to get potentially difficult counterexamples. As this is still a preparatory work, neither the inter-annotator agreement nor any other evaluation could be done so far.

In the next future, parts of the spoken corpora should get tectogrammatical parsing. The manual annotation is supposed to adopt this new feature of the annotation scheme, and we will try to incorporate it into our statistically trained automatic parsing tools.

Conclusion

The confrontation of our current annotation scheme with spoken dialog data has raised issues of ellipsis restoration and textual coreference in non-sentential utterances. We have found common relations between non-sentential utterances and their contexts, and we have integrated them into our semantic annotation scheme without violating its general principles. A tentative manual annotation of these relations in a small corpus suggests that such annotation is feasible. Further investigation on larger data along with machine-learning experiments is intended.

Acknowledgements

This work was funded in part by the Companions project (www.companions-project.org) sponsored by the European Commission as part of the Information Society Technologies (IST) programme under EC grant number IST-FP6-034434, as well as by the Czech Science Foundation (GA405/06/0589), and by the Czech Ministry of Education (MSM0021620838, MŠMT ČR LC536).

References

- Jay Bradley, Oli Mival, and D. Benyon. 2008. A Novel Architecture for Designing by Wizard of Oz. In: Proceeding of CREATE08, British computer Society, Covent Garden, London, 24-25 June 2008.
- Lou Burnard. 2000. Reference Guide for the British National Corpus (World Edition). Oxford University Computing Services. Available from <ftp://sable.ox.ac.uk/pub/ota/BNC>.
- Silvie Cinková, Jan Hajič, Jan Ptáček. 2008. An Annotation Scheme for Speech Reconstruction on a Dialog Corpus. In Fourth International Workshop on Human-Computer Conversation. Bellagio, Italy: [http://www.companions-project.org/events/200810_bellagio.cfm],2008:1-6.
- Raquel Fernández, Jonathan Ginzburg, and Shalom Lappin. 2007. Classifying Non-Sentential Utterances in Dialogue: A Machine Learning Approach. Computational Linguistics, Volume 33, Nr. 3. MIT Press for the Association for Computational Linguistics.
- Erin Fitzgerald and Frederick Jelinek. 2008. Linguistic Resources for Reconstructing Spontaneous Speech Text. In: LREC 2008 Proceedings.
- Jan Hajič, Silvie Cinková, Marie Mikulová, Petr Pajas, Jan Ptáček, Josef Toman, Zdeňka Uřešová. 2008. PDTSL: An Annotated Resource For Speech Reconstruction. In Proceedings of the 2008 IEEE Workshop on Spoken Language Technology. IEEE, 2008.
- Jan Hajič, Marie Mikulová, Martina Otradovcová, Petr Pajas, Nino Peterek, Pavel Češka, Miroslav Spousta. 2009. PDTSL - Prague Dependency Treebank of Spoken Language - Czech, Institute of Formal and Applied Linguistics, Charles University in Prague.
- Eva Hajičová (ed.) 1995. Text And-Inference-Based Approach to Question Answering. Prague.
- Petr Sgall, Eva Hajičová, and Jarmila Panevová. 1986. The Meaning of the Sentence in Its Semantic and Pragmatic Aspects. Dordrecht:Reidel Publishing Company and Prague:Academia.