

# Data collection from a WoZ- experiment, annotation scheme development and first results.

Núria Bertomeu  
IGK, Saarland University

# Introduction

- Overall goal: provide an analysis of discourse phenomena in QA dialogues.
- Discourse phenomena considered: elliptical questions, bridging, anaphora, nominal ellipsis.
- Questions: do these phenomena really occur? What are the relations between discourse phenomena and antecedents?
- Need for quantitative and qualitative data.
- Previous approaches:
  - n WoZ experiments were done in the 80s, but there are no quantitative data available. (Carbonell, 1983)
  - n Proposals to annotate discourse structure in QA have been made, but there is no annotated corpus (Chai and Jin, 2004)

# Outline

- Experimental setup.
- Annotation scheme and discussion.
- Some initial descriptive statistics.
- Conclusions.

# Experimental setup

- Interaction modality: chat interface.
- 3 tasks and 10 subjects per task (+ 3 pilots).
  - <sub>n</sub> Find 3 traineeships at 3 different projects at 3 different institutions dealing with different subjects.
  - <sub>n</sub> Find 3 conferences in the winter-term and 3 in the summer-term taking place in different countries and dealing with different topics.
  - <sub>n</sub> Obtain some information for a report about language technology in the last 10 years in Europe.
- Querying LT-WORLD in English.

# Experimental setup

- Instructions and examples in German.
- Example dialogues were about a different task and contained sentences both exhibiting discourse phenomena and not exhibiting them.
- 1 hour for obtaining the information + 15 mins. to fill in a form with the results.
- Subjects had different nationalities and pursued degrees in different areas.

# Experimental setup

- The wizard was interacting with the subjects through a chat interface.
- She used an application which allowed her to easily construct the queries and generate the answers.
- She was told to answer the questions of the subjects and ask for clarification when there was a misunderstanding.
- She displayed requests to wait when the query took long to process.

# Corpus

- 33 logs / 17 annotated.
- 125.534 alphanumerical strings.
- 2.534 turns.
- 1.174 user turns.

# Annotation scheme

## ● Annotation levels:

<sup>n</sup> **turns** → speaker and time.

<sup>n</sup> **pos**

<sup>n</sup> **questions**

<sup>n</sup> **utterances**

<sup>n</sup> **entities**

automatically annotated



# Annotation scheme: questions

- User questions corresponding to database queries.  
Relational questions are annotated as 1 single question.
- *semantic message*:
  - <sub>n</sub> wh-question vs. polar question.
- *query*:
  - <sub>n</sub> SERQL-query.
- *question-to-question-relation*:
  - <sub>n</sub> refinement, theme-entity, theme-property, paraphrase, overlap.
  - <sub>n</sub> pointer to a previous question.
- *question-to-answer-relation*:
  - <sub>n</sub> refinement, theme.
  - <sub>n</sub> pointer to a previous answer.

# Some examples: refinement

- **Refinement from question:** a question which asks for the same type of entity as some previous question, but where the restricting conditions are different, asking, thus, for a subset, superset, overlapping set or disjoint set of the same class.
- **Refinement from answer:** a question which asks for a subset of the entities given in the previous answer.

(14:16:22) **USER:** How many institutions concerning language technologies are there?

(14:16:52) **LT-WORLD:** 3810.

(14:17:36) **USER:** and how many in Europe?

(14:18:34) **LT-WORLD:** 1130.

(14:19:16) **USER:** How many of them came to exist during the last 10 years?

(14:20:13) **LT-WORLD:** The database does not contain information about founding dates.

(14:20:56) **USER:** How many financing institutions, companies are there?

# Some examples: theme

- **Theme-entity from question:** a question is about the same entity as some previous question.

(15:28:01) **USER:** What kind of topics will be discussed at AAAI Spring Symposium on Semantic Web meets eGovernement?

(15:28:57) **LT-WORLD:** Semantic Web. Semantic Knowledge Mangament. Ontology. Mapping. Knowledge Retrieval.

(15:29:24) **USER:** What is the dead-line for applicants?

(15:30:06) **LT-WORLD:** 2005-10-21.

- **Theme-property from question:** a question asks for the same property as the previous question but for another entity.

- **Theme from answer:** a question asks about an entity first introduced in some previous answer.

(16:59:14) **USER:** I need projects at the University of Athens in Signal Processing.

(16:59:55) **LT-WORLD:** RainbowII. MEMPHIS. MKBEEM. CATCH-2004. PHYSTA. MANDES. VISION.

(17:00:27) **USER:** Description of MANDES, please.

(17:01:16) **LT-WORLD:** The database does not contain information about MANDES.

(17:01:28) **USER:** What about VISION?

# Annotation scheme: utterances

## ● Criteria to utterance segmentation:

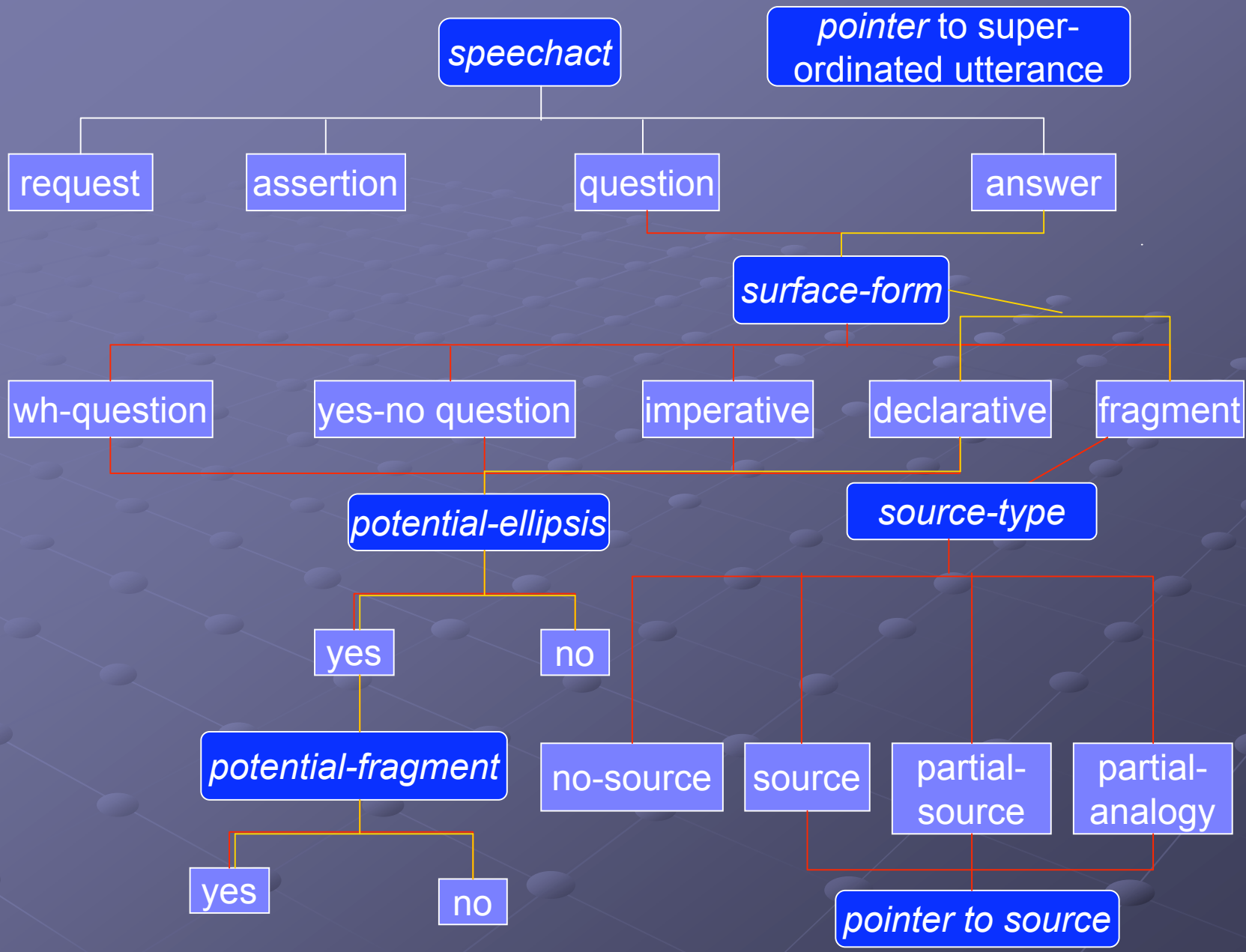
- n punctuation, when available;
- n unique speech act;
- n unique surface-form.

## ● *speech-act*:

- n question, answer, assertion, request.
- n pointer to a super-ordinated utterance, when it applies.

## ● **Questions and answers:**

- n *surface-form*.



# Some examples: utterance subordination

- Question-answer pairs, dependent questions, requests and assertions related to some question. It introduces structure in the discourse in terms of discourse goals.

(14:42:56) **USER:** how many inventions linked with language technology were there over the last ten years?

(14:44:14) **LT-WORLD:** 749.

(14:44:29) **USER:** Can you give me examples?

(14:44:48) **LT-WORLD:** How many patents would you like to see?

(14:45:30) **USER:** the 10 most recent ones?

(14:46:46) **LT-WORLD:** Annotation of an image based on associated data. Speech synthesis apparatus and method. Machine translation apparatus and method. System of Finite State Machines. ...

(14:46:51) **USER:** where were they announced?

(14:47:24) **LT-WORLD:** Please wait...

(14:49:00) **LT-WORLD:** Annotation of an image based on associated data, in United\_Kingdom. Speech synthesis apparatus and method, in United\_States. Machine translation apparatus and method, in United\_States.

# Some examples: surface-form

- Declarative:

(18:40:04) **USER:** i need deadline, topic and time of AAAI Spring Symposium on Semantic Web meets eGovernment.

- Imperative:

(18:07:14) **USER:** give me three different conferences in winter 2005

- Fragment:

- n No-source:

(18:21:03) **USER:** homepage of AAAI Spring Symposium on Semantic Web meets eGovernment.

- n Source:

(18:27:03) **USER:** where is European Joint Conferences on Theory and Practice of Software? ←

(18:27:42) **LT-WORLD:** Vienna, Austria.

(18:27:49) **USER:** and AAAI Spring Symposium on Semantic Web meets eGovernment? ←

- n Partial source:

(15:37:19) **USER:** When does the conference start? ←

(15:37:42) **LT-WORLD:** 2005-12-16.

(15:37:52) **USER:** Topics discussed? ←

# Annotation scheme: entities

- *entity type:*

- n **question specific** → abstract, restriction.
- n general → entity, generic.

- *semantic class:*

- n organisation, project, technology, location, ...

- *grammatical function:*

- n none, subject, object, object2, other.

- *co-reference type:*

- n identity, subset, other → set of markables
  - *potentially less explicit:* yes, no.
- n bridging, ellipsis nominal antecedent, possessed → pointer to markable

- *parallel class:*

- n abstracts and their instantiations in the answers → set of markables

obligatory

optional



# Some examples: abstract and restriction

- Questions are traditionally treated as open propositions, where some entity is  $\lambda$ -abstracted.
- The abstract corresponds to the linguistic realization of the entity which is  $\lambda$ -abstracted in a question.
- The restriction corresponds to the linguistic realization of the conditions holding on the abstract.

(11:04:41) **USER:** when does the third global wordnet conference take place?

( $\lambda t$ , where  $t$  is a time span and the third global wordnet conference takes place during  $t$ .)

(17:14:10) **USER:** topics of smartweb?

( $\lambda x$ , where  $x$  is a topic and smartweb is on  $x$ .)

(11:50:06) **USER:** define authoring tools

( $\lambda x$ , where  $x$  is a definition and  $x$  holds of the technology authoring tools.)

(17:48:21) **USER:** Are there conferences on translation?

( $\lambda x$ , where  $x$  is a conference and  $x$  is on translation.)

# Some examples: parallel class

- Instantiations of abstracts are annotated as entities and marked as parallel to those.
- Distinguish the new information from the old information provided in the answer and identify cases of overanswering.

(13:01:48) **USER:** List me **the names** of the projects of the University of Nijmegen

(13:02:26) **LT-WORLD:** SemiPort. DIOGENE. CLIF.

(13:02:40) **USER:** **When** did these projects start and **when** will they end?

(13:04:15) **LT-WORLD:** SemiPort: 2002-04-10 - 2005-06-30.

DIOGENE: 1999-05-01 - 2003-04-20. CLIF: 2003-08-01 - 2006-07-31.

# Some examples: co-reference

## ● Identity:

(13:08:40) **USER:** At which university are projects to language resources in the moment?

(13:09:32) **LT-WORLD:** Carnegie Mellon University. Dauphine University.

(13:09:56) **USER:** List me the names of these projects

(16:00:24) **USER:** are there organizations for language technology in europe?

(16:02:02) **LT-WORLD:** LIMSI Spoken Language Processing Group. ...

(16:03:27) **USER:** how many groups among them?

(16:03:57) **LT-WORLD:** 53.

## ● Superset / subset:

(12:55:22) **USER:** please give me a list of technologies, that are uses in europe

(12:55:42) **LT-WORLD:** Acoustic Modelling in Speech Recognition. Answer Extraction. Authoring Tools. ...

(12:56:30) **USER:** which are the most important

(13:31:47) **USER:** now tell me hoe many conferences about these technologies where there in the last 5 years

(13:32:28) **LT-WORLD:** 630.

(13:33:44) **USER:** please name 3 and where were they held ?

# Nominal ellipsis

- An empty word is inserted in the base-data.
- A markable at the entities level is created.
- The elided material behaves then as a normal entity, establishing identity, subset, or parallel relations with other entities.
- If no such a relation exists between the elided material and the antecedent, the 1st points to the 2nd through a relation *ellipsis\_nominal\_antecedent*.

# Some examples: nominal ellipsis.

## ● Identity:

(14:55:00) USER: what fields cover the first ten [ ]?

(11:26:25) **USER:** is there also a possibility for joining [ ] only two or three days and for handing in a paper [ ] before [ ]?

(11:22:16) **USER:** when does it take place and where [ ]?

## ● Ellipsis nominal antecedent:

(16:53:14) USER: I am searching for projects.

(16:53:35) LT-WORLD: Would you like to see all projects?

(16:54:02) USER: How many [ ] are there?

# Some examples: bridging

- **Bridging:** Definite NPs denoting an entity which hasn't been introduced in the discourse, but which stands in some kind of relation to an entity being spoken about.

(18:10:08) **USER:** tell me the dates of **these conferences**.

(18:10:55) **LT-WORLD:** 2006-03-27. 2006-01-20. 2005-12-20.

(18:12:22) **USER:** what are **the homepages**

# Some quantitative data

UTTERANCES	QUESTIONS	SURFACE-FORM				
		FRAGMENT	WH	YES-NO	DECLARATIVE	IMPERATIVE
1321	586	118 (20.13%)	336 (57.33%)	70 (11.94%)	13 (2.21%)	49 (8.36%)

- 42.62% of the questions are not direct *wh-questions*, although most of them are semantically *wh*.
- It's important to find a way to detect the abstract in the absence of a *wh*-word.

QUEST	REF. Q	REF. A	THEME E Q	THEME P Q	THEME A	RELATED TO Q, A	PARAPHRASE
544	64	17	187	60	164	61	30

- These data, however, don't say much about how the discourse is thematically structured, because the utterances may not be consecutive.
- Next step: look at distances between thematically related utterances.
- Differences across subjects: some construct the discourse thematically, some jump from theme to theme and then go back to the previous one. Sometimes they forget to ask about something.
- Hypothesis: ellipsis is possible when 2 utterances are related thematically and there is no intervening theme between them.

# Discourse phenomena

ELLIPTICAL QUESTIONS	NO-SOURCE	SOURCE	PARTIAL-SOURCE	PARTIAL-ANALOGY	TOTAL (SOURCE)	POTENTIAL ELL.QUEST.	TOTAL
118	39	24	39	16	79 (27.81%)	205 (72.18%)	284

- Ellipsis is an optional phenomenon.
- The use of ellipsis diverges a lot from subject to subject.  
Mean=4.6. High variance → between 26 and 0.
- Differences across different fields of study. E.g. Computational linguists are less naïve.

IDENTITY	ELLIPSIS	PRON.	DEICT. NP	DEF NP+ ELLIPSIS	DEF NP	DEICT. PRON.	TOTAL	POT-LESS EXPLICIT
1030	29	65	31	13	102	23	263	80

- These data are not concluding, since among the 80 potentially less explicit references there are some which are already expressed by discourse phenomena
- However, there are <343 occurrences of entities which can be referred to by discourse phenomena, and in 263 of them this is the case.
- In >76.67% of the cases where it is possible to make an implicit reference this is done.



# Discourse phenomena

## ● Bridging:

- n 59 occurrences.
- n High variance across subjects: most don't use it at all, some use it a lot.
- n Mean=3.47. Values between 16 and 0.

# Conclusions

- The chat modality has features of both spoken and written language.
  - n it is spontaneous instead of planned.
  - n it is almost synchronous (delay, questions are sometimes answered when another question has already been posed).
  - n for some subjects it is a formal interaction, for other is rather informal (formal formulation vs. no punctuation, no caps, more ellipsis).
  - n there is the possibility of copy and paste, which substitutes other instruments of linguistic economy.

# Conclusions

- Different attitudes towards the machine:
  - n Of 18 subjects, 12 believed they were talking to a machine, 5 to a human, and 1 was undecided.
  - n Some don't trust that the machine will be able to understand the discourse phenomena.
  - n Some are more familiar with copying and pasting than with formulating new utterances.
  - n Some find it easier to query a database than to formulate questions in natural language.
  - n Others behave linguistically like with another human and found that the machine did a good job :-)

# Conclusions

- That elliptical questions occur in 27% of the cases makes it worth to attempt their resolution in a QA system.
- The number may get higher when subjects trust more the capabilities of the machine.
- 76.67% of implicit co-reference tells that it is really difficult to say everything explicitly.

# Next steps

- Look at the distances between fragments and sources and to their thematic relatedness.
- Annotate what fragments inherit from the sources.
- Make a typology of different realisations of abstracts.
- Look at the subsets and other kinds of co-reference.
- Let annotate the corpus by a second annotator.