Lauri Carlson

Parsing Spoken Dialogue

Abstract

This paper describes an approach to parsing spoken conversation. The idea is to extend the dialogue game model of dialogue (Carlson 1983) with schemas from classical rhetorical stylistics, and implement a parser for them by extending the `cparse` categorial grammar parser (Carlson 2005) with appropriate Lambek style categorial inference rules (Lambek 1958).

1. Written monologue vs. spoken dialogue

Fred Karlsson has taken pains to empirically show just how far real life sentences fail to reach the bottomless depths of context free grammar complexity. Spoken language syntax appears even flatter and fragmentary compared to written. All the less does one expect to find in real speech artful periods like[1]

(1) *Lektorn: Om dig, Eriksson, om vilken jag ej kunnat tänka mig något dylikt, har jag, då du, då jag till följd av iråkad snuva måste nysa, brast i skratt, fått en högst ofördelaktig tanke.*
'Lecturer: Of you, Eriksson, whom I would not have thought capable of it, I have, now that you, when I had to sneeze due to a head cold, burst into a laugh, formed a highly unfavourable opinion'

My reactionary point in this note is, nevertheless, that spoken language is closer to written language than has been fashionable to think in recent times.[2] With Teleman (1983), I submit that the differences have less to do

[1] The example originally comes from a cartoon in *Strix* nr 1 (1902). I found it in Lars Magnusson (2005).
[2] The Swedish written vs. spoken controversy (Linell 1982, Telemann 1983, Allwood & al .1990, Anward 2001) is at least partly an ideological one. For instance Linell's (2003) definition of a spoken language construction as a configuration, in the form of a

A Man of Measure
Festschrift in Honour of Fred Karlsson, pp. 128–137

with grammar than with the different contingencies of monologue vs. dialogue and the differences between written vs. spoken channel.[3]

The point can be derived from my dialogue game model (Carlson 1983). From the dialogue game point of view, a conversation is a projection of a largely implicit dialogue game into one explicit play of one. Looking at it from the other end, spoken discourse is a partly submerged iceberg, well formed if it can be extended to a complete dialogue game.

The point of this extension is a logical one: the reconstructed dialogue allows the application of a logical and game theoretical definition of conversational coherence in terms of rational play of a game (Carlson 1983: xiv). This goal provides a semantic and pragmatic boundary condition on the success of the grammatical reconstruction of utterances as well: the reconstructed grammar of spoken language should make the logic of the embedding dialogue explicit.

It has been pointed out earlier that spoken language syntax appears more complete and connected if one does not restrict the unit of analysis into a single turn, but considers fragments of language in their context. A stylized example of this interplay of grammar and game structure is the way Donald Duck's nephews talk in turns, constructing well-formed sentences in cooperation. This ability of turn co-construction is no fiction, but happens all the time in everyday conversation (Local 2005). Conversely, different speakers' turns separated on their individual channels reveal much more finished phrases than appears from their joint record (Auli Hakulinen, p.c.).

According to this model, (spoken) dialogue differs from (written) monologue principally in that dialogue moves can be constructed by several speakers and span several speaking turns, and conversely, an explicit dialogue turn may manifest traces of several implicit moves, not all of them explicit. The task of parsing dialogue splits into the two tasks of reconstruction of moves and parsing them.

In this view, the bits and pieces of spoken language surface utterances are not directly governed by a, special looser spoken language syntax on the level of individual moves, but result from interplay of more or less

---

recurring template, of two or more morphosyntactically and prosodically constituted and linked expression elements, associated with a relatively stable semantic-pragmatic (functional) potential for carrying out communicative tasks is not in conflict with my suggestions.

[3] My claim is limited to the traditional qualitative formal language notion of grammatical coverage, apart from the statistics of constructions (Biber 1986).

written language like phrasal grammar with real-time dialogue game structure and planning. Though the result of this interplay might also be described directly in terms of a spoken language grammar, my hypothesis is that the modular view allows for better explanatory power and more elegant grammar.[4]

2. Grammar

Formally, from the point of view of parsing, my thesis boils down to the idea that a parser or generator for spoken language can be obtained by applying a number of spoken language specific processes (or rule schemas) to (perhaps a subset of) written language grammar.

It immediately turns out that this idea is not at all novel, either. Such processes have already been identified in the long rhetorical tradition starting with Aristotle and listed as classical rhetorical devices. Here is a short list.

1) ellipsis
2) aposiopesis
3) asyndeton
4) anakolouthon
5) apokoinou
6) repetition
7) correction[5]

I exemplify these processes with examples drawn from Swedish.[6]

Ellipsis leaves out a beginning or end of a phrase, or more generally, allows suppressing (usually low-information) parts of one. Spoken Swedish seems to allow dropping phrase initial grammatical words like pronouns and conjunctions rather freely.

---

[4] I do not suggest spoken syntax consists of performance errors, rather, it represents an adaptation of grammar to conversation. As Miller & Weinert (1998) point out, similar constructions occur in spontaneous speech across languages, which indicates that systematic, language independent mechanisms are at work.

[5] See e.g. http://humanities.byu.edu/rhetoric/ or for a longer list of schemes. Thanks to Helena Pirttisaari (p.c.) for spotting this term.

[6] The Swedish examples originate from Magnusson (n.d.), They are only used as illustration of well known spoken language phenomena, so nothing hinges on their accuracy or specifics.

(2) *– jag tror att det är apelsinträd.*
'I think it is an orange tree'

(3) *– jaså, satt dom själv?*
'I see, planted them yourself?'

The opposite rhetorical device of interruption, or leaving one's sentence unfinished, is known as **aposiopesis**. Here, the omitted part may include the main message of the planned contribution.

(4) *å de…ja tyckte skolan gick ju bra...de va inget speciellt,* ***lärarna va inga sp-****...ja har inget minne av att dom va elaka.*
'and it – I thought school went well .. it was nothing special, the teachers were no mo- … I have no recollection that they were nasty'

Asyndeton, unmarked coordination or ellipsis of conjunction, allows chaining similar phrases without explicit connectives. This device is not unknown in written language grammar, though less frequent.

(5) *och eh...då åkte vi ner till morfar han bodde i själva samhället, nere i stationssamhället nere i lidnäs.*
'and um … then we drove down to grandpa he lived downtown, down at the station down in Lidnäs'

Repetition (epanalepsis) is just what the name suggests: a part of a phrase is repeated several times in the same form or in different variants, perhaps because it is difficult, or because it is important.

(6) *...sen ..hade min yngsta moster hon gjorde..ett lekebo.....****i-i-i förå-förrådet*** *där dom hade tvättstugan,*
'… then … my youngest aunt she made … a playhouse … in-in-in the storeroom where they had the laundry'

Anakolouthon is the result of starting a sentence in one way and ending it in another. When there is a shared phrase in the middle, we have the special case known as **apokoinou**.[7]

[7] This example is from Norén (2003b). Norén's prototypical cases of apokoinou are of the form A B B\A and serve the function of affirmation. By the theory of thematic functions in Carlson (1983: 202, 211), this is the predicted function of the chiasmus A B , B A, where the first occurrence of B is new information, the second contrastive.

(7) *Dags för de:n eh andra perioden, Kanada alltså i ett-noll-le:dning, Sverige har* ***har*** *numerärt över:läge* ***i ytterligare femtioåtta sekunder ska*** *Sean >Donovan< (0.4) sitta utvisad.}*
'Time for the uh second period, Canada leading one-nought, Sweden has a power play for another fifty-eight seconds is how long Sean Donovan will serve penalty'

Repair (correctio) is an attempt to take back part of a sentence an replace it with another part. Editing noises are made at the juncture to indicate what is to be ignored or replaced.

(8) *nä han måste antagligen för han sa det haka inte upp* (.) *dej på det sa han för att* (.) ***eller haka inte upp er på det*** *för att jag- liksom måste* (.) *följa programmet sa han.. så han menar väl att...*
'No he must presumably for he said don't get stuck on that he said because (.) or don't you guys get stuck on it because I- like I must (.) follow the program he said … so he must mean that …'

This analysis already shows that the spoken language devices are not formally well separated. Many devices are special cases of others.

On the other hand, no special purpose processes are called for to parse complete phrases of any category. Sentences need no special privilege in parsing theory.

(9) *...de va också en historisk byggnad* | *som hade vart...*| ***jaa men va hade de vart*** | ***nåt slags tempel av nå slag tror ja*** | *som dom hade bevarat den ena delen å sen byggt till å gjort större å så där...*
'… It was also a historical building | that had been… | well what had it been | some kind of a temple or something I think | that they had preserved one part and then built on it and made it bigger or something'

This is an example of one turn containing a small self-addressed question-answer dialogue at the marked junctures. There is no call to parse it into one sentence.

In general, object grammar need not parse the explicit conversational turns, but just the implicit moves. Those moves are reconstructed from the explicit turns with the help of rhetorical schemes and circumstantial knowledge of the surrounding dialogue game.

In formal language terms, spoken discourse resembles the result of applying shuffle and quotient operators on the concatenation of the different participants' moves.

3. Parsing

Janne Bondi Johannessen and Fredrik Jørgensen (2005) make a similar point describing the Norwegian spoken language project TAUS from the 1970's. In this project, spoken language was described as a deviation from written. The project distinguished the following types of deviations from written grammar:

- lexical corrections
- syntactic corrections
- missing material
- interruption
- blend
- false start
- repetition
- exclamation

These types more or less match the rhetorical devices listed earlier. Interruptions and missing material were the commonest classes (50% and 30% of the lot, respectively). Missing material was mostly old information (often, closed category items), while interruptions were unpredictable from the context (new information was missing). Allowing for these deviations, 75% of the spoken language utterances were parsable with a written language grammar. The new NoTa project plans to turn the TAUS findings into a parsing procedure.

My plan is to extend my written language parser (Carlson 2005) with Lambek or Steedman style categorial grammar schemata constrained by associated conditions and penalties. For instance, an aposiopesis

```
This is a sentence and –
```

would be parsed by extending the parser (or grammar) with the Lambek categorial equation

$$x \Rightarrow y = x/z \Rightarrow y/z$$

so as to yield the extended grammar

```
S ⇒ S and S
S/S ⇒ S and S/S
S/S ⇒ S and
```

The penalty of the rule should be proportional to the recoverability of the divisor `z` in the context of the dialogue. Similarly, apokoinou can be represented by the Lambek style equation

```
x y\z = x/y y y\z
```

where B is the shared string, *koinon*. This makes apokoinou a recoverable special case of ellipsis, hence presumably associated with a lower penalty than the general rule.

Asyndeton is a special case of coordination. Repetition is the special case of asyndeton of identical constituents. Repetition of apparent non-constituents can be subsumed as the coordination of elliptic ones, in the style of Steedman (1990, 2000).

Repair is syntactically a coordination connected by with edit markers according to the Lambek style schema

```
y = x x\1 y
```

where `x\1` is the category of edit markers.

4. Disambiguation

It soon becomes obvious that the difficulty in spoken language parsing is not finding a parse, but choosing the right, or best one. Part of the problem is that formally, the languages generated by the different devices are not disjoint. Ellipsis and asyndeton alone are enough to cover all spoken language syntax (they allow generating the universal language W* from any grammar G on vocabulary W).

As the formal analysis shows, some of the devices can be treated as special cases or combinations of the others. Some cases of *anakolouthon* could equally well be treated as asyndeton with initial ellipsis, or just as a final and an initial ellipsis in a row. The choice between devices is sooner functional than structural. One of the first tasks in filling in the detail of the proposal would to clarify the structural and functional criteria of the taxonomy.

Even after that clarification, a problem of disambiguation is likely to remain. It therefore becomes important to find disambiguating clues. There are two approaches to this problem. One is to use higher level information from dialogue context to guess at the function of construction. This approach has been taken in the framework of conversation analysis. Norén (2003b) applies conversation analytic methods to sort out different rhetorical extensions, with particular reference to the *apo koinou* construction.

Another, more concrete avenue is to look for low level disambiguating information from phonetic realization. This comprises not only prosody but also timing.

Merle Horne (2005) reported on interruptions ending in a function word which promises a constituent beginning with it. She suggests that the length of the pause after the function word is related to length of phrase being planned. Another suggestive finding was that there are apparent built-in timing restrictions on speech processing. Apparently, there is a phonological loop of an approximate length of 2 seconds. Such chunks may contain silent pauses but not breathing. Thus a parser should predict phrase boundaries within a window of about 2 seconds from a breather or another such boundary.

John Local (2005) demonstrated further spoken language punctuation signs. In his English data, a glottal stop before a pause marks it not as turn final, while creaky voice allows an end of turn. Similarly, aspirated stops are turn final, whereas unreleased ones are not. Creaky voice seems to be a good predictor of end of turn in Finnish as well (Martti Vainio, p.c.).

Local's point for speech synthesis was that one should be able to predict finely tuned properties of phonetic realization from the interplay of language with the syntax of spoken interaction. Conversely, there is good hope that these and other phonetic indications serve to choose between competing parse hypotheses, possibly far better than written language punctuation.

For an automatic parser, a corpus of annotated text (a spoken language treebank) would allow looking for such operational contextual clues for the different constructions. Such a corpus is part of the PhD. thesis plan of Anna Dannenberg. A good guide to Finnish spoken language constructions now exists in the new large Finnish grammar of Hakulinen & al. 2003.

If the dialogue game approach described above is on the right track, spoken language parsing will not succeed well in a sentence window. It must take into account the dialogue game the speech forms a part of, and

pay attention to the fine phonetic realization of the speech signal, including absolute timing.

References

Allwood, Jens, Joakim Nivre & Elisabeth Ahlsén (1990) Speech management : On the non-written life of speech. *Nordic Journal of Linguistics* 13: 3–48.

Anward, Jens (2001) Talspråk och grammatik. *Språkvård* 3: 36–38.

Biber, Douglas (1986) Spoken and written textual dimensions in English: resolving the contradictory findings. *Language* 62: 384–414.

Carlson, Lauri (1983) *Dialogue Games: An approach to discourse analysis*. Dordrecht: Reidel.

—— (2005) CPARSE Manual. URL: http://www.ling.helsinki.fi/~lcarlson/cparse.html

Hakulinen, Auli, Maria Vilkuna, Riitta Korhonen, Vesa Koivisto, Tarja-Riitta Heinonen & Irja Alho (2004) *Iso suomen kielioppi*. Suomalaisen Kirjallisuuden Seuran Toimituksia 950. Helsinki: Suomalaisen Kirjallisuuden Seura.

Horne, Merle (2005) Speech technology in a general language processing framework. Paper read at NODALIDA 2005, Joensuu, May 2005.

Johannessen, Janne Bondi & Fredrik Jørgensen (2005) Annotation of spoken language data. Paper read at NODALIDA 2005, Joensuu, May 2005. URL: http://www.hf.uio.no/tekstlab/treebank_workshop/program.htm 31.05.2005

Lambek, Joachim (1958) The mathematics of sentence structure. *American Mathematical Monthly* 65: 154–170.

Linell, Per (1982) *The Written Language Bias in Linguistics*. Studies in Communication 2. Linköping: Department of Communication Studies.

—— (2003) Grammatiska konstruktioner i samtalspraktiken. In Bengt Nordberg, Leelo Keevallik Eriksson, Kerstin Thelander & Mats Thelander (eds.) *Grammatik och samtal: Studier till minne av Mats Eriksson*, pp. 161–171. Skrifter utgivna av Institutionen för nordiska språk 63. Uppsala: Uppsala University, Department of Scandinavian Languages.

Local, John (2005) What can talk-in-interaction teach us for speech synthesis? Paper read at NODALIDA 2005, Joensuu, May 2005.

Magnusson, Lars (2005) Grammatikaliteten i talad svenska. Unpublished paper. URL: http://hem.passagen.se/larsmagnussonb/ 31.05.2005

Miller, Jim & Regina Weinert (1998) *Spoken Spontaneous Language: Syntax and discourse*. Oxford: Oxford University Press.

Norén, Niklas (2003) *Apokoinou i samtal. Samtalsgrammatisk konstruktion och resurs för komplexa kommunikativa handlingar*. Arbetsrapporter från Tema K; 2003:2. Linköping: Tema Kommunikation, Linköpings universitet. Also URL: http://www.tema.liu.se/people/nikno/texter.html.

Steedman, Mark (1990) Gapping as constituent coordination. *Linguistics and Philosophy* 13: 207–263.

—— (2000) *The Syntactic Process*. Cambridge, MA: The MIT Press.

Teleman, Ulf (1983) *Har tal- och skriftspråk olika grammatiker?* Skrifter från institutionen för nordiska språk i Lund. Nordlund 3. Lund: Institutionen för nordiska språk i Lund.

Contact information:

Lauri Carlson
University of Helsinki
Department of Linguistics
FI-00014 Helsingin yliopisto
lauri(dot)carlson(at)helsinki(dot)fi
http://www.ling.helsinki.fi/~lcarlson