

# Syntax of Spoken Language

Lauri Carlson

University of Helsinki

Department of General Linguistics

lauri.carlson@helsinki.fi

This note is an application of the dialogue game model of Carlson (1983) to the syntax of spoken language. My hypothesis in this note is that syntaxes of spoken and written language are closer to one another than first meets the eye.

In my view, spoken language differs from written language principally in that explicit utterances in it can span several dialogue situations, (Carlson 1983), not all of them explicit, and conversely, an explicit dialogue situation may manifest fragmentary traces of several implicit ones. From this point of view, spoken language grammar as developed here is but a projection of a largely implicit dialogue game into one explicit play of one.

This view thus takes view of discourse as a partly submerged ‘iceberg’, well formed if it can be expanded to a complete dialogue game, down to the level of individual utterances. The bits and pieces of spoken language surface syntax are not primarily governed by a special looser spoken language syntax, but result from an interplay of dialogue game structure with more or less written language like phrasal syntax. (There may be spoken and written language specific constructions too.)

An example of this is the way Donald Duck’s nephews talk in turns, constructing a well-formed sentence in cooperation. It has been pointed out (Auli Hakulinen, p.c.) that spoken language syntax will appear more complete and connected if one does not restrict the unit of analysis into a single turn, but considers fragments and their context together. Syntax is reconstructed as the result of applying a shuffle operator on different participants’ moves and silence.

The idea of considering spoken language an extension of the written (or conversely, written language as a regimentation of the spoken) is not novel: in fact, one can consider it implicit in Aristotelian rhetorics. Similar ideas have been presented more recently as well, e.g. Sandig 1978 (reported in Saukkonen 1984:

Formally, from the point of view of automatic parsing, this 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 to a subset of written language grammar. Such processes include stylistic devices named in the long rhetorical tradition:

1. ellipsis
2. asyndeton
3. anakolouthon
4. repetition

## 5. repair

**Ellipsis** leaves out a beginning or end of a sentence, or more generally, allows suppressing (usually low-information) parts of one. Ellipsis is really needed only for clearly unfinished phrases; noun phrases for instance can be parsed as such. Formally, a quotient operator  $A/B$  or  $A \setminus B$ , alias categorial grammar composition. For instance Swedish spoken language syntax seems to allow dropping sentence initial grammatical words like conjunctions and pronouns more freely than written language. Ellipsis alone is enough to cover all spoken language syntax (it allows generating the universal language  $W^*$  given any grammar  $G$  on vocabulary  $W$ ). It therefore makes sense to look for specialisations.

**Asyndeton**, or unmarked coordination, allows chaining similar phrases without syntactic glue. Formally, concatenation  $A.B$ .

**anakolouthon (apokoinou)** is the result of starting a sentence in one way and ending it in another, with a shared part in the middle. In grammar terms, it can be represented as a regular expression  $A/B.B \setminus C$  where  $A$  is the shared string in the anakolouthon. Many cases of anakolouthon could equally well be treated as asyndeton with anaphor drop, or just an end and a beginning ellipsis in a row.

**Repetition** is just what the name suggests: a part of a sentence is repeated several times, perhaps because it is difficult or because it is important. Formally  $AB^{\bar{C}}$  equals  $ABC$ .

**Repair** is an attempt to take back part of a sentence and replace it with another part. Can't be done, for *quod semel emissum est volat irrevocabile verbum*. Instead editing noises are made to indicate what is to be ignored or replaced. Syntactically a coordination conjoined with the edit markers.

In transducer terms, this corresponds to  $A.B.B^{\bar{1}}.C$  which equals  $AC$ . Examples:

- 1) Ellipsis: vad är huvudstaden? -- washington disi.
- 2) Asyndeton: å så skulle vi in | börja riinga på ringklockan | han öppna inte dörren
- 3) anakolouthon: vi bodde på ett hotel som **hette tembo hotel**. låg alldeles på stranden...

The dialogue game idea of reconstruction is essential in cases where the explicit concept does not supply the missing part, but it can be inferred:

- A. Jag har ringt å ringt ann- christin men ja kan inte begripa var hon håller hus
- B. B: -- stuckit hem till rom

Here, we are entitled to interpolate the question A': vad har hon gjort? Hon har... to fill in the ellipsis. Thus conversely, this model of spoken language syntax also supports the reconstruction of dialogue structure.

å de här ä ingenting som man vill ha upp I skallen, för då **tror man kan** påverka ens humör.

This is a clear anakolouthon (just one man instead of two).

**Han som har hand om nödslakten i uddevalla** | hans pappa de va en sån här riktig galning som bara jobba...  
då ringde dom till mej ett par ganger **tjejerna själva ...**  
Everyone **knows** Helen Liddell **how hard she works**

These are left and right dislocation, respectively, but they also found in literary written language. The last case is grammatical except punctuation, functionally a sentence internal left dislocation.

De handlar då om en tjej som **jag inte kommer ihåg va hon heter---**  
så när rom väl spytt så ha rom gjort de så få rom gå å hämta sina påsar själv  
**som dom stackars anställda får gå å städa på golven**  
I only wear shoes **that I'm not thrown forward on my toes.**  
Some people lift the phone **that you think you've interrupted a suicide attempt.**

These are loose relative clauses: there is no syntactic trace or no correlate.

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...

Here is a repair marked by *eller* followed by an anakolouthon. (han sa ... sa han).

The above deviations from sentence syntax can be simulated with general rules without changing the object grammar, for instance anakolouthon can be parsed by an all-paths bottom up parser which produces the overlapping maximal partial parses.

An example analysis of a complete recorded phone dialogue in Finnish can be found in Appendix A, with a syntactic reconstruction using the above process primitives. Morphological spoken language features are not heeded.

It soon becomes obvious that finding plausible parses for spoken language utterances is not difficult, given the suggested traditional rhetorical extensions to grammar. Rather, the problem is the opposite: there is an *embarras de richesse* of competing reconstructions. Additional sources of information are needed to choose between hypotheses.

Norén (2003b) applies conversation analytic methods to sort out different rhetorical extensions, with particular reference to the *apo koinou* construction.

For an automatic parser, a corpus of annotated text (a spoken language treebank) would allow looking for more mechanisable contextual clues for the different constructions. This is where a spoken language treebank enters the scene.

## References

Carlson, Lauri. Dialogue Games: An Approach to Discourse Analysis. Reidel, Dordrecht 1983.

Hakulinen- Vilkuna- Korhonen- Koivisto- Heinonen- Alho, Iso suomen kielioppi. SKS 2004

Norén, Niklas, 2003a: Apokoinou som metod för hantering av turkonkurrens och samtidigt tal i samtal. I: Bengt Nordberg, Leelo Keevallik Eriksson, Kerstin Thelander, & Mats Thelander (red.), Grammatik och samtal. Studier till minne av Mats Eriksson. Uppsala: Institutionen för nordiska språk, Uppsala universitet.

Niklas Norén, 2003b: Apokoinou i samtal. Samtalsgrammatisk konstruktion och resurs för komplexa kommunikativa handlingar. (Arbetsrapporter från Tema K, 2003:2), Tema Kommunikation, Linköpings universitet. [PDF-fil]

Saukkonen, Pauli, 1984. Mistä tyyli syntyy. WSOY 1984.  
<http://www.tritonia.fi/vanha/ov/wsoy/luku5.htm>

Sandig, Barbara, 1978. Stilistik. Berlin 1978.

## Appendix A

Syntactic analysis of a sample dialogue from the Interact project corpus (City of Helsinki transport helpdesk)

clip 1 00:01:35:48 - 00:01:50:12 (nro 1)

L. [Liikenneneuvonta.][Puhelu äänitetään asiakaspalvelun parantamiseksi.] ...  
[Haloo?]

A. [<< M. I. >> päivää.]

L. [Päivää.]

clip 1 00:01:50:12 - 00:02:01:24

A. [On sellanen asia kun mun pitäs päästä Nuottalahdesta **sinne Nihtisillankujalle nel(jältä?)**] [**sinne Kiloon** niin mitenkä mä parhaiten pääsen sinne julkisilla - ?

*Asyndeton. Allowed by written syntax except for punctuation perhaps.*

Comments (Häiriöitä.)

clip 1 00:02:01:24 - 00:02:35:47

A. ... kulkuneuvoilla?]

*Interruption, grammatical without it.*

L. [**Sinne ei - sinne ei** kyllä ihan suoraan mene,] [katsotaan,] [hetkinen vaan.] ... [Mihin aikaan päivästä te muuten olisitte menossa?]

*Repetition. Ellipses.*

A. [No siinä joskus aamupäivällä.]

L. [Joo] ... [Tässähän olis kyllä **tää - tää** viitonen mikä menee ruuhka- aikana sinne]

*Repetition*

A. [Joka tarkoittaa aamua?]

*Ellipsis. Grammatical together with previous sentence, if not allowed by written syntax as is.*

L. [Se tarkoittaa sitte aamua ja iltapäivää]

clip 1 00:02:35:47 - 00:02:50:37

A. [Joo]

L. [Ja sit 35 menee kyllä **tähän Nihtisiltaan siihen aika lähelle**] [**tähän Nihtisillan...**]

A. [joo]

L. **... kohdalle ihan,**] [et tällähän pääsis kyl suoraan sitten.]

*Asyndeton.*

A. [Koska vaan]

L. [Se menee koko päivän] [joo]

A. [joo]

clip 1 00:02:50:37 - 00:02:59:36

A. [No mut sitä **kannattaa** - ]

*Ellipsis (cut off by next speaker)*

L. [Onks teillä se aikataulu?]

A. [Ei mull oo tässä aikataulua - ]

L. [Mihin aikaan teidän pitäis **olla?**]

*Ellipsis (adverbial, old information)*

A. [No varmaan mä lähtisin siinä joskus kymmenen, yhdentoista aikaan]

clip 1 00:02:59:36 - 00:03:15:47

L. ... [Täsön Matinkylän Kalastajanmäestä **lähdössä 10.10,** -

*Ellipsis (subject, old information)*

A. [joo]

L. ... 10.40, 11.10, 11.40.]

A. [Joo nää varmasti riittää.] [Mitenkäs sit sieltä pääsee takasin,] [**onks sitä- ?**]

*Ellipsis*

clip 1 00:03:15:47 - 00:03:28:02

A. [**miten -**]

*Ellipsis*

L. [No tää tulee täältä Kauniaisten kautta ja pysäkki tulee sit siihen] - [**tää tosiaan tää ohitusaika** on tosta Kauniaisista Tunnelitieltä et se on varmaan noin kolme minuuttia tohon Nihtisiltaan sitte]

*Asyndeton. Grammatical in written language except punctuation, spoken language et is a consecutive conjunction alone, written language might require niin että.*

A. [joo]

clip 1 00:03:28:02 - 00:03:40:36

L. [Mihin aikaan te olette tulossa?]

A. [No varmasti siinä ei mene kun **oisko** puol tuntii tai jotain semmosta] [ja sitten tulisin takasin] - [kun mä **käyn hakemassa mun**] - [**tai meen viemään** passin -

*Oisko is a spoken language adverb. Repair, grammatical except for ellipsis of object.*

L. [just joo]

A. - **sinne passitoimistoon**]

*Discontinuous move. Grammatical without the(apparent?) interruption.*

clip 1 00:03:40:36 - 00:03:53:06

L. [Ja tässon Kauniaisten kohdalla 10.26, 10.56, 11.26 ja 56] [**et ihan tätä -**

A. [Joo] - [just et se on] -

L. - **puolen tunnin väli näin päivällä**]

*Repair (tätä to be ignored. Edit marked by a pause.)*

A. [just joo]

clip 1 00:03:53:06 - 00:03:58:21

A. [No mut tää oli hirveen hyvä.] [Kiitos paljon]

L. [Kiitos]

A. [Kiitos kuulemiin]

L. [Kuulemiin]

clip 1 00:03:58:21 - 00:04:11:44 (nro 2)

(lapsen itkua, puhelimesta tai taustalta)

L. [Liikenneneuvonta.] [Puhelu äänitetään asiakaspalvelun parantamiseksi.]

A. [Päivää.]

L. [Hyvää päivää.]

clip 1 00:04:11:44 - 00:04:26:35

A. [**Tuota Tikkurilaan täältä Myyrmäestä?**] ... [Peltolantie 2 << D >> on tää osote.]

*Ellipsis (verb, inferrable)*

L. [Ahah,] [joo,] [no 55 tulee sit siihen Kielotielle.]

clip 1 00:04:26:35 - 00:04:38:15

A. [**Kielotielle?**] [Miss] - [**onkse Peltolantie missä se on,**] [onkse ihan siinä] -

*Ellipses (false starts). A possible anakolouthon around Peltolantie.*

L. [**No se on**] - [**se tulee siihen**] - [**se on**] - [mun täytyy ihan katsoa et onkse ihan peräti poikkitie siinä] - [miten se meni,] [hetkinen]



*Ellipses (false starts).*

A. [joo]

clip 1 00:04:38:15 - 00:05:01:04

L.... [Se oli Peltolantie 2?]

A.[Niin]

L.... [No tää tulee aika lähelle kyllä,] [tässon **tää kävelykatu tää Tikkuraitti** siinä välissä]

*Asyndeton. Grammatical in written language except for punctuation.*

clip 1 00:05:01:04 - 00:05:05:00

A. [Joo] [**elikkä viisvitos ella**] - [**mites ku** mun pitäis olla 12.30 siellä?]

*Ellipses. These types found in written language as well.*

clip 1 00:05:05:00 - 00:05:16:02

A. ... [Milläs bussilla mä **ehin?**]

*Ellipsis (if even that).*

L. [**No se on Myyrmäestä on** ykstoista nollakaksi.]

*anakolouthon round Myyrmäestä.*

clip 1 00:05:16:02 - 00:05:23:54

A. [**Joo ei**] [**mut ei**] [täytyyhän sielt myöhemminkin lähtee eiks niin?]

*Ellipses (allowed by written syntax).*

L. [No sit on 11.30 seuraava.]

A. [Joo- o?]

clip 1 00:05:23:54 - 00:05:37:34

L. [Et kylhän tää nyt menee sinne sitten]...

A. [Paljon **sinne** menee muute **Tikkurilaan** täältä?]

*Double adverb (allowed by written syntax).*

L. [Kyl se sen puol tuntia voi mennä]

A. [**Meneeks kaheltatoista?**]

*Ellipsis (allowed by written syntax).*

L. [No sit menee 12.02]

A. [No sillä mä - **yri- yrittää** sit - **jos kaks**] - [se on varmaan kakstoista kolmekymmentä **siellä on**] - ?

*Repair, ellipsis, anakolouthon round siellä. The morphological recogniser could be extended with initial reduplication to recognise false starts.*

L. [No kyllä tällä varmaan ehtii sit] - [Tänään on kyllä huono keli **et**]

*Ellipsis. et is a consecutive conjunction in spoken language, perhaps also a particle.*

clip 1 00:05:37:34 - 00:05:43:58

A. [jaa?]

L. [On kaikkennäköstä viivästymistä ollu]

A. [Joo] [**eli viisivitonnen sitten.**] [Kiitos sulle]

*Ellipsis, or a free noun phrase.*

L. [Kiitos]

A. [joo hei]

L. [Hei]

(nro 9)

L. [Liikenneneuvonta.] [Puhelu äänitetään asiakaspalvelun parantamiseksi]

A. [Ai jaaha,] [no kiitos juu,] [mä tota jo ajattelin et oottekste **hylänny**] - [hetk(in)en vaan (?)] - [hylänny kun tota mä yritin sinne ja sielt tuli tyyttyyttää] [ja mä oisin kysynyt semmosta bussiyhteyttä **kun Suomensuontie 1 Laajasalosta niin ku menee Malmin sairaalaan**] [no se on 86 joka menee sieltä Laajasalosta] [ni millon voi tota noin niin vaihtaa **mihin mistä mihin** tota et pääsee Malmin sairaalaan]

*Repair, marked by interjection. Niin is a resumptive particle, ku a relative pronoun. Repair in mihin – mistä mihin - .*

L. [No nyt on sitten ] [Jos tulette metrolla niin voi tulla Sörnäisiin **ja**] –

*Ellipses: self-interruption (false take) and interruption.*

A. [Niin mut kun ensin täytyy mennä tota tällä bussilla tohon Herttoniemen metroon] [et onko siinä mahdollisuutta?]

L. [Niin, niin no siin täytyy ottaa se metro väliin] [et siit ei suoraan mene ihan sinne] - [Malmille menee kyllä 79 mut tää ei mene ihan sairaalalle, et se menee siihen Malmin asemalle]

A. [**Malmin asemalle?**] [Juu] - [no juu sitä just mentiin ja oli niin kauhea työ eikä tiedetty **mikä on siitä sitte jatkais** mm tota noin niin **siihen**] -

*Ellipsis (or free noun phrase), anakolouthoun around siitä, ellipsis (interruption).*

L. [Niin se on] - [**se tulee nimittäin kun tää tulee** Pukinmäen kautta niin se **menee menee** sitten tavallaan **siihen Malmin tän aseman**] [siinähan on se terminaali niin sen alapuolella on pysäkki et siint joutuis ne portaat nousemaan ylös]

*False start, repetition, ellipsis (self-interruption)*

A. [Ai jaa] [sen on niin vaikee, mul on jalatkin niin huonot **ja muut**]

*Ellipsis (allowed in written syntax too).*

L. [Mut **miten miltä** tuo metro tuntuu] [et jos tulette sillä tuohon Sörnäisiin tai Hakaniemeen jos se on tutumpi paikka]

*Repair*

A. [No Hakaniemi ois melkeen **tutumpi juu**]

*Asyndeton*

L. [Joo,] [niin siitä sitte pääsette bussilla 73]

A. [73?]

L. [Joo] [se lähtee Hakaniemestä ihan ja tulee sit **siihen ku tulette metrolta ulos siihen bussipysäkeille** torin laitaan **niin siitä**]

*Asyndeton, ellipsis. kun relative pronoun*

A. [Se on siinä torin laidassa tää 73 lähtee?]

*Asyndeton*

L. [Kyllä]

A. [Ettei se oo niinkään siitä kun **on toisia busseja menee** siitä Metallitalon edestä?]

*anakolouthon round toisia busseja*

L. [Ei vaan se tulee ihan tähän kun tullaan metrosta ulos]

A. [*Se* tää menee sit ihan sairaalaan?]

*Asyndeton*

L. [Kyllä,] [sillon pysäkki siin sairaalan kohdalla]

A. [Aha,] [no se onkin paras juttu sitten,] [että tota niinkun **se on niin vaikee sinne kun mentiin** tässä näin niin **siinä sai bussia ja asemalla ja** (ei tietty?) mihin me mennään ja kukaan ei tiennyt mitään kun kysy ja sitte piti olla taksia siin välissä ja] -

*Asyndeton, ellipses*

L. [Voi harmi,] [siin **ois kyllä tämmönen 69 lähtee** sitten ihan siitä 79:n päätepysäkin läheltä ja sillä pääsee sit myös siihen Malmin sairaalaan] -

*anakolouthon (or asyndeton: ellipsis of relative pronoun?)*

A. [Ai jaa 79 siihen asemalle ja sit 69?]

L. [Joo]

A. [Onks ne samalla pysäkillä?]

L. [No se ei oo **ihan ? ihan** sama mut ne on kyllä ihan lähekkäin]

*Repetition*

A. [Lähekkäin siinä niin?] [No juu kiitos nyt tästä avusta.]

L. [Kiitos, kuulemiin]

A. [Kuulemiin]