Annotating Turn Competition in Multi-Party Conversations

The use of inter-annotator agreement in Conversation Analysis

Myrthe Faber
University of Groningen / University of Sheffield
Statistics Seminar 13.05.2009

Overview

- › Overlapping speech in conversations
  - What is turn competition?
- › Identification of turn competition in previous studies
  - Classification
    - Interruption studies
    - Conversation Analysis
  - Evaluation of classification
- › This study: intuitive coding vs sequential coding

Extract 1. Ford & Thompson 1996 (p.151)

```
K: It was like the other day uh. (0.2)
    Vera (.) was talking on the phone to
    her mom?
C: Mm hm.
K: And uh she got off the pho:ne and she was
    incredibly upset?
C: Mm hm.
```

Turn taking in conversations

Two observations:

1. People usually do not talk at the same time.
2. People usually do not leave many long silences.

$\Rightarrow$ They minimize overlap and they minimize silence.

Extract 2. Fragment from the LA phone call corpus

```
Cor:   I: didn't tell her you were:. I just said
       yeah, Angela thinks (you'uv) some of her
       stuff! you know.=
Ang:  =well then: she took it the wrong way an
       I'd like you to like clear it up,
       (.)
  →   °for    [me.°] (h)
Cor:          [I wi]ll but- (0.4)
       °<I honestly don't think that->° (1.4)
```

Extract 3. from Annotation Guidelines (Kurtic, unpublished)

```
A :     the network group is almost entirely Germans and
        Spaniards.
B :     Well Oh. But the thing is, I think that these
        people are of high enough level in their in their
        language [PROFICIENCY THAT]
A :                [    I SEE        ].
B :     And I 'm not objecting to accents.
```

Extract 4. from Annotation Guidelines (Kurtic, unpublished)

```
B : He You mentioned this
    [LAST TIME, THAT THAT IF IF YOU 'RE STRAIGHT DOWN THE
     MIDLINE],
C :[YEAH, WE HAV NEED TO PUT IT ON A LITTLE TURNTABLE],
B : then then the r the left - right 's gonna be
    different,
D : I I I I I th
E : Well, it's-
B : and and and in his case, I mean, he 's closer to it
    anyway.
```
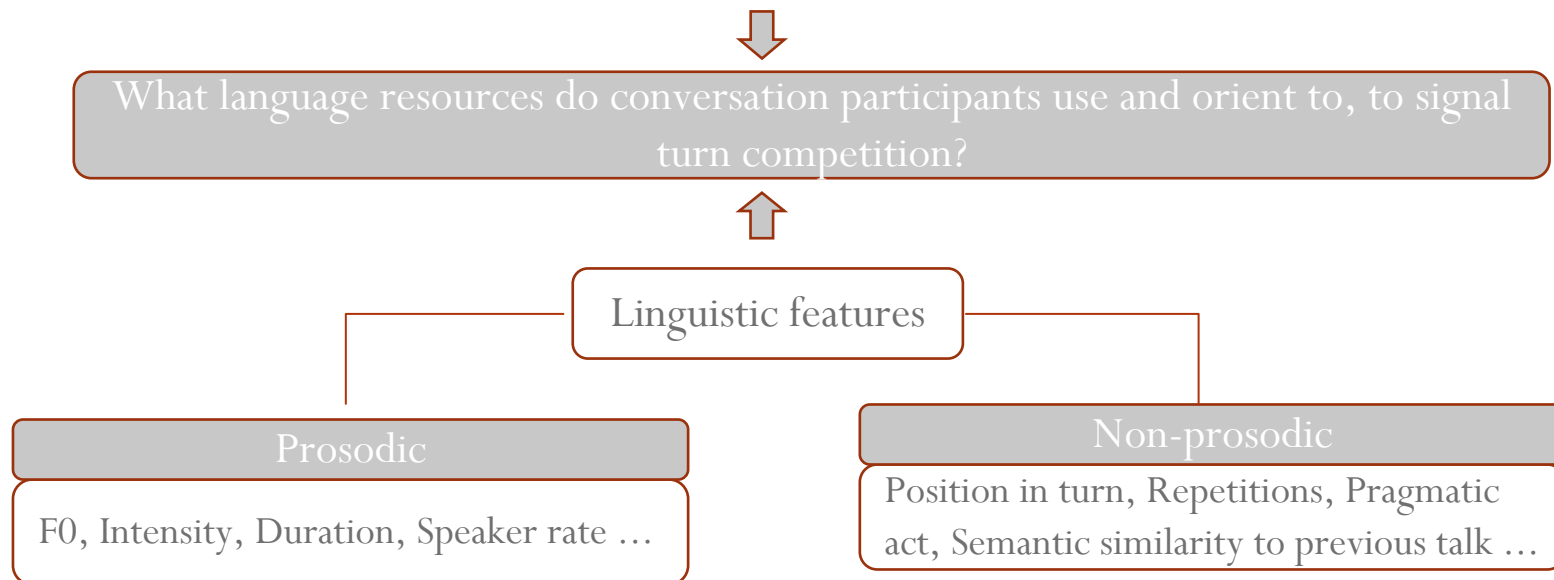
university of
groningen

Overlapping speech can be turn competitive or non-competitive

What language resources do conversation participants use and orient to, to signal turn competition?

Linguistic features

Prosodic

F0, Intensity, Duration, Speaker rate …

Non-prosodic

Position in turn, Repetitions, Pragmatic act, Semantic similarity to previous talk …

(Kurtic, unpublished)

What is turn competition?

Preliminary definition:

Turn competition is a conversational action that takes place when either or both conversation participants demonstrate the aim to prevent the other party from either keeping or taking over the current turn **(Kurtic, unpublished)**.

Extract 5. Segment 275 from ICSI Bmr_008

```
me011:   and now list the ones in first grade.
me011:   and now list the ones your frien[ds (.)speak  ]
me018:                                 [     Sort    ] of
         like as soon as you get to the cases on the edge
         the complexity just shoots[ up.  ]
fe008:                             [ Yeah ] that 's right
me011:                             [Right ]
fe016:  Mm - hmm.
```

Extract 6. Segment 461 from ICSI Bmr_008

```
me013:        And uh what the [so I'd]
me018:                          [ What ] about
              that error that (.) that (0.3)
              uh (.) the (0.3) the supposed lub
```

Extract 7. Segment 61 from ICSI Bmr_008

```
fe008:      You can always get more specific.
fe008:      An[d  it  may  be ]
me011:        [I mean so what ] would you suggest
me018:      I don't know.
```

Research on turn competition in three fields:

Studies on conversational interruption	Aim: Do men interrupt women more frequently than other way around?
Conversation Analysis (CA)	• Interest in overlapping speech as conversational phenomenon • Investigate linguistic resources that participants use to signal competition in overlap & strategies for overlap resolution
Computational Discourse Modelling	**Dialogue Act annotation** of large conversational corpora (e.g. Switchboard (Jurafsky et al. 1998), ICSI Meeting Corpus (Shriberg et al. 2004)) for training and testing statistical models

(Kurtic,unpublished)

Interruption studies – Methods of classification

Annotation scheme in the form of a decision tree:

- Set of if-then rules (Roger et al. 1988, Beattie 1981)

  - First (relevant question) in scheme:

Did the second speaker disrupt the first speaker's utterance?

$\Rightarrow$ The annotator has to have some intuition about what it means for an overlap to be interruptive.

Appendix Flow-chart for scoring simultaneous speech using the Simultaneous Speech Coding System
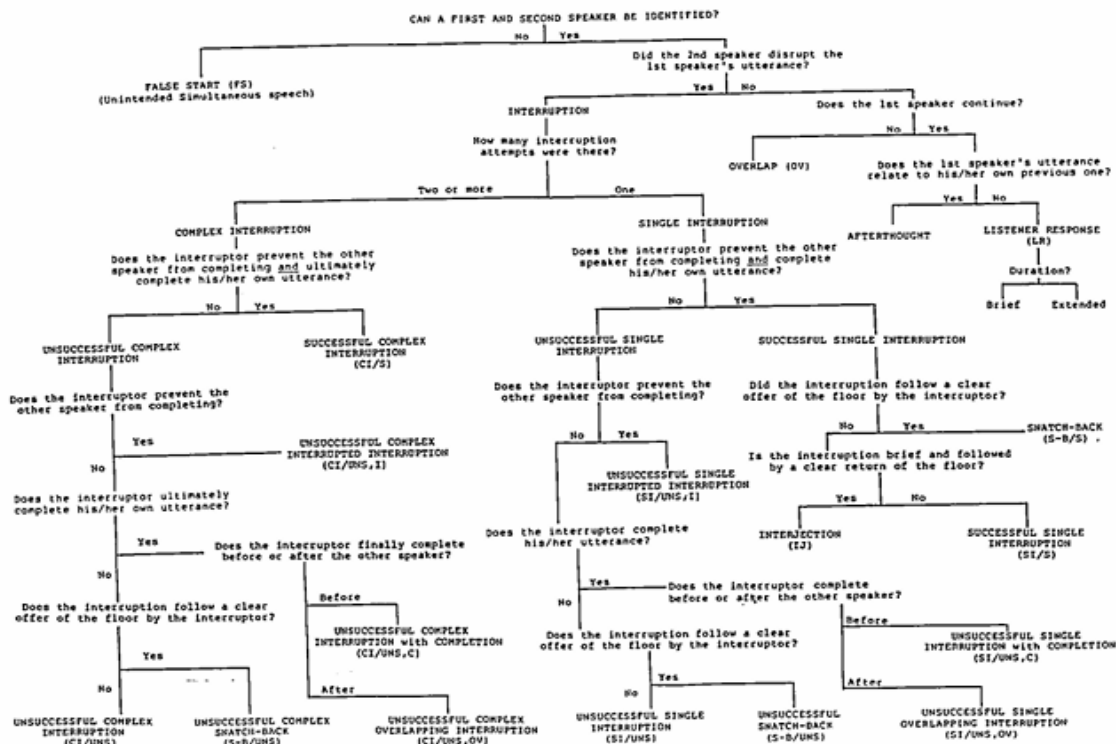
Interruption studies – Methods of evaluation

Report %-age agreement between two independent annotators
- Roger et al. (1988): 75-95%; Beattie (1981): 88%

Report agreement between two independent annotators as Cohen's kappa
- Beattie (1981): kappa = .85

Conversation analysis – Method of "classification"

- Grounded in the principle of CA (Heritage 1989):

Contributions to interaction are both context-shaped and context-renewing.

$\Rightarrow$ Each turn is a part of conversational sequence in which it occurs, and the action it incorporates can only be interpreted based on how conversation participants themselves interpret it in that particular sequence (Kurtic, unpublished).

Conversation analysis – Method of classification

- Consequences:

$\Rightarrow$ Categories can only be made if they are justifiable by their relevance to participants

$\Rightarrow$ The process of arriving at the categories is mostly more revealing of the structure of conversation than the categorization itself (Wootton 1989)

$\Rightarrow$ Classification (coding or annotation) is rarely reported in CA studies

Conversation analysis – Method of "classification"

› What is competitive overlap Schegloff (2000, 2002):
- Overlaps in which the conduct of participants indicates that they are treated as **problematic**
- *To ask whether something "is" an interruption is to ask whether it is [..]* **complainable...** (Schegloff 2002)
- There are classes that are generally non-competitive: Continuers, Choral and Collaborative productions and terminal overlaps
  - However, occasionally, these can also be treated as problematic and thus competitive
  - $\Rightarrow$ It is the conduct of participants in the conversation sequence that decides on overlap competitiveness

Conversation analysis – Method of evaluation

Basing analytic claims in demonstrable actions of conversation participants gives ground to assume the reliability of analytic claims a-priori (Wootton 1989)

ICSI Meeting Recoder dialogue act annotation (Shriberg et al. 2004)

Competitiveness classification - summary

› There are two different ways of identifying turn competition

Analysts' intuition:

Both dialogue act classification and interruption studies essentially rely on the fact that analyst's intuition is good enough to discriminate between competitive and non-competitive incomings

Participants' orientations:

CA based approaches ground their decisions on competitiveness in sequential analysis that reveals whether an overlap is treated as "problematic" by participants

Competitiveness classification - summary

> There are two different ways of defining reliability (Wootton 1989)

- Agreement metric

- Basing analytic claims on observable actions of conversation participants

› **This research (Kurtic, unpublished):**

› How good are people really in distinguishing between competitive and non-competitive overlaps?

› Does this differ when we let them decide based on intuition as opposed to deciding based on analysis of conversational sequence?

› In both cases Cohen's kappa statistics is used to measure the agreement between annotators as the indicator of how good the annotation is.

Intuition based competitiveness classification

› Online experiment:
- 10 participants
- Presented with transcript, competitiveness definition and able to listen
- 40 overlaps drawn from the set of 665 overlaps precategorized for competitiveness:
  - 20 competitive
  - 20 non-competitive
  - In each group 10 shorter than mean duration and 10 longer, to avoid bias towards short overlaps that are more frequent
  - Chosen by random selection

› http://ext.dcs.shef.ac.uk/~u0065/main/start.jsp

Intuition based competitiveness classification

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
A1	X	.55	.65	.19	.60	.44	.35	.39	.39	.39
A2		X	.51	.37	.56	.22	.60	.28	.28	.57
A3			X	.23	.64	.28	.50	.22	.33	.33
A4				X	.25	.41	.55	.34	.45	.23
A5					X	.51	.45	.34	.45	.45
A6						X	.20	.39	.39	.27
A7							X	.55	.35	.35
A8								X	.20	.32
A9									X	.32
A10										X

Intuition based competitiveness classification

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
A1	X	.55	**.65**	.19	**.60**	.44	.35	.39	.39	.39
A2		X	.51	.37	.56	.22	**.60**	.28	.28	.57
A3			X	.23	**.64**	.28	.50	.22	.33	.33
A4				X	.25	.41	.55	.34	.45	.23
A5					X	.51	.45	.34	.45	.45
A6						X	.20	.39	.39	.27
A7							X	.55	.35	.35
A8								X	.20	.32
A9									X	.32
A10										X

Intuition based competitiveness classification

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
A1	X	.55	.65	.19	.60	.44	.35	.39	.39	.39
A2		X	.51	.37	.56	.22	.60	.28	.28	.57
A3			X	.23	.64	.28	.50	.22	.33	.33
A4				X	.25	.41	.55	.34	.45	.23
A5					X	.51	.45	.34	.45	.45
A6						X	.20	.39	.39	.27
A7							X	.55	.35	.35
A8								X	.20	.32
A9									X	.32
A10										X

Intuition based competitiveness classification

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
A1	X	.55	.65	**.19**	.60	.44	**.35**	**.39**	**.39**	**.39**
A2		X	.51	**.37**	.56	**.22**	.60	**.28**	**.28**	.57
A3			X	**.23**	.64	**.28**	.50	**.22**	**.33**	**.33**
A4				X	**.25**	.41	.55	**.34**	.45	**.23**
A5					X	.51	.45	**.34**	.45	.45
A6						X	**.20**	**.39**	**.39**	**.27**
A7							X	.55	**.35**	**.35**
A8								X	**.20**	**.32**
A9									X	**.32**
A10										X

Sequential analysis for competitiveness classification

> › 3 annotators from previous 10
> > • All had training in CA and/or specifically on this task
> > • 419 overlap instances from one meeting

Sequential analysis for competitiveness classification

› 3 annotators from previous 10
  • All had training in CA and/or specifically on this task
  • 419 overlap instances from one meeting

Results:

	Sequential (kappa)	Intuitive (kappa)
A5– A5	.71	.68
A5 – A7	.56	.45
A5 – A9	.63	.45
A7 – A9	.67	.35

Conclusions

- **How good are people in distinguishing between competitive and non-competitive overlaps?**

$\Rightarrow$ Not very good

$\Rightarrow$ The classifications thus lack reliability:

$\Rightarrow$ They are not grounded in analysis of participants' treatments of turns in the conversational sequence

$\Rightarrow$ They also don't offer good inter-annotator agreement

Conclusions

- **Does this differ when we let them decide based on intuition as opposed to deciding based on analysis of conversational sequence?**

  ⇒ Yes, the agreement seems to improve when longer conversational sequences are available and sequential analysis can be conducted

  ⇒ It seems that grounding decisions in sequential analysis can improve reliability of dialogue act classification in both senses of reliability

Thank you!