Conversations in party games: Construction and analysis of the Werewolf corpus

Laurent Prévot & Yao Yao

Variamu 4th Workshop, January, 6th-7th, 2016, Hong Kong
Overview

- Background
- Corpus construction
  - French
  - Chinese
- Preliminary analysis
  - General description
  - Turn-taking behavior in multi-talker conversations
- Future research
Introduction

Despite the large number of corpora, there is still a lack of corpora that document dialogues involving more than two interlocutors.

Why studying multi-talker conversations?

• Complex behavior regarding turn-taking, interaction among talkers [Grothendieck et al., 2009, Laskowski, 2010]
  • Social status, inter-personal relationship, role in conversation
  • Social convention, cultural norm, …
Difficulty in creating a multi-talker conversational corpus

Naturalness of speech
- Difficult to conduct a truly engaging discussion with subjects who may not be familiar with each other
- Thus naturally-occurring group meetings are often recorded

Technical requirement
- High-quality recording facilities
- Alignment of multiple recording channels
Interactional Linguistics

Very close to interactional linguistics [Couper-Kuhlen and Selting, 2001] and constrative conversational analysis [Maynard, 1990, Clancy et al., 1996]...

- how people are interacting with each other through language
- the study of the linguistic structures of such interaction

Focus on

- analysis of spontaneous spoken data
- objects studied are multidimensional (lexis, syntax and prosody,...)
  - turn-taking, discourse particles, discourse syntactic positions, repairs, fragments, spoken language constructions

but not sharing the methods:

- Conversational Analysis
- light-weight quantitative descriptions (sometimes)
Shortcomings and proposal

- datasets are too small for serious quantitative analysis
- analyses are generally sporadic and/or not replicated/replicable
- datasets are not comparable enough across languages

~

- joint design from beginning (including pilot studies and comparability checks on the way)
- large corpus collection
- automatic or replicable (at very least inter-coder agreement checked) analysis only

can be realized only within a long term stable joint team for each language considered
Quantitative Comparative Interactional Linguistics

- Quantitative requires significant amount of data (statistical significance)
- QCIL: Approach in a systematic data-driven way on large comparable corpora
- Existing works:
  - [Ward and Tsukahara, 2000]: Turn-taking and prosody in English and Japanese
  - [Levitan et al., 2015]: Entrainment in English, Mandarin, Spanish and Slovak
  - ...
General framework

Same situation encoded in comparable corpora
  • same communicative needs
  • same time pressure
  • same interpersonal relationships
  • (remain interindividual variation)

Significant differences observed due to:
  • linguistic / interactional structures
  • socio-cultural constraints

Commonalitites / Universals?
  • At interactional level [Levinson, 2006]
Current study

Werewolf Corpus = A corpus of multi-talker conversations in social situations
  • Natural speech elicited with the Werewolf game in a group setting
    • Conditions for both genuine/cooperative speech and deceptive speech
  • Parallel corpus
    • French
    • Chinese (Mandarin)
The Werewolf game

A popular role-playing game...

- **Roles**
  - 6 villagers, 1 or 2 (secret) werewolf, (1 police), 1 game master

- **Procedure**
  - Night and day alternate
  - during night, werewolf eat a villager
  - during day, everyone discuss to find the werewolves and then vote → persuasion, (dis)agreement, argument, deception...
  - Game ends when there is no more villager or werewolf
Pilot data (French)

- Recorded in LPL (Aix)
- 9 games, 26 participants
- Recorded with head-mounted microphones directly plugged into a Zoom mixer. Fully synchronized
- 3 games transcribed at the word level, total 55 min → used for current analysis
Example of French transcription

quoi que non tu es végétarien donc heu dan un jeu de

oh le gars oh cette logique en bois

et au lieu de c'est grillée + c'est comme je vais accuser quelqu'un

ah merde je me suis trompé
Pilot data (Chinese)

- Recorded in CBS language labs (PolyU)
- 4 games, 8 participants, total 48 min
- Recorded with microphones plugged to desktop computers. Recording process controlled by a classroom control software installed on all computers. Fully synchronized
- All 4 games transcribed at the word level → used for current analysis
### Example of Chinese transcription

<table>
<thead>
<tr>
<th>对(啊)应该</th>
<th>(mmm)是谁</th>
<th>这一排为什么</th>
<th>(哎哟)$cla$你</th>
</tr>
</thead>
<tbody>
<tr>
<td>这个学妹</td>
<td>(哦)LA</td>
<td>到底是*</td>
<td>(哦)</td>
</tr>
<tr>
<td>谁找的</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(嗯)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>我觉得如果既然杀学妹的话*</td>
<td>*谁找</td>
<td>LA</td>
<td>不是有两次(嘛)一个是杀</td>
</tr>
<tr>
<td>(哦)其</td>
<td>但是她刚刚把</td>
<td></td>
<td></td>
</tr>
<tr>
<td>我[是 sh]我</td>
<td>对就因为我是觉得是这这一</td>
<td>一个是杀手警察我</td>
<td></td>
</tr>
<tr>
<td>(哦)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

29.227299
Do people overlap with each other in group conversations?

- Percentage of the recordings when X speakers are producing sounds (including laughter [@] or not)

- Most of the time, more than one person is producing some sound...
Three types of utterances

- General idea
  - **Main speech** - expressing some thought
  - **Listener response** - e.g. “ouais”, ”hmm”, as response to a previous (or concurrent) speaker, often short and limited to a small set of lexical items
  - **Non-linguistic sounds** - mostly laughter

- Identifying listener responses
  - Core response items extracted from high-frequency short utterances (FR: ≤3 words; CH: ≤ 2 syll)

<table>
<thead>
<tr>
<th></th>
<th>FR</th>
<th>CH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mh, ouais, oui, d’accord, ok, voilà, bon, ben, ah, non, ...</td>
<td>ang, e, en, eng, mmm, woo, 啊, 哎, 敖, 嗯. 对, 好, 哦, 咦, ...</td>
</tr>
</tbody>
</table>

- How many different ways of saying *en*?
Three types of utterances (cont’d)

- Counts of different types of utterances

![Bar chart showing counts of different types of utterances for 'fr' and 'zh'.]
Four basic measures calculated for each utterance

- **t_offset**
  - duration of overlapped main speech (negative value) or silence (positive value)

- **nb_spkr_before**
  - # of spkrs talking within 2s before the current turn begins

- **nb_spkr_overlap**
  - # of spkrs overlapped with the current speaker

- **talking_before**
  - whether the current spkr is engaged in main speech within 800ms before the current utterance (0 or 1).
t_offset

![Box plot of t_offset](image)
Normalized nb_spkr_before
Normalized nb_spkr_overlap
Normalized talking_before

![Talking Before Diagram](image_url)
Turn-taking patterns

- **Chaotic period**
  - initiating a turn in a chaotic period (when many people are talking)
  - nb_spkr_before > 3

- **Legal turn-taking**
  - initiating a new turn of main speech after all other speakers finish
  - nb_spkr_before ≤ 3, talking_before = 0, t_offset > 1

- **Turn grabbing**
  - initiating a new turn of main speech before all other speakers finish
  - nb_spkr_before ≤ 3, talking_before = 0, t_offset < 0

- **Turn holding**
  - continuing a previous turn (by the same speaker), with or without someone else interrupting
  - nb_spkr_before ≤ 3, talking_before = 1
Turn-taking patterns (cont’d)

Counts of different types of turn-taking behavior

![Bar chart showing counts of different types of turn-taking behavior for 'fr' and 'zh' languages. The chart includes bars for 'chaos', 'grab', 'hold', and 'legal'.]
Summary

The preliminary analysis shows

- variety of utterances in spontaneous speech
- high frequency of cross-talk in group conversations

More questions

- Linguistic cues of turn-initiating and turn-grabbing?
- Do turn-taking patterns vary with roles in the game (werewolf, villager, etc.)?
- Do turn-taking patterns vary with other speaker characteristics (sex, language background, seating distance, etc.)?
- Do speakers maintain an underlying speech rhythm during turn transitions?
Features of the Werewolf Corpus

- Fiercely spontaneous and interactional language structures
- Comparable data in French and Chinese
- Speech data for analyzing attitudes and emotion (e.g. laughter)
- Variety of discourse types (deceptive speech, argumentation, etc.)
- Group evolution through the interaction
Next stage of the project

• Develop a common protocol to be used for data collection, annotation and storage on both sites
• Record and transcribe more games
• Topics for further research
  • Deceptive vs. genuine speech
  • Rapport vs. attack
  • Cross-cultural variation


References II

On the human” interaction engine”.

Acoustic-prosodic entrainment in slovak, spanish, english and chinese: A cross-linguistic comparison.
In 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue, page 325.