Toward a scary comparative corpus: The Werewolf Spoken Corpus
Laurent Prevot, Yao Yao, Arnaud Gingold, Bernard Bel, Kam Yiu Joe Chan

To cite this version:
Laurent Prevot, Yao Yao, Arnaud Gingold, Bernard Bel, Kam Yiu Joe Chan. Toward a scary comparative corpus: The Werewolf Spoken Corpus. THE 19TH WORKSHOP ON THE SEMANTICS AND PRAGMATICS OF DIALOGUE, 2015, Gothenburg, Sweden. pp.204, 2015. hal-01231889

HAL Id: hal-01231889
https://hal.archives-ouvertes.fr/hal-01231889
Submitted on 15 Nov 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Toward a scary comparative corpus
The Werewolf Spoken Corpus
Laurent Prévot, Yao Yao, Arnaud Gingold, Bernard Bel, Kam Yiu Joe Chan

1 Aix Marseille Université, Laboratoire Parole et Langage, France
2 The Hong Kong Polytechnic University, Chinese and Bilingual Studies, Hong Kong

Context
- Dialogue and multilogue differ significantly in terms of dialogue structure [3]
- Not much multilogue corpora available
- No comparable multilogue corpus available, see however [4] for English
- Existing multilogues [1, 5, 2] have been created within cooperative communicative situation
- Not much quantitative / systematic study of cross-linguistic / intercultural variation of this kind of natural data
- More theoretically: how different sources of variation blends within actual performances

Elicitation Protocol
- Role: 2 Werewolf, 6 villagers, 1 game master
- Alternation of night / day phases
  - Night: Werewolves decide to make disappear a villager (silent)
  - Day: everyone vote to kill someone (discussion phase)
- Game ends when either all villagers or werewolves have been killed

Recording Setting
- Usually difficult to record clean speech data in ‘one-location’ multilogues
- Here a circle of people (French), 2 rows of 4 people facing each other (Mandarin)

Objectives
- Explore linguistic and cultural variation is non-cooperative multilogue situation (deceptive speech, persuasive speech, competitive speech)
- First step: gather data and check first variables (overlap, turn-taking, amount of speech) variations across speaker situation

Data curation
The Wolrewolf pilot corpus and its annotations were stored and described so as to allow a wide dissemination

- http://hdl.handle.net/11041/ortolang-000900
- http://hdl.handle.net/11041/ortolang-000908

Basic info about a game
- So far 4 games for each language, 7-27 minute per game

References

Acknowledgements
This work has been carried out thanks to the support of the A*MIDEX Variam project (ANR-11-IDEX-0001-02), the Ortolang (ANR-11-EQPX-0032) French Government program and of the Faculty of Humanities of The Hong Kong Polytechnic University. We would like to thank all the students in Aix and Hong-Kong that participated in the recordings and in the transcription.