Toward a scary comparative corpus: The Werewolf Spoken Corpus

Laurent Prevot, Yao Yao, Arnaud Gingold, Bernard Bel, Kam Yiu Joe Chan

Toward a scary comparative corpus
The Werewolf Spoken Corpus

Laurent Prévot[1], Yao Yao[2], Arnaud Gingold[1], Bernard Bel[1], Kam Yiu Joe Chan[2]
[1] Aix Marseille Université, Laboratoire Parole et Langage, France
[2]The Hong Kong Polytechnic University, Chinese and Bilingual Studies, Hong Kong

Context

- dilogue 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 perfomances

⇝ Record and analyse a comparable corpus of Werewolf social game in France and in Hong Kong.

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 WEREWOLF 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`).

References

[1] Susanne Burger, Victoria MacLaren, and Hua Yu. The isl meeting corpus: the impact of meeting type on speech style. In *INTERSPEECH*, 2002.

[2] Jean Carletta, Simone Ashby, Sebastien Bourban, Mike Flynn, Mael Guillemot, Thomas Hain, Jaroslav Kadlec, Vasilis Karaiskos, Wessel Kraaij, Melissa Kronenthal, et al. The ami meeting corpus: A pre-announcement. In *Machine learning for multimodal interaction*, pages 28–39. Springer, 2006.

[3] Jonathan Ginzburg and Raquel Fernández. Action at a distance: the difference between dialogue and multilogue. In *Proceedings of DIALOR*, volume 5, 2005.

[4] Hayley Hung and Gokul Chittaranjan. The IDIAP wolf corpus: exploring group behaviour in a competitive role-playing game. In *Proceedings of the international conference on Multimedia*, pages 879–882. ACM, 2010.

[5] Adam Janin, Don Baron, Jane Edwards, Dan Ellis, David Gelbart, Nelson Morgan, Barbara Peskin, Thilo Pfau, Elizabeth Shriberg, Andreas Stolcke, et al. The icsi meeting corpus. In *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*, volume 1, pages I–364. IEEE, 2003.

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

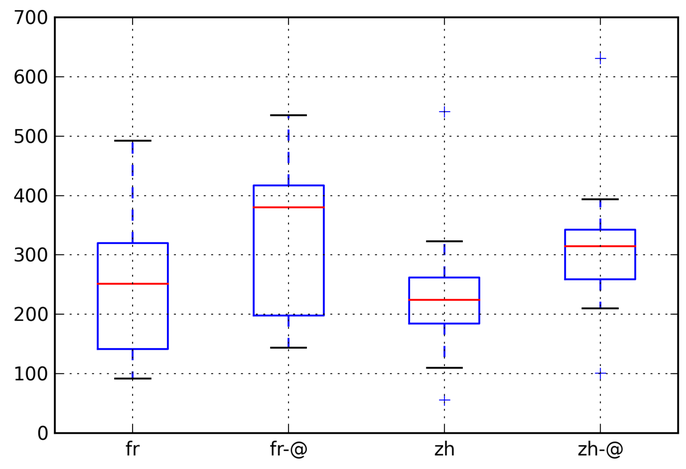
Recording Setting

French side

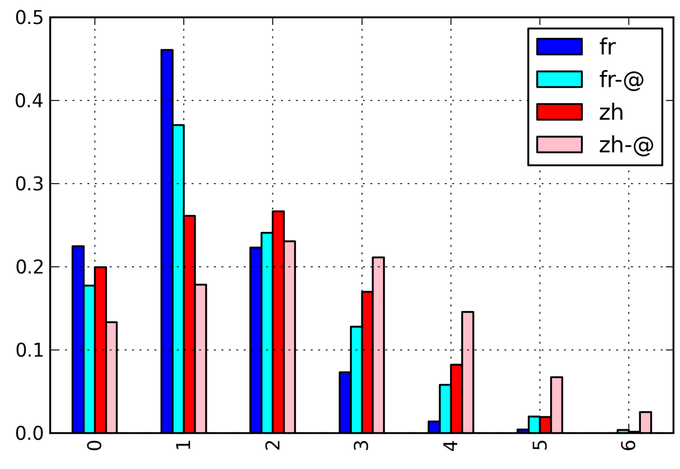
- Usually difficult to record clean speech data in 'one-location' multilogues
- Here a circle of people (French), 2 rows of 4 people facing each other (Mandarin)

Basic info about a game

- So far 4 games for each language, 7-27 minute per game

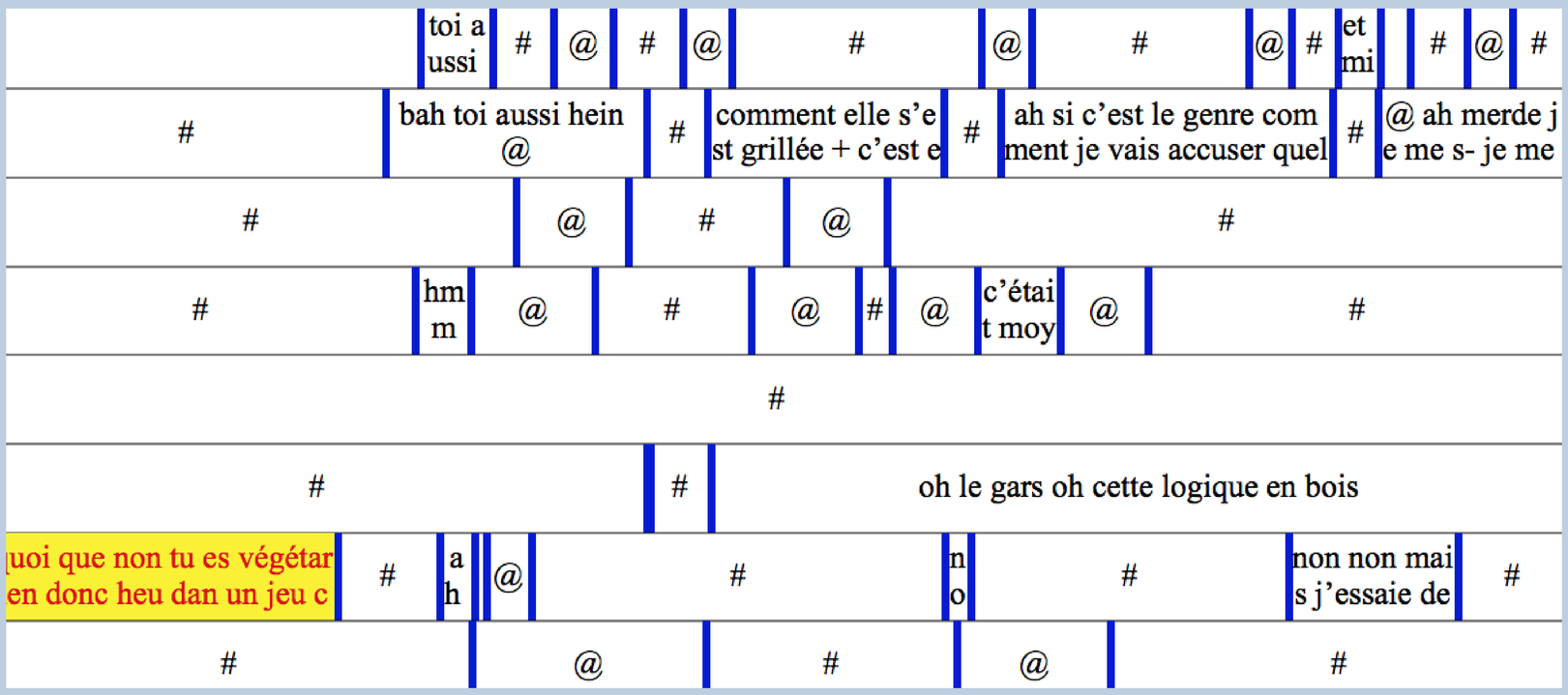


Actual Speaking Duration (s)



# of simultaneous speakers (%),1Hz)

Illustration



Acknowledgements

This work has been carried out thanks to the support of the A*MIDEX Variamu 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.