

Toward a scary comparative corpus: The Werewolf Spoken Corpus

Laurent Prévot

Aix-Marseille Université, CNRS
LPL, UMR 7309, Aix-en-Provence, France
laurent.prevot@univ-amu.fr

Yao Yao

The Hong Kong Polytechnic University
Chinese and Bilingual Studies
Hung Hom, Hong Kong
ctyaoyao@polyu.edu.hk

Arnaud Gingold and Bernard Bel

Aix-Marseille Université, CNRS
LPL, UMR 7309, Aix-en-Provence, France
firstname.lastname@lpl-aix.fr

Kam Yiu Joe Chan

The Hong Kong Polytechnic University
Chinese and Bilingual Studies
Hung Hom, Hong Kong
joe.ky.chan@polyu.edu.hk

1 Introduction

Despite the large number of corpora—both written and spoken—that are currently available, there is still a lack of corpora that document dialogues involving more than two interlocutors. To our knowledge, there is no multilogue corpus recorded in different languages in order to perform comparative studies. This abstract describes the first steps in building such a multilogue comparable corpus for French and Mandarin languages.

Generally speaking, the development of multi-person conversational corpora is hindered by two major obstacles. First, it is relatively hard to elicit natural, multi-person conversations in a laboratory setting. While a two-person dialogue may be easily convened by an experimenter (often as an interviewer who proposes topics for discussion; e.g. the Buckeye corpus, it is much more difficult if not impossible to conduct a truly engaging discussion with a group of invited subjects who may or may not know each other very well. For this reason, existing multi-person conversation corpora mostly use recordings from naturally-occurring group meetings (e.g. research group meetings, business conferences, etc.; see ISL, ICSI (Janin et al., 2003) and AMI (Carletta et al., 2006) meeting corpora), in which case the genre of speech is limited to professional conversations that happen in work places.

The second challenge is posed by the technical difficulty of recording multi-person conversations. For one thing, group meetings are usually recorded in their natural environment, i.e. regular conference rooms, which are often not sound-proof, and the speech data are collected by a few microphones placed in different spots of the room. As a result, the recordings may contain ambient noise, and it is hard to separate different talkers'

speech, especially when two or more talkers speak at the same time. In addition, even if it is possible to record in a sound-treated room with multiple recording devices one set for each talker ideally, there must also be some master control mechanism that ensures that the separate tracks of recordings can be aligned properly later, but such mechanism is rarely provided in regular conference rooms.

Given the above problems, we propose to create a new spoken corpus that uses a party game to elicit multi-person conversations in a non-professional, social setting.

Moreover, one of our objective is to perform comparative quantitative analysis of this kind of data.

2 Elicitation protocol

The game we use is the Werewolf game (also known as the *Mafia* game), which is often played by a group of people (typically more than 5) at parties. Participants of the game are randomly assigned to different roles (e.g. the “murderer”, the “judge”, the “innocent”) by drawing from a deck of cards. Apart from the judge, participants playing other roles are supposed to keep their identities only to themselves. In each round, the undercover “murderer” can “kill” one innocent player, and those who are still alive (including the “murderer”) will vote who they think is the “murderer”. Thus, the innocent players task is to figure out the identity of the “murderer”, while the “murderer” will try to hide their identity and direct suspicion to other players. The game reaches an end either when the “innocent” players guess correctly who is the “murderer”, or when the “murderer” successfully “kills” all the “innocent” players without being caught.

The Werewolf game requires very little instrument but encourages verbal communication as

participants need to exchange information and opinions in order to achieve their goals. Thus, the game is often played as an ice-breaker among new friends or a pastime among familiar friends, which makes it an ideal game for eliciting natural and informal multi-person conversations among familiar or unfamiliar participants in a laboratory setting. The nature of the game involving deception and persuasion also makes it a perfect venue for studying these phenomena in conversations.

All the participants are sitting on chairs forming circle. Such a setting has the advantage of naturalness but despite usage of headset-like microphone, we still face a relatively high level of spill. The game master was recorded only for the Mandarin dataset¹.

3 Basic facts about one Werewolf game

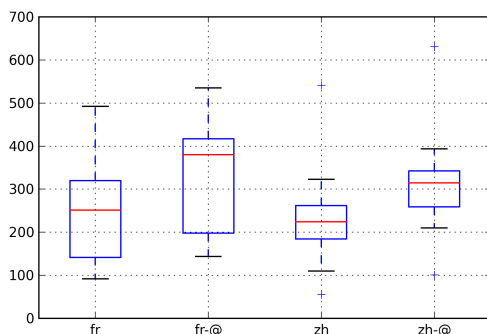


Figure 1: Actual Speaking Duration (s)

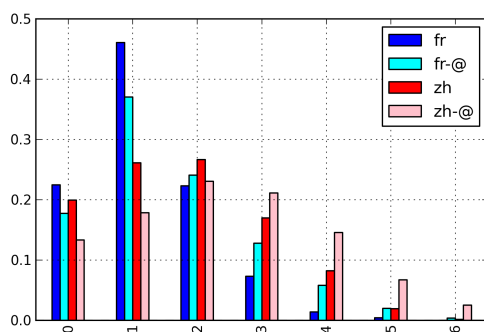


Figure 2: Number of speaker simultaneously speaking (%), sampling rate = 1Hz

For the time being we ran 4 games for each language. One game of each language has been fully

¹The French set-up limited the number of recorded participants to 8.

transcribed. Our games lasted between 7 to 27 minutes (for an actual speech duration of 33 minutes and 3346 tokens in the later case). This important variation is partly due to games in which 'werewolves' are identified early in the game. We checked some basic properties of the sessions recorded in terms of participants involvement and simultaneously speaking. Figure 1 shows the figures for duration of speech for different speaker (excluding the game master) in the longest French and Mandarin games, including or not laughter (-@ in the figures) as verbal behavior. As expected this value is subject to strong inter-individual variations. Figure 2 provides statistics about the number of people simultaneously speaking (excluding the game master). We observe that overlapping speech amount for 35% to 65% of the data which already make it an original data set to work with.

4 Data management and curation

The WEREWOLF pilot corpus and its annotations were stored and described so as to allow a wide dissemination (permanent identifiers : <http://hdl.handle.net/11041/ortolang-000900> and <http://hdl.handle.net/11041/ortolang-000908>). Participants signed a consent form and agreed that their interactions could be publicly disseminated. The work of curation made it possible to store data in sustainable formats and facilitate reuse.

Acknowledgments

This work has been carried out thanks to the support of the A*MIDEX project (ANR-11-IDEX-0001-02) French Government program. We would like to thank all the students that participated in the recordings and in the transcription.

References

- Jean Carletta, Simone Ashby, Sebastien Bourban, Mike Flynn, Mael Guillemot, Thomas Hain, Jaroslav Kadlec, Vasilis Karaiskos, Wessel Kraaij, Melissa Kronenthal, et al. 2006. The ami meeting corpus: A pre-announcement. In *Machine learning for multimodal interaction*, pages 28–39. Springer.
- Adam Janin, Don Baron, Jane Edwards, Dan Ellis, David Gelbart, Nelson Morgan, Barbara Peskin, Thilo Pfau, Elizabeth Shriberg, Andreas Stolcke, et al. 2003. 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.