Constructing a Chinese Backchannel Corpus: Combining Morphological Analysis and Generative AI

Yusuke Mochizuki, Eiko Yamamoto
Gifu Shotoku Gakuen University / Matsuyama University, Gifu Shotoku Gakuen University
mochizukiyusuke0409@gmail.com, eiko@gifu.shotoku.ac.jp

Abstract

This study explores a method for constructing a corpus of naturally occurring conversations among native speakers of Chinese by integrating natural language processing (NLP) with generative AI. The analysis of natural conversation requires not only the transcription of speakers' utterances but also the systematic annotation and retrieval of interactional phenomena produced by listeners, such as backchannels, fillers, overlapping speech, and interruptions. However, the Chinese word segmentation and part-of-speech tagging tool Jieba does not provide a dedicated tag for backchannels. Expressions such as *"嗯" (ng)* and *"哦" (o)*, for example, may be assigned tags associated with states or interjections, making it difficult to identify backchannels automatically on the basis of part-of-speech tags alone. To address this limitation, this study proposes a hybrid extraction model that combines candidate identification through morphological analysis, context-sensitive classification using generative AI, and manual verification. It also examines cases in which the same linguistic form may serve multiple discourse functions, including backchannels, responses, and fillers, and discusses both the advantages and limitations of using generative AI for the construction and annotation of natural conversation corpora.

Keywords: naturally occurring Chinese conversation, backchannels, *Jieba*

1. Introduction

Spoken interaction is not a one-way activity in which a speaker simply conveys information to a listener. Rather, it is a jointly constructed process in which speakers and listeners continuously respond to and interpret one another's actions as the interaction unfolds. To capture the nature of naturally occurring conversation, it is therefore necessary to examine not only speakers' primary utterances but also how and when listeners display understanding, agreement, acceptance, surprise, and other responses. Backchannels, in particular, constitute an important form of listener behavior through which recipients demonstrate their ongoing participation in the interaction, and they are therefore indispensable for understanding the organization of conversation. A substantial body of research has examined backchannels from a variety of perspectives, including their forms, functions, distribution, and interactional roles (e.g., Clancy et al.,1996).

At the same time, although many existing large-scale Chinese corpora provide powerful facilities for searching textual information, they do not necessarily allow researchers to directly retrieve the phenomena characteristic of naturally occurring interactions such as backchannels, fillers, overlapping speech, interruptions, and silence—together with their interactional functions. For example, forms such as 嗯, 啊, 哦, and 对 can readily be searched as character strings, but not every occurrence functions as a backchannel. The same form may signal recipient acceptance, occur in a repair initiation or a response to a question, or function as a filler at the beginning of a turn. Identifying backchannels therefore requires analysis that goes beyond surface form alone and takes the surrounding interactional context into account.

Moreover, even when generative AI systems are instructed to produce natural-sounding dyadic conversations in Chinese, backchannels do not necessarily occur with the frequency or in the sequential positions observed in naturally occurring human interaction. In preliminary trials using ChatGPT and Google Gemini, backchannels appeared infrequently in the generated conversations, and occurrences of "嗯"(ng) remained limited even when the systems were explicitly instructed to include backchannels. This suggests that generative AI cannot yet be straightforwardly used as a substitute for naturally occurring conversational data. Rather, it should be employed as an analytical aid that is grounded in authentic interactional data.

Against this background, the present study examines a method for identifying backchannels in naturally occurring first-encounter conversations among native speakers of Chinese collected by the author, combining morphological analysis with generative AI. The study has three main objectives: first, to clarify the limitations of automated processing based on morphological analysis; second, to investigate the extent to which generative AI can identify potential backchannels by taking contextual information into account; and third, to propose a reproducible workflow for corpus construction that incorporates manual verification.

2. Data and Analytical Focus

The data analyzed in this study consist of 690 minutes of naturally occurring first-encounter conversations between native speakers of Chinese, collected by the author in Beijing. First-encounter conversations are particularly suitable for the present analysis because the participants have relatively little shared knowledge, making interactional activities such as topic initiation, displays of understanding, agreement, assessment, and affiliation relatively explicit. Moreover, the roles of speaker and listener are not fixed: a form used as a listener response at one moment may subsequently serve as a resource for initiating the same participant's own turn. Such data therefore provides an appropriate basis for examining

functional distinctions among backchannels that cannot readily be captured through simple lexical searches.

The main forms examined in this study include *“嗯”(ng)*, *“啊”(a)*, *“哦”(o)*, *“是”(shi)*and *“对”(dui)*. However, backchannels are not defined here in terms of particular lexical forms. Rather, they are provisionally defined as brief listener responses to a preceding utterance that displays understanding, acceptance, agreement, or encouragement for the current speaker to continue. Accordingly, even instances of the same form, such as 嗯, are treated separately when they do not respond to the immediately preceding utterance but instead function as resources for initiating the speaker's own turn, or when they convey substantive propositional content comparable to "yes" or "no." Such instances are distinguished from backchannels in the present analysis.

2.1 Preliminary Morphological Analysis

When transcripts of naturally occurring conversations are processed using the Chinese word segmentation and part-of-speech tagging tool *Jieba*, forms such as *“嗯”(ng)* and *“哦”(o)* are not necessarily assigned tags that reflect their conversational functions. Examples are provided below.

<example1>

[When morphological analyses were performed, the backchannels are enclosed in parentheses to explicitly indicate their positions.]

（1）我叫【姓名】，是那个国际关系学院大二的学生（哦）[1]。

(My name is [Full Name], and I'm a second-year student at the School of International Relations.)

Results of Morphological Analysis:

我/r 叫/v 【姓名】/nr，/x 是/v 那个/r，/x 国际/n 关系/n 学院/n 大二/m 的/uj 学生/n （/x 哦/zg /x ）/x 。/x

<example2>

[When morphological analyses were performed with backchannels transcribed in their original positions.]

Results of Morphological Analysis:

我/r 叫/v 【姓名】/n，/x 是/v 那个/r ，/x 国际关系学院/nt 大二/m 的/uj 学生/n 哦/zg 。/x

In this example, *“哦”(o)* is assigned the tag "zg," which is associated with states. Forms such as *“嗯”(ng)* and *“哦”(o)* may also be assigned the tag "e," which denotes interjections. However, neither "zg" nor "e" is a dedicated tag for backchannels. Therefore, searching solely on the basis of these tags does not allow backchannels to be extracted selectively. Conversely, searching for character strings such as *“嗯”(ng)* and *“哦”(o)* can retrieve a broader range of potential backchannels, but it also captures instances functioning as fillers, turn-initial elements, or repair initiators.

2.2 Limitations of Morphological Analysis

What is important here is not to point out errors in morphological analysis itself. General-purpose morphological analyzers are designed primarily to segment words and assign part-of-speech information; they are not intended to directly classify interactional functions. The issue, therefore, is not so much that "Jieba misidentifies backchannels," but rather that morphological information alone is insufficient to determine whether a given form functions as a backchannel in discourse.

For example, *“对”(dui)* is highly likely to be treated as a backchannel when it occurs independently as an expression of agreement with the preceding speaker. When it carries more substantive semantic content, however, it needs to be analyzed as serving a different function. Similarly, *“嗯”(ng)* may occur as a brief response with falling intonation, as a repair initiator with rising intonation, or as a filler used to hold the floor while the speaker formulates an upcoming utterance. Identifying backchannels therefore requires reference to at least the surrounding utterances, the sequential position of the form, speaker change, and whether or not a subsequent utterance follows.

3. A Generative AI–Assisted Approach

One of the main advantages of generative AI over morphological analysis is its ability to take into account the broader linguistic context surrounding a given expression and to assess the discourse function that the expression performs within that context. For example, rather than presenting the system with an isolated instance of *嗯”(ng)*, the preceding and following turns, together with the relevant conversational context, can be provided as input. This makes it possible to distinguish between *嗯”(ng)* functioning as a backchannel and *嗯”(ng)* used with more substantive semantic content.

Generative AI does not necessarily produce exactly the same judgment or explanation each time it is given identical input. Moreover, unless explicit classification criteria are provided, it may overinterpret an expression on the basis of its general semantic meaning or assumptions about what would constitute a natural conversational exchange. Accordingly, in the present study, generative AI is not treated as the final arbiter of classification. Rather, it is employed as an auxiliary tool for context-sensitive classification, serving as an intermediate step between morphological analysis and manual analysis.

4. A Hybrid Model for Backchannel Extraction

This study proposes a five-stage model consisting of: (1) candidate extraction through string matching and

[1] Backchannels are indicated by ( ).

morphological analysis, (2) filtering and organization based on formal criteria, (3) context-sensitive classification using generative AI, (4) final manual verification, and (5) annotation of the resulting data in the corpus. The aim of this approach is not to eliminate human intervention entirely, but rather to efficiently narrow down the cases requiring manual inspection while systematically accumulating explicit criteria for classification.

Procedure	Main Function
(1) String Search and Jieba Analysis	Broadly extract candidate instances containing forms such as *“嗯” (ng)*, *“啊” (a)*, *“哦” (o)*, *“是” (shi)*and *“对” (dui)*
(2) filtering and organization based on formal criteria	Annotate information such as utterance position, standalone use, speaker change, and the presence of a following utterance.
(3) context-sensitive classification using generative AI	Provisionally classify candidates as backchannels, responses, fillers, etc. by referring to the surrounding context.
(4) final manual verification	Researchers make final judgments, focusing particularly on borderline cases and misclassified instances.
(5) annotation of the resulting data in the corpus	Store information on form, function, position, speaker, and related features in a searchable format

Table 1: A Hybrid Model for Backchannel Extraction

5. Methodological Implications

When using generative AI for classification, it is preferable not to rely solely on open-ended responses, but rather to specify the classification categories and the criteria for judgment in advance. For example, the present study provisionally established the following categories: “A: backchannel,” “B: response to a question or other eliciting action,” “C: repair initiation,” and “D: filler or turn initiation.” The generative AI system was then asked to determine which category each target expression belonged to and to identify the preceding and following utterances that provided the basis for its judgment. By standardizing the output format in this way, subsequent quantitative analysis and manual verification were made more efficient.

The proposed approach offers several potential advantages. First, it makes it possible to narrow down, from a large volume of naturally occurring conversational data, the candidate instances that require closer examination by the researcher. While string searches and morphological analysis are fast and highly reproducible, they are less well suited to identifying discourse functions. Generative AI can serve as an intermediate stage in the analytical process by taking contextual information into account and provisionally classifying candidate instances.

Second, generative AI can be prompted to articulate the reasoning underlying its classifications. For example, it may provide an explanation such as: “This instance is classified as a backchannel because it responds to the immediately preceding utterance and is not followed by a substantive utterance by the same speaker.” Such explanations make it easier for researchers to assess the validity of individual classifications. Moreover, when the same decision criteria recur across multiple instances, these criteria can be explicitly formulated and incorporated into the annotation guidelines.

6. Potential Applications to Linguistic Research

Once functional annotations for backchannels have been systematically incorporated into the corpus, it becomes possible to analyze not only their frequency of occurrence but also their relationship to factors such as position within the utterance, turn transition, the type of preceding utterance, speaker characteristics, and the stage of conversational development. For example, one could quantitatively examine differences in the distribution of 嗯, 哦, and 对, functional differences between turn-internal and turn-boundary positions, and the relationship between listener responses and topic development, and then conduct a more detailed interactional analysis of individual cases.
In this respect, the significance of the present study lies less in using generative AI to artificially produce conversations than in enhancing the analytical potential of naturally occurring conversational data. Phenomena found in natural conversation may not be adequately represented in edited written texts or in prototypical conversations generated by AI. By taking authentic conversational data as the empirical basis and combining formal processing through natural language processing, context-sensitive analysis using generative AI, and human analysis of interaction, it is possible to establish a framework for investigating forms of linguistic behavior that have been difficult to retrieve and analyze in conventional text-centered corpora.

7. Conclusion

This study examined a model for constructing a corpus of naturally occurring Chinese conversation by combining morphological analysis with generative AI, with a particular focus on backchannels. Morphological analysis tools such as *Jieba* are useful for extracting candidate forms; however, part-of-speech tags alone cannot determine whether forms such as *“嗯” (ng)* and *“哦” (o)* are functioning as backchannels in a given conversational context. Generative AI, by contrast, has the potential to provide provisional classifications of discourse functions by taking the surrounding context into account. Nevertheless, AI-generated classifications should not be treated as final judgments. Rather, it is essential to combine string-based searches and morphological analysis, context-

sensitive classification using generative AI, and manual verification.

Future research will extract a substantial number of candidate instances from naturally occurring conversational data and examine the agreement between human and generative AI classifications, as well as patterns of misclassification. It will also investigate differences in classification accuracy between conditions in which audio information is available and those in which only textual information is provided, differences across multiple generative AI models, and the effects of prompt design on classification outcomes. Ultimately, the proposed approach will be extended beyond backchannels to the interactional annotation of other phenomena essential to naturally occurring conversation, including fillers, overlapping speech, and interruptions. The broader goal is to establish a corpus-construction methodology that can support systematic research on naturally occurring Chinese conversation.

8. Bibliographical References

Clancy,Patricia M., Sandra A. Thompson, Ryoko Suzuki, and Hongyin Tao. (1996). The conversational use of reactive tokens in English, Japanese, and Mandarin. Journal of Pragmatics, 26:355–387.