

Corpus-Assisted Learning in Intermediate-Advanced Chinese Contextual Conversations

Hsin-Jou Yeh

Department of Chinese Studies, Universität Tübingen



Zusammenfassung

This project integrates a corpus-based teaching module into an advanced Chinese speaking course (“Sprachverwendung in sozialen Kontexten”) at the University of Tübingen. The target learners are 5 B.A. and 3 M.A./M.Ed. students of Sinology at the CEFR B2 level.

The project addresses the limitations of textbook-based learning and students’ reliance on translations or AI tools by introducing corpus tools (Sketch Engine) into the classroom. Students learn to analyze authentic language use, verify their own expressions, and apply these skills in a final oral presentation.

The results show that corpus-based learning enhances students’ awareness of language use and supports learner autonomy.

At the same time, challenges such as the complexity of authentic data highlight the need for structured guidance and balanced integration.

Herausforderung

The project addresses four main challenges in current Chinese language teaching.

First, textbook examples and contexts are often limited and constructed based on the authors’ intuition, which may not accurately reflect authentic language use.

Second, vocabulary explanations frequently rely on English translations, which can be misleading. Learners may overlook subtle differences between Chinese and English usage or assume that words sharing the same English translation can be used interchangeably.

Third, students often lack the ability to critically evaluate their own language use. With the increasing availability of translation tools and AI-based applications, they tend to rely heavily on technology, which can result in unnatural expressions.

Finally, error correction in the classroom is typically reactive and teacher-driven. While teachers provide immediate feedback, such corrections are often not retained. In contrast, developing learners’ ability to actively and consciously verify and adjust their own language use is crucial for long-term improvement.

Zielsetzung

This project aims to enhance the pragmatic competence of Sinology students through Corpus-Assisted Language Teaching (CALT) by achieving the following objectives.

1. Providing students with authentic language input and diverse linguistic exposure:
Enabling students to access authentic language samples from various fields to strengthen their understanding of authentic contexts, as well as reducing dependence on rigid textbook expressions and fostering more natural and flexible language use
2. Developing students’ ability to recognize and apply language in different contexts:
Helping students master essential linguistic features such as collocations, formulaic expressions, and syntactic structures to improve fluency, as well as encouraging students to analyze corpus data and adjust their expressions to suit different real-world contexts
3. Improving students’ language output to resemble native-like expressions:
Using corpus-based analysis and application exercises to develop students’ ability to produce natural language expressions and reduce direct translation-based errors, as well as enhancing students’ adaptability in both formal and informal communication settings
4. Promoting student collaboration and cooperative learning:
Encouraging students to work in groups to analyze and interpret corpus data through discussions and research projects, as well as implementing peer-to-peer teaching strategies where students share insights on corpus-based learning to enhance both analytical and communicative skills
5. Cultivating students’ research abilities and fostering inquiry-based learning:
Guiding students to explore corpus data, identify patterns, and formulate their own linguistic rules, as well as introducing students to digital humanities tools and data-driven linguistic research methodologies

Lösungsansatz

The course is structured in two main components. During the first 60–75 minutes, students engage in speaking practice based on *New Era Spoken Chinese: Intermediate Level*. The remaining 15–30 minutes are dedicated to guided corpus-based activities using Sketch Engine.

Students are introduced to key functions such as Word Sketch (to explore collocations of a given word), Word Sketch Difference (to compare near-synonyms), and Concordance (to examine large numbers of authentic examples in context).

The instructor uses vocabulary from the textbook as a starting point, guiding students to verify whether textbook collocations reflect common usage in real language data. In addition, learners’ own language errors produced during class are used as prompts for corpus exploration, enabling students to observe differences between similar expressions and refine their usage.

Furthermore, students are required to use corpus tools in preparing their final oral presentation. In addition to presenting on a chosen topic, they must demonstrate how they have used the corpus in their learning process. Two weeks before the final presentation, individual consultations are held, during which students can discuss their progress and address any difficulties encountered when working with corpus tools.

Innovationscharakter

The innovative aspect of this project lies in four dimensions.

First, it provides students with access to a large amount of systematically-collected authentic language data, going beyond the limitations of textbook-based input. While traditional teaching materials often rely on constructed examples, the use of corpus data allows learners to observe real language use and identify frequent patterns and collocations in context.

Second, the project redefines the roles within the classroom. Instead of passively receiving knowledge, students are positioned as active researchers, while the instructor assumes a facilitative role. Learners are guided to formulate questions, explore corpus data, and interpret linguistic patterns, thereby engaging more deeply with the learning process.

Third, the approach introduces an evidence-based learning strategy. Students are encouraged to verify their linguistic intuitions and the results obtained from dictionaries, translation tools, or AI applications through corpus data. This process not only improves the accuracy of language production, but also fosters critical thinking and a more reflective approach to language use.

Furthermore, corpus-based learning creates opportunities for sustained and autonomous learning. Students are able to continue using corpus tools beyond the classroom, applying them to new contexts and even to other languages. In this way, the project supports the transferability of learning strategies and promotes long-term learner development.

In this sense, the project fosters not only linguistic competence, but also methodological awareness and learner autonomy, which are essential for advanced language learning.

Erfahrungen

The project has led to three main positive outcomes.

First, students developed the ability to independently verify their own language use, which enhanced their awareness of collocational patterns and improved the accuracy of their expressions.

Second, the use of corpus tools exposed learners to a wider range of authentic vocabulary and usage patterns, supporting vocabulary expansion beyond textbook input.

Third, the project fostered learner autonomy. Students not only applied corpus tools in their final presentations, but also extended their use beyond the classroom, for example by using Sketch Engine for English learning or exploring additional functions such as the thesaurus feature.

At the same time, a central challenge lies in students’ limited vocabulary when working with authentic corpus data. Even at the B2 level, learners frequently encountered unfamiliar words, which sometimes made interpretation difficult and required additional support. This challenge is further reinforced by a limitation of the tool itself: Sketch Engine does not provide integrated translation support. As a result, students need to rely on external translation tools, suggesting that corpus-based learning is most effective when combined with other resources.

Building on this limitation, future research could explore the development of a Chinese–German parallel corpus or a pedagogically oriented corpus. Such resources may better support learners in linking meaning and usage, while also making corpus-based learning more accessible and suitable for instructional purposes.

Studentische Perspektive

Student feedback indicates a generally positive perception of corpus-based tools, particularly in supporting the understanding of authentic language use and vocabulary differentiation.

Several students highlighted the usefulness of example sentences and the ability to compare similar words. As one student noted:
“Providing example sentences for words [...] Sketch Engine can show when to use which word, if you have two words with similar meaning [...] Also finding similar words can be useful.” (S1)

In addition, students appreciated the visualization features of Sketch Engine, which transform tabulated data into graphical representations and make relationships between target words and their collocations more transparent. As one student commented:
“I really liked the visualization aspect.” (S2)

At the same time, feedback revealed certain limitations. One student pointed out that while Sketch Engine is helpful for understanding how a word is used, it presupposes prior knowledge of word meaning. As a result, learners tend to combine corpus tools with translation applications:
“Sketch Engine can help us better understand how to use a word, but we need to know its meaning first, so I also use translation tools.” (S3)

Regarding the level of instruction, more advanced students suggested that corpus-based learning could already be introduced at earlier stages, such as A2 or B1.

In terms of implementation, students expressed a preference for a moderate integration of corpus-based activities. Most suggested that using such tools once or twice per month would be sufficient, rather than on a weekly basis.

Finally, students identified a need for more structured guidance when learning how to use corpus tools. As one student suggested:
“A quick crash course on every function at the beginning of the course and exercises to make use of specific functions to deepen the understanding.” (S2)
Similarly, another student remarked:
“I would suggest to spend more time introducing how to use Sketch Engine with examples.” (S3)