

Hung-Hsin Hsu

A corpus-based contrastive study of questions in Mandarin and French

Keywords: question types; question structures; comparable corpus of fiction; Taiwan Mandarin; French

The topic of questions has received relatively little attention in contrastive studies to date (Coveney 2011; Curry/Chambers 2017), especially in pairs involving languages other than English. This paper aims to fill this gap by examining questions in Mandarin and French, specifically regarding question types (i.e. polar questions, disjunctive questions, and *wh*-questions; see examples below extracted from my own corpus) and their respective question structures (e.g. declarative forms, subject-verb inversion, question particles *est-ce que* in French or *ma* and *ne* in Mandarin, A-not-A constructions, fragments.)

(1) Mandarin

- a. *Nǐ zài gǎoxiào ma?*
 2SG PROG joke Q
 ‘Are you joking?’
- b. *Jiàoshǐ shì nán de nǚ de?*
 driver COP man DE woman DE
 ‘The driver is a man or a woman?’
- c. *Shénme rén huì chóushì lǎorén?*
 what people AUX hate elderly
 ‘What kind of people hate the elderly?’

(2) French

- a. *Serons-nous en contact?*
 COP.FUT-1PL PREP contact
 ‘Will we be in contact?’
- b. *Était-il en migration lui aussi ou en fuite?*
 COP.PST-3SG PREP move 3SG also or PREP run
 ‘Was he on the move too, or on the run?’
- c. *Qu’y a-t-il là-dedans?*
 what LOC have.PRS-t-3SG in-there
 ‘What is in there?’

Previous findings of a pilot study using Mandarin and French web data from the zhTenTen 17 and frTenTen 17 corpora (Jakubíček et al. 2013) show that in the two languages polar questions are more frequent than the other question types, which is in line with the theory of social economics of questions (Levinson 2012) and also with the results of studies devoted to English (Stivers 2010; Siemund 2017). Additionally, interesting cross-linguistic differences emerge from the web corpora, showing that *wh*-questions are significantly more frequent in French than in Mandarin, possibly due to the high number of different structures available

in French to express *wh*-questions (i.e. inversion, declarative forms, *est-ce que* particle, *wh*-in-situ, and fragments). Admittedly, these different structures, with the exception of *wh*-in-situ, can also be used in polar and disjunctive questions. However, *wh*-questions have few more possible variations observed, such as cleft sentences (Guryev 2017; Larrivée/Guryev 2021). These various options provide language users with a wider range of choices when producing *wh*-questions. However, these findings remain tentative given the diversity of registers included in web corpora, which may potentially jeopardize the cross-linguistic comparability of the data. Consequently, to get a better view of cross-linguistic differences and confirm the observations of the pilot study, I build a 1-million-word self-compiled bilingual comparable corpus that comprises contemporary fictional texts, including detective and fantasy stories. The corpus was uploaded on Sketch Engine (Kilgarriff et al. 2014) for automatic extraction of questions in the two languages using the CQL query [word = “\?”] (see Biber et al. (1999) and Axelsson (2020) for a similar approach). The query returns over 3,000 occurrences in each language, as shown in Table 1. The questions are then coded for question type and syntactic structure. Based on the results of the pilot study of web corpora, two hypotheses are tested in the current study: (1) polar questions are the most frequent question type in the two languages and (2) *wh*-questions are more widespread in French than in Mandarin. In my presentation, I will report on the results of the analysis with a view to sketching the main commonalities and differences between questions in Mandarin and French.

	Mandarin	French
Raw frequencies	3,211	3,573
Relative frequencies per 100,000 tokens	770	649
Corpus size (tokens)	417,049	550,807

Table 1: Question marks in the self-compiled bilingual comparable corpus

References

- Axelsson, Karin (2020): Questions in English and Swedish fiction texts: A study based on parallel and comparable corpus data. In: *Languages in Contrast*, 20(2), pp. 235–262. <http://doi.org/10.1075/lic.00017.axe>.
- Biber, Douglas/Johansson, Stig/Leech, Geoffrey/Conrad, Susan/Finegan, Edward (1999): *Longman Grammar of Spoken and Written English*. Longman.
- Coveney, Aidan (2011): L’interrogation directe. In: *Travaux de linguistique*, 63(2), pp. 112–145. <http://doi.org/10.3917/tl.063.0112>.
- Curry, Niall/Chambers, Angela (2017): Questions in English and French Research Articles in Linguistics: A Corpus-Based Contrastive Analysis. In: *Corpus Pragmatics*, 1(4), pp. 327–350. <http://doi.org/10.1007/s41701-017-0012-0>.

ICLC-10 2023 - A corpus-based contrastive study of questions in Mandarin and French

Guryev, A. (2017): La forme des interrogatives dans le corpus suisse de SMS en français: étude multidimensionnelle. Université de la Sorbonne Nouvelle (France) & Université de Neuchâtel (Suisse). <https://www.theses.fr/2017USPCA045>.

Jakubiček, Miloš/Kilgarriff, Adam/Kovář, Vojtěch/Rychlý, Pavel/Suchomel, Vít (2013): The TenTen Corpus Family. In: 7th International Corpus Linguistics Conference. *7th International Corpus Linguistics Conference*, pp. 125–127.

Kilgarriff, Adam/Baisa, Vít/Bušta, Jan/Jakubiček, Miloš/Kovář, Vojtěch/Michelfeit, Jan/Rychlý, Pavel/Suchomel, Vít (2014): The Sketch Engine: ten years on. In: *Lexicography*, 1(1), pp. 7–36. <http://doi.org/10.1007/s40607-014-0009-9>.

Larrivé, Pierre/Guryev, Alexander (2021): Variantes formelles de l'interrogation. Présentation. In: *Langue française*, 212(4), pp. 9–24. <http://doi.org/10.3917/lf.212.0009>.

Levinson, Stephen C. (2012): Interrogative intimations: On a possible social economics of interrogatives. In: de Ruiter, Jan P. (ed.): *Questions: Formal, Functional and Interactional Perspectives*. Cambridge: Cambridge University Press (Language Culture and Cognition), pp. 11–32. <http://doi.org/10.1017/CBO9781139045414.003>.

Siemund, Peter (2017): Interrogative clauses in English and the social economics of questions. In: *Journal of Pragmatics*, 119, pp. 15–32. <http://doi.org/10.1016/j.pragma.2017.07.010>.

Stivers, Tanya (2010): An overview of the question–response system in American English conversation. In: *Journal of Pragmatics*, 42(10), pp. 2772–2781. <http://doi.org/10.1016/j.pragma.2010.04.011>.

Contact information

Hung-Hsin Hsu

Center for English Corpus Linguistics (CECL)/

Institut Langage et Communication (ILC)

Université catholique de Louvain (UCL), Belgium

hung-hsin.hsu@uclouvain.be

Graduate Institute of Linguistics

National Chengchi University (NCCU), Taiwan

108555501@g.nccu.edu.tw