

How subjectivity is expressed in discourse: Evidence from collocation analyses

Yipu Wei¹; Jacqueline Evers-Vermeul²; Ted Sanders²

¹ Peking University; ² Utrecht University

Background & Research questions To communicate in a coherent way, speakers choose words to express the relations between consecutive discourse segments (Sanders et al. 1993: 94, Schilperoord & Verhagen 1998). For instance, connectives such as *because* and *so* are used to provide information on the type of coherence relation (e.g., temporal, adversative, or causal). In several languages, connectives also code information on subjectivity, i.e., the involvement of a locutionary agent (Finegan, 1995). For example, the Dutch connectives *want* ‘because’ and *dus* ‘so’ (Spooren, et al., 2010) and Mandarin Chinese *kejian* ‘so’ (Li, et al., 2013) specifically express subjective relations. The former type of connectives (such as English *so* and *because*) are underspecified in terms of subjectivity, and the latter type are specific subjective connectives. In expressing discourse relations, people also need to continuously choose other linguistic elements to express their opinions and feelings in these relations, such as who is responsible for the reasoning in the causal relation and what the speaker’s attitude is. When making those choices, speakers and writers decide how informative they should be in order to provide sufficient cues for others to comprehend them. At the same time, they should also avoid being too wordy (see *hearer economy* and *speaker economy*, Horn, 1984).

Language users do not only rely on connectives to express subjectivity in coherence relations. Perspective markers such as *John thinks*, *according to Peter* and *perhaps* also provide information on the involvement of a locutionary agent. On-line processing studies suggest that the processing effects of connectives are interfered by the presence of these perspective markers (Canestrelli, Mak, & Sanders, 2013). However, stance marking is not restricted to one specific type of stance – the epistemic stance in the case of the processing study, indicating that someone’s mind is involved in the construction of the coherence relation; two other dimensions can be distinguished: attitudinal stance and style stance (Conrad & Biber, 2000). This paper aims to extend the understanding of the relation between stance markers and connectives encoding different degrees of subjectivity.

To achieve this explorative goal, we perform collocation analysis on natural language data, which can advance our knowledge on the properties of a connective on the basis of its contextual features. Specific questions are raised: Do connectives marking different subjectivity degrees (specific subjective connective vs. underspecified connective) differ in the types of collocates? Specifically, do connectives differ in the types of perspective markers they co-occur with? Do the collocation patterns differ across various contexts and genres?

Method We focused on two Chinese causal connectives: the specific subjective *kejian* ‘so’, and a connective that is underspecified in terms of subjectivity and can be used in both objective and subjective relations: *suoyi* ‘so’ (Li et al., 2013). A distinctive-collocate analysis was performed by measuring the association strength between these connectives and other discourse elements. We retrieved data from the CCL corpus, a large, balanced Modern Chinese written corpus. Association scores (G^2 and Delta-P) were calculated based on contingency tables of observed and expected frequencies (Evert, 2008; Gries, 2013). Top 100

collocates ranked by G^2 were considered significant collocates of either of the connectives. The Delta-P value was used as a secondary criterion: words in attraction to *suoyi* are above the threshold of 0, and words in repulsion to *suoyi* (as collocates of *kejian*) are below this threshold (<0). The collocation patterns were investigated and compared in different contexts (global vs. local; preceding vs. following) and genres (narrative vs. non-narrative).

Results & Conclusion The underspecified connective *suoyi* prefers contexts with perspective markers expressing epistemic stance: cognition verbs (*think, know*), communication verbs (*say, tell*) and modal verbs (*should, may*). *Kejian* co-occurred more often with perspective markers related to attitudinal stances, such as markers of expectedness (*surprisingly, unexpectedly*) and importance (*important*). The collocation patterns were consistent across different contexts and genres with small variations. Given the findings in corpus-based research, we conducted a follow-up reading experiment to investigate how different types of perspective markers affect the on-line processing of subjective relations marked by different connectives. The results from the reading experiment suggests that both types of stance markers can facilitate the processing of subjective relations. No interaction of connective type and the type of stance markers was found in the on-line reading.

Selected references

- Canestrelli, A. R., Mak, W. M., & Sanders, T. J. M. (2013). Causal connectives in discourse processing: How differences in subjectivity are reflected in eye-movements. *Language and Cognitive Processes*, 28(9), 1394–1413.
- Conrad, S., & Biber, D. (2000). Adverbial marking of stance in speech and writing. In S. Hunston & G. Thompson (Eds.), *Evaluation in text: Authorial stance and the construction of discourse* (pp. 56–73). Oxford: Oxford University Press.
- Evert, S. (2008). Corpora and collocations. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics. An international handbook* (pp. 1212–1248). Berlin: Mouton de Gruyter.
- Finegan, E. (1995). Subjectivity and subjectivisation: An introduction. In D. Stein & S. Wright (Eds.), *Subjectivity and subjectivisation: Linguistic perspectives* (pp. 1–15). Cambridge: Cambridge University Press.
- Gries, S. T. (2013). 50-something years of work on collocations: What is or should be next... *International Journal of Corpus Linguistics* 18(1): 137-165.
- Horn, L. R. (1984). Toward a new taxonomy for pragmatic inference. In D. Schiffrin (Ed.), *Meaning, form, and use in context: linguistic applications* (pp. 11–42). Washington, D.C.: Georgetown University Press.
- Li, F., Evers-Vermeul, J., & Sanders, T. J. M. (2013). Subjectivity and result marking in Mandarin: A corpus-based investigation. *Chinese Language and Discourse*, 4(1), 74–119.
- Sanders, T. J. M., Spooren, W. P. M. S., & Noordman, L. G. M. (1993). Coherence relations in a cognitive theory of discourse representation. *Cognitive Linguistics*, 4(2), 93–133.
- Schilperoord, J., & Verhagen, A. (1998). Conceptual dependency and the clausal structure of discourse. In J.-P. Koenig (Ed.), *Discourse and cognition: Bridging the gap* (pp. 141–163). Chicago: University of Chicago Press.
- Spooren, W. P. M. S., Sanders, T. J. M., Huiskes, M., & Degand, L. (2010). Subjectivity and causality: A corpus study of spoken language. In S. Rice & J. Newman (Eds.), *Empirical and experimental methods in cognitive/functional research* (pp. 241–255). Chicago: University of Chicago Press.