# Corpus Approaches to Lexicogrammar 2022

## LxGr2022

8-9 July 2022

## SYMPOSIUM ABSTRACTS

# Not all subjects are "raised" equally:
## The alternation between the constructions with *sembrare* (*seem*) in written and spoken Italian

Flavio Pisciotta

University of Bologna

Traditionally, *sembrare* (*seem*) has gained interest for its peculiar syntactic and semantic behaviour. In particular, scholars have claimed that the raising construction with *sembrare* (a) is truth-conditionally equivalent to the impersonal construction in (b) (Davies & Dubinsky 2004: 4):

> a. Gianni sembra amare Piero.
> 'Gianni seems to love Piero'

> b. Sembra che Gianni ami Piero.
> 'It seems that Gianni loves Piero'

This has led to a transformational analysis of (1a), where the subject *Gianni* is selected by the subordinate verb and then raised to be the syntactic subject of *sembrare*. This perspective implicitly maintains a free variation between the two constructions. Nonetheless, many scholars working in constructionist frameworks have found that a number of formal alternations are motivated by differences in the functional properties of the alternating constructions (e.g., Gries 1999; 2003; Stefanowitsch 2003; Perek 2015; De Vaere et al. 2018). Following a Construction Grammar approach, we hypothesize that *sembrare*+infinitive (1a) and *sembra che* (1b) are allostructions, i.e., formal variants of a more abstract construction that differ with respect to some semantic or pragmatic properties (Cappelle 2006).

To test this hypothesis, we focus on selection properties by analyzing whether any subject is a good candidate to appear in the constructions under exam. 216 occurrences of the two constructions were selected from the written Italian corpus CORIS (Rossini Favretti et al. 2002), and 189 occurrences from two spoken Italian corpora: LIP (De Mauro et al. 1993) and KIParla (Mauri et al. 2019). We tested four variables related to "raised" and "unraised" subjects, namely length, animacy, accessibility, and type of anaphoric link represented by the subject (Nissim et al. 2004). We used the Mann-Whitney test to assess the length of the subjects, and the chi-squared test with Monte Carlo simulation of p-values for each of the other three variables. Written and spoken occurrences were tested separately.

Our results show that there are, in fact, differences in the selection of the subject. In the data from the written corpus, the subjects of *sembrare*+infinitive tend to be longer than the subjects of *sembra che* (p-value = 0.005374). In the data from the spoken corpora, the type of anaphoric link is highly significant (p-value = 9.999e-05): the standardized residuals show that *sembra che* rarely selects a relative pronoun as its subject, while it is very commonly selected by *sembrare*+infinitive. Furthermore, *sembrare*+infinitive never selects a 1st/2nd person pronoun as its subject, unlike *sembra che*. These two restrictions on the anaphoric links proved to be valid also for our written data, even if they did not provide a significant result; this is probably due to their lower frequency in the written corpus.

These results support the hypothesis that *sembrare*+infinitive and *sembra che* are allostructions, although the functional differences between the two constructions need to be inquired more thoroughly. Moreover, this study highlights how the analysis of spoken data can be valuable for syntactic theory, providing examples of phenomena that would otherwise be lost.

**References**

Cappelle, B. (2006) Particle placement and the case for 'allostructions'. *Constructions*, SV 1(7), pp. 1-28.

Davies, W. D. & Dubinsky, S. (2004) *The grammar of raising and control: A course in syntactic argumentation*. Malden: Blackwell Publishing.

De Mauro, T., Mancini, F., Vedovelli, M. & Voghera, M. (1993) *Lessico di frequenza dell'italiano parlato*. Milano: Etaslibri. Available at: http://badip.uni-graz.at/it [accessed 13 April 2022].

De Vaere, H., De Cuypere, L. & Willems, K. (2018) Alternating constructions with ditransitive 'geben' in present-day German. *Corpus Linguistics and Linguistic Theory*, 17(1), pp. 73-107.

Gries, S. Th. (1999) Particle movement: A cognitive and functional approach. *Cognitive Linguistics*, 10(2), pp. 105-145.

Gries, S. Th. (2003) *Multifactorial analysis in corpus linguistics: A study of particle placement*. London: Continuum.

Mauri, C., Ballarè, S., Goria, E., Cerruti, M. & Suriano, F. (2019) KIParla corpus: a new resource for spoken Italian. In: Raffaella Bernardi, Roberto Navigli & Giovanni Semeraro (eds.), *Proceedings of the 6th Italian Conference on Computational Linguistics CLiC-it, November 13-15, 2019, Bari, Italy*. Available at: https://kiparla.it/search/ [accessed 13 April 2022]

Nissim, M. Dingare, S., Carletta, J. & Steedman, M. (2004) An Annotation Scheme for Information Status in Dialogue. In: Maria Teresa Lino, Maria Francisca Xavier, Fátima Ferreira, Rute Costa & Raquel Silva (eds.), *Proceedings of the 4th Conference on Language Resources and Evaluation (LREC2004), May 2004, Lisbon, Portugal*. Available at: http://www.lrec-conf.org/proceedings/lrec2004/pdf/638.pdf [accessed 13 April 2022].

Perek, F. (2015) *Argument Structure in usage-based Construction Grammar*. Amsterdam: John Benjamins.

Rossini Favretti, R., Tamburini, F. & De Santis, C. (2002) CORIS/CODIS: A corpus of written Italian based on a defined and a dynamic model. In: Andrew Wilson, Paul Rayson & Tony McEnery (eds.), *A Rainbow of Corpora: Corpus Linguistics and the Languages of the World*, pp. 27.38. München: Lincom-Europa. Available at: http://corpora.dslo.unibo.it/TCORIS/ [accessed 13 April 2022]

Stefanowitsch, A. (2003) Constructional semantics as a limit to grammatical alternation: The two genitives of English. In: Günter Rohdenburg & Britta Mondorf (eds.), *Determinants of Grammatical variation in English*, pp. 413-443. Berlin-Boston: De Gruyter Mouton.