

LANCSBOX OPTIONS FOR RESTRICTIVE FOCUSING ADVERBS STUDIES (BASED ON ICE CORPORA)

Adverbs represent the most heterogeneous class of lexicon, which previously was considered a “wastepaper basket” for all the words that did not fit the main categories, viz. nouns, verbs or adjectives [12]. Focusing adverbs (FAs), singled out as a class at the end of the XX-century, are characterized from a two-plane perspective based on their role for the truth condition of the utterance and sentence presupposition [15]. Referring to this class in modern grammars scholars also label such lexical units as “focusing modifiers” [7; 8;13], “focusing particles” [3;11; 14], “scalar operators” [9]. The current paper proposes the term focusing adverbs agreeing with T. Nevalainen that these elements differ syntactically from “typical particles” or modifiers because they are 1) associated with another clause element and phrasal categories; 2) “most of these class items are marked by derivational suffix *-ly*”, which connects them lexically to the class of adverbs [16, p. 4]; 3) they have diachronically and synchronically related homonyms in other word classes or can have different functional variants [1; 2].

FAs fall into two major groups: restrictives (*only, just, exactly* etc.) and additives (*also, even, too, etc.*) [14]. Typical of this type in Present-Day English (PDE) is their role in marking sentence focus and their predominantly preadjacent position to the focused part, due to the rigid SVO word order.

The current study of the restrictives centers around the search term *only* in three subcorpora of International Corpus of English (ICE) – British, American and Canadian – compiled with the aim to collect the material for the comparative investigation of the English language worldwide. Every ICE corpus contains 1 million words (i.e. 500 texts of ca. 2,000 words each) exemplifying spoken and written English of the 1990s and later. To ensure the compatibility among the corpora components the research team follows a common corpus design and annotation scheme in order to obtain relatively balanced data for comparison [17].

The automated analysis of the lexeme *only* in ICE Corpora is made by means of #LancsBox, a new generation software package for the study of language data and corpora [6]. The main features of the software are 1) working with user’s data or existing corpora (can be loaded in txt, .xml, .doc, .docx, .pdf, .odt, .xls, .xlsx and other formats); 2) visualizing language facts; 3) analyzing the data in any language; 4) automatic annotation of data for part-of-speech (POS); 5) working with major operating systems (Windows, Mac, Linux) [5]. The main asset of the software lies in automated research on word associations, identifying collocates based on three traditional criteria: distance (specifying the span around a node word, ‘collocation window’), frequency (an important indicator of typicality of word association) and exclusivity [5]. The other criteria in line with S. Gries [10] are directionality (the strength of the attraction between two words), dispersion (the distribution of the node and the collocates in the corpus) and type-token distribution among collocates (takes into account the strength of collocational relationship and the level of competition for the slots around the node word from other collocate type). Additionally, the developers of #LancsBox take into account the connectivity between individual collocates [6].

To simplify the data search and visualize the results obtained the following tools from #LancsBox package have been used: KWIC (provides co-textual information about the token under investigation. It generates a list of all instances of a search term in a corpus in the form of a concordance. Double clicking on the node opens a pop-up window with a larger number of the texts which allows investigating the word in a broader context), Words (which main function is to seek words belonging to the same word class), GraphColl (provides information on the collocational patterning of the node search, visualizing both right and left collocates in a collocation network graph according to three parameters: strength, frequency, position). The Words tool allows in-depth analysis of frequencies of types, lemmas and POS categories as well as comparison of corpora using the keywords technique. The Ngrams tool allows the in-depth analysis of frequencies of ngram types, lemmas and POS categories as well as comparison of corpora using the key ngram technique [4; 5].

The study of the distribution figures for search term *only* in ICE-Britain has indicated that it is exemplified 1583 times in the corpus (the word is found in 14.013 tokens per 10k in 447 out of 500 texts). The similar frequency figures are characteristic of ICE-USA: 576 sentences (13.262 per 10k in 175 out of 401 texts). Interestingly, this distribution in ICE-Canada has proven lower, viz. 1180 tokens (10.537 per 10k, 422 out of 506 texts).

Collocates analysis of *only* has shown that three language variations have 6-top frequent collocates in common, which testifies to the high similarity in terms of *only* usage in all PDE variations. The study of the major collocates based on collocation frequency, defined as 01-Freq (5.0), L1-R1, C: 5.0-NC: 5.0, has allowed the assumption that *only* oft-times functions as an adjective with the most recurrent collocate ‘*the only*’ registered in all three corpora. The presence of a definite article indicates that this collocate belongs to an DP with N as its head: [DP[D[the]][Adj[only]] [N[...]]]. Therefore, it has been hypothesized that being a part of the DP *only* normally operates as an adjective. However, further KWIC investigation based on contextual analysis has evidenced that collocates analysis does not correlate with adjectival or adverbial distribution of *only* in text. Thus, the meaning of the latter is registered twice more often in comparison to adjectival sense and all three ICE corpora indicate the predominant usage of *only* as a focusing exclusive (74.58–83.03%), with the share of adjectival meaning varying from 13.54% to 21.02% in different ICE Corpora. Particularly this tendency is obvious for American variation, while British and Canadian English suggest that *only* has developed other additional meanings, viz. conjunction, connector *except*, etc.

Word order analysis has determined the low frequency of inverted element arrangement in the sentences in all PDE variations, with the dominant SVO and adverb pre-modifying position in relation to the constituent it refers to (80.19–92.44%). The information-structural analysis of ICE Corpora states that highlighting information focus and new information by *only* is its dominant function in PDE with *only* referring to VP group in 64.02–70.32% of instances. The figures for *only* marking identificational focus and given information vary from 4.52% to 13.42%, demonstrating the lowest recurrence in ICE-Canada, while in ICE-Britain and ICE-USA the frequency figures are almost identical (12.62% and 13.62% respectively). Marking by *only* a contrastive focus is significantly lower in British and Canadian English (9.64% and 7.1% correspondingly), when compared to the texts from ICE-

USA (16.46%). When marking emphatic focus and either old or new information *only* is observed in sentences with inverted word-orders or is placed in post-position to the element it refers to.

Acknowledgements

This work is supported by die Nationale Akademie der Wissenschaften Leopoldina (Das Leopoldina Ukraine Distinguished Fellowship) provided to Olena Andrushenko.

REFERENCES

1. Andrushenko, O. (2022a). The landscape of Middle English focusing adverb even. *Litera: Journal of language, literature and culture studies*, 32 (2), p. 861–883.
2. Andrushenko, O. (2022b). The Scope of *just*: evidence from information-structure annotation in diachronic English Corpora. In *Proceedings of the 6th international conference on computational linguistics and intelligent systems (COLINS 2022)*, Gliwice, Poland, May 12-13, 2022, Volume I: Main Conference. Sharonova, N., Lytvyn, V., et al. (eds.), p. 677–696. Available at: <https://ceur-ws.org/Vol-3171/paper51.pdf>.
3. Benazzo, S., Patin, C. (2017). French additive particle aussi: Does prosody matter? In *Focus on additivity: adverbial modifiers in Romance, Germanic and Slavic Languages*. De Cesare, A.-M. & Andorno, C. (eds.). Amsterdam, Philadelphia: John Benjamins Publ., p. 107–136.
4. Brezina, V., Weill-Tessier, P., & McEnery, T. (2020). *#LancsBox 5.x and 6.x [software]*. Available at: <http://corpora.lancs.ac.uk/lancsbox>.
5. Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge: Cambridge University Press.
6. Brezina, V., McEnery, T., & Wattam, S. (2015). Collocations in context: A new perspective on collocation networks. *International journal of corpus linguistics*, 20 (2), p. 139–173.
7. De Cesare, A.-M. (2017). On 'additivity' as a multidisciplinary research field. In *Focus on additivity: adverbial modifiers in Romance, Germanic and Slavic languages*. De Cesare, A.-M. & Andorno, C. (eds.). Amsterdam, Philadelphia: John Benjamins Publ., p. 1–22.
8. Favaro, M. (2020). From focus marking to illocutionary modification: Functional developments of Italian *solo* 'only. In *Information-structural perspectives on discourse particles*. P.Y. Modicom, O. Duplâtre (eds.). Amsterdam, Philadelphia: John Benjamins Publ., p. 111–135.
9. Gast, V. (2017). The scalar operator 'even' and its German equivalents. Pragmatic and syntactic factors determining the use of 'auch', 'selbst' and 'sogar' in the Europarl corpus. In *Focus on additivity: adverbial modifiers in Romance, Germanic and Slavic languages*. De Cesare, A.-M. & Andorno, C. (eds.). Amsterdam, Philadelphia: John Benjamins Publ., p. 202–235.
10. Gries, S. h. (2013). 50-something years of work on collocations: What is or should be next.... *International journal of corpus linguistics*, 18 (1), 137–166.
11. Grubic, M., Wierzba, M. (2021). The German additive particle noch: testing the role of topic situations. *Glossa: A journal of general linguistics*, 6 (1): 1– 29.

12. Hummel, M., Valera, S. (2017) *Adjective adverb interfaces in Romance*. Amsterdam; Philadelphia: John Benjamins Publ.

13. Kim, M.-J. (2020). *The syntax and semantics of noun modifiers and the theory of universal grammar: A Korean perspective*. Amsterdam; Philadelphia: John Benjamins Publ.

14. König, E. (2017) Syntax and semantics of additive focus markers from a cross-linguistic perspective: A tentative assessment of the state of the art. In *Focus on additivity: adverbial modifiers in Romance, Germanic and Slavic languages*. De Cesare, A.-M. & Andorno, C. (eds.). Amsterdam, Philadelphia: John Benjamins Publ., p. 23–43.

15. Nakagaw, N. (2020). *Information structure in spoken Japanese: particles, word order and intonation*. Berlin: Freie Universität.

16. Nevalaine, T. (1991). *BUT, ONLY, JUST: focusing adverbial change in Modern English 1500- 1900*. Helsinki: Societe Neophilologique.

17. Scheube, F. (2006). *The international corpus of English (ICE)*. München: GRIN Verlag.