Benchmark keywords in European integration discourse

A. Kryvenko

Kyiv National Linguistic University, Kyiv, Ukraine

Abstract. This paper adopts a keyword approach to the identification of nodal points – sites of a particular discursive concentration, which have to do with the reproduction and construction of theories of European integration in written academic discourse. It lays out a comprehensible methodology for singling out relevant keywords in comparable specialized sub-corpora on European integration, which allows reducing subjectivity during the selection process. Quantitative data obtained from multiple sub-corpora allow for a layered comparison and potentially facilitate a qualitative analysis of the construal of European integration in various discursive practices.

Keywords: European integration discourse, corpus-assisted discourse studies (CADS), specialized corpus, nodal points, keyword analysis.

Introduction. The institutionalization of professional discourse and its growing role in social practices have been in the spotlight of multidisciplinary discourse studies for a few decades. A linguistic take on discourse treats it primarily as "a means of talking and writing about and acting upon worlds, a means which both constructs and is constructed by a set of social practices within these worlds, and in so doing both reproduces and constructs afresh particular social-discursive practices, constrained or encouraged by more macro movements in the over-arching social formation" (Candlin 1997: viii). At the same time, it is inevitable for research focusing on the constitutive role of discourse – and particularly language – in the dynamics of social formations to draw on works in other humanities (Wodak 2018: 3–6). On the other hand, the linguistic turn in modern social theory is paving new avenues for further interdisciplinary insemination. One case in point is the process of European integration, which has claimed its own field of research. European integration studies, traditionally grounded in political science and international relations, have recently developed discursive approaches to the analysis of European integration (Waever 2009: 163) through making meaning of Europe, organizing Europe and drawing its borders (Wodak 2018: 11-12) by means of language as well as other modes of social semiotics.

Making meaning of Europe or the EU in general and European integration in particular has been the object of numerous investigations focusing, among other issues, on legitimation strategies in political speeches on Europe (Weiss 2002), the discursive construction of European identities (Wodak & Weiss 2004; Krzyżanowski 2010), conceptual metaphors of Europe (Musolff 2004; Ya-vorska, Bohomolov 2010), and narratives of Europe's common past (Forchtner & Kølvraa 2012). However, research on socially loaded meanings in terms of post-structuralist discourse theory is a challenging endeavor, since meanings are "fragile and contested construction(s) of the discourse participants" (Wodak 2018: 4); they are formed in use while reappearing in other texts, being decontextualized and recontextualized (ibid.: 7). Although socially loaded meanings lack a transcendental center, they may be determined "within a relational system which is provisionally anchored in nodal points that are capable of partially fixing a series of floating signifiers" (Torging 2005: 13-14). Within this framework of research, a nodal

point is defined as "the site of a particular discursive concentration – the point at which several associative chains are condensed" (Reyes 2005: 242). In order to single out nodal points in discourse it has been suggested to look for series of semantically similar terms rather than the recurrent use of a single term in relevant texts (ibid.). However, this technique lacks procedural clarity, imposes significant limitations on the volume of language data that can be analyzed this way (which, in its turn, questions the representability of a certain discourse in the texts under research), and unwarrantably reduces the scope of the relational system within which "floating signifiers" are determined.

In what follows, I investigate how keyness in specialized corpora can be of use in determining nodal points in European integration discourse.

Literature review. In corpus linguistics, keywords are statistically significant words which are used in the texts of a target corpus more frequently (or less frequently, if negative keyness is investigated) than in the texts of a reference corpus (Scott and Tribble 2006; Baker 2006; for other meanings of the term *keyword* see, e.g. Stubbs 2010). Keywords analysis has increasingly been used for the exploration of specialized discourses with the intention to unearth words having a measure of saliency, which might indicate the propositional content of a corpus and its overall semantic cohesion. By extension, keywords might point to sets of meanings and values, which are of importance to some particular communities or institutions (Groom 2010: 59) and therefore be handy in research into socially loaded discourses.

One of the fundamental challenges in applying a keyword approach is narrowing down a list of automatically generated candidate items, which might contain hundreds or even thousands of keywords, for further analysis, even when tight statistical constraints are imposed (ibid.: 60-61). It is commonly done by either cutting off anything below the top 20 – 100 items and/or by taking into account only certain lexical or grammatical categories of keywords depending on the research objectives. A more efficient way of controlling keyword lists for undesired variables is avoiding a general reference corpus for the automatic generation of keywords altogether and using another specialized corpus for this purpose instead. For instance, in their examination of the construal of Europe in the British press over time, Marchi and Taylor (2009:

208-209) mine two compatible sub-corpora representing two different periods and use each corpus as the other's reference corpus in their calculation of keywords. However, despite applying this technique along with the threshold for keyness set at a low probability value, the full keyword lists still needed to be "manually (and therefore subjectively)" (ibid.) sifted for the relevant items.

The aim of this paper is to lay out a comprehensible methodology for singling out relevant keywords in comparable specialized sub-corpora on European integration, which allows reducing subjectivity during the selection process. It is claimed that developing more rigorous procedures for singling out keywords for further analysis is an important step in determining nodal points in European integration discourse.

Data and methods. The methodological framework of my study rests on the recently established trend to combine quantitative and qualitative approaches to discourse analysis in the spirit of Corpus Assisted Discourse Studies (CADS), which aim "at identifying distinctive features and investigating non-obvious meanings within specific discourse types" (Marchi and Taylor 2009: 203). Importantly, CADS-based analysis is comparative by nature.

The data for this study come from an ad-hoc built specialized corpus of texts theorizing European integration in the field of European integration studies. Although it is standard practice to explore research articles in corpus-based studies of academic discourse (Swales 2004; Malavasi and Mazzi 2010; Parodi 2015), they might not be suitable for some types of corpus queries. In fact, it has been pointed out that research articles in various disciplines tend to report on unusual results and exceptions rather than on the norm (Fruttaldo and Cambria 2017: 299). Hence, when the purpose is to uncover conventional and recurrent patterns of meaning-making in academic written discourse, such as making meaning of European integration, a corpus of research articles is hardly the best choice for data mining, since a keyword analysis of such a corpus is likely to produce skewed and misleading results. Also, there is enough empirical evidence showing disciplinary diversity when it comes to preferences in academic genres, language choices or epistemic conventions, namely ways of constructing, formulating, negotiating, and disseminating knowledge (Malavasi and Mazzi 2010: 170; Parodi 2016: 492ff). In fact, the field of European integration studies is quite peculiar in this regard, as it lies at the crossroads between political science and international relations. It is characterized by theoretical dispersion and a multitude of approaches grounded in a variety of ontological and epistemological assumptions (Wiener and Diez 2009: 16), which, in a way, is iconic to the multiperspectival character of the European Union.

With this in mind, it was decided to compile a corpus of texts that are recognized as a standard reference in the field of European integration studies, can work as a point of access to the disciplinary information in its development over time as well as capture some conventional and recurrent lexical patterns describing various aspects of European integration. The best candidate available for this role proved to be *European Integration Theory* (Wiener and Diez 2009), a multi-authored volume with chapters on federalism, neofunctionalism, liberal intergovernmentalism, governance approaches, policy networks, institu-

tionalism, social constructivism, discursive approaches, gender, normative theory, and critical political economy. Its academic discourse genre was identified as largely belonging to the "Disciplinary Text" type with some elements of the "Textbook" type (in terms of Parodi 2015). This volume predominantly presents, elaborates, discusses, and reflects upon various theories of European integration. At the same time, it does include a guide to further readings and study questions at the end of each chapter, which were added in the second edition. The editors of the volume are quite explicit about their communicative purposes and their target audience: not only to educate students of integration, but also to contribute to the general public debate and to further "the theorizing of European integration itself" (Wiener and Diez 2009 preface: vii).

For the extraction of keywords, each chapter of the volume with the exception of the preface and the introduction was treated as a target sub-corpus for this research, whereas the whole text of the volume (including the preface and the introduction as well as the bibliography and the index) was used as a reference corpus for each of the target sub-corpora. Keywords were automatically calculated for each of these eleven sub-corpora with the help of AntConc 3.4.4w functions based on log-likelihood (LL), with the average value filter. The LL value, or keyness, is high where there is a significant disparity in the relative frequencies of a word in the target and reference corpora; however, it should be kept in mind that the LL values are "greatly affected by the size of the corpora being analyzed" (Pojanapunya, Todd 2016: 14). Since the size of each target sub-corpus varies, ranging on average between 8,000 and 10,000 tokens (with only two chapters noticeably diverging from this range), the LL values in eleven keyword lists were compared only for the sake of ranking words according to their keyness. The total size of the reference corpus is 148545 tokens as counted by AntConc.

Other AntConc tools extracting word lists, collocates, clusters/n-grams and concordances were employed for further analysis of the selected keywords.

Results and discussion. Comparisons of keyword lists produced for each sub-corpus, as described above, are a good starting point for tracing the linguistic establishment of European integration across different theoretical approaches to the subject matter while preserving a greater distance between the observer and the data (Partington 2006: 268), when needed.

The application of the threshold value at the level of the average value in the calculation of keyness resulted in anywhere between 240 and 460 keywords with most lists containing over 300 keywords. It comes as no surprise that some of the highest-ranking keywords in each of the sub-corpora are related to the names of the respective theoretical approaches or their branches. For instance, *institutions, institutionalist, institutionalism, institutionalists, institutionalisms, institutional* are ranked 3rd, 4th, 12th, 29th, and 43rd in the keyword list of the "New Institutionalisms and European Integration" sub-corpus and *rational, choice, historical,* and *sociological* (ranking 1st, 2nd, 6th and 36th) refer to rational choice, historical, and sociological types of institutionalism. The surnames of the cited authors contributing to various approaches in the field of

European integration studies also have a relatively high keyness in the sub-corpora. More importantly for linguistic analysis, the third clearly distinguishable group of the keywords points to the semantic fields, which are core to the sub-corpora under investigation and have to do with their "aboutness". In the "Gender and European Integration" sub-corpus, semantically related keywords include *gender* (1st), *equality* (2nd), *women* (3rd), *feminist* (4th), *masculinity* (6th), *family* (12th), *men* (16th), *gendered* (17th), *feminists* (20th), *patriarchal* (22nd), *male* (24th), *femininity* (30th), *masculine* (36th), and *masculinist* (47th) with the respective ranks above. In the "Critical Political Economy" sub-corpus, in addition to *economy, economic, macroeconomic, and economics*, the other indicative keywords are *capitalism, capital, monetary, productivity, production, growth, labour, investment, currency,* and *mercantilism*, ranking 2nd, 3rd, 6th, 11th, 14th, 21st, 29th, 33rd, 34th, and 47th respectively.

Less obvious to the naked eye looking through the keyword lists are dispersion patterns of the keywords across the target sub-corpora. However, a regular wordlist compiled out of all eleven sub-corpora keyword lists reveals that six keywords are found in five sub-corpora, sixteen keywords are found in four sub-corpora and ninety-two keywords are found in three sub-corpora. Grammatical words aside, the two most salient keywords are *EU* and *actors* both in terms of their positive and negative keyness, as shown in Table 1. The other prominent keywords, which appear in any three sub-corpus keyword lists, include but are not limited to *agenda, benefits, CEE (Central and Eastern European), commitment, costs, diversity, effects, EP (European Parliament), Europe, governments, identity, interests, member, negotiations, preferences, states,* and *values*.

Table 1. Dispersion, ranking and raw frequencies of the selected positive and negative (-) keywords across the sub-corpora

	Fede-ralism	Neo-funct.	Inter-govern.	Gover-nance	Policy netw.	Institu-tionalism	Con-struct.	Discur-sive	Gender	Norm. theory	Crit. polit. econ.
EU		-2nd/27			6th/152	73rd/110	117th/87	-2nd/22	15th/123	43rd/146	-1st/23
actors	-9th/1	97th/31		13th/36	10th/42	100th/30	21st/32	-84th/9			-4th/1

A further comparison of the mentions of the EU, the European Union and Europe in the sub-corpora revealed that although the total number of the EU mentions exceeds the mentions of Europe almost three times and the mentions of the European Union over eleven times, three sub-corpora strongly prefer *Europe* over the *EU* and the *European Union*. Only in the sub-corpus on liberal intergovernmentalism does the raw frequency for *Europe* equal the sum of the raw frequencies for the *EU* and the *European Union*. Importantly, collocates for the search terms *EU* and *Europe* in the -5 to +5 span noticeably vary. The top lexical co-occurrences for the search term *EU* include *policy/policies, member, states, gender, institutions, enlargement, making, governance*, whereas *Europe* eagerly co-occurs with *federal, political, eastern, state, building, modern, nation, central, social, choice,* and *concepts*. These results are not quite consistent with the findings reported by Marchi and Taylor (2009: 205-211, 222) for the British press, which suggests that not only are the terms for naming Europe treated as non-interchangeable, but also the reasons for these choices may have to do with genre sensitivity as well as the conceptual structure of the texts being grounded in the nature of the discourse.

The analysis of collocates and n-grams as well as a closer reading of concordance lines confirm the importance of *actors* as yet another "site of a particular discursive concentration" in European integration discourse. The actors are linguistically specified via premodifying adjectives such as *social, political, economic, national, supranational, regional, local, private, public,* and *promiscuous*; appositives *including corporate actors such as national governments, firms or interest groups*; postmodifying prepositional phrases like *in the target countries, in the real world of the EU, in the political process*, etc. The actors are further linguistically established with

the help of some other keywords including *interests, preferences, commitment,* and *values*. Negative keyness of the term *actors* is coupled with its rather negative semantic prosody in the relevant sub-corpora (see Table 1): *the more fine-grained and pragmatic question about whether supranational actors exercise autonomous authority in the political process* ("Critical Political Economy"); *all three approaches acknowledge, admittedly to differing degrees, the reality of supranational actors and institutions* ("Federalism"); *another aspect of the relationship between actors and discourses is whether actors are fully in the grip of discourses; a more radical deconstructivist would instead want to emphasize how language plays games with actors* ("Discursive approaches"). Negative keyness of the term *actors* in a few sub-corpora under study is likely to highlight epistemological facets of several theoretical approaches to European integration.

Conclusions. This paper intended to create a background for uncovering and further exploring the nodal points pertinent to European integration discourse with the help of more rigorous procedures provided by corpus linguistics. In particular, the keyword analysis proved a good starting point for identifying particular discursive concentrations as condensed associative chains. Quantitative data obtained from multiple sub-corpora allow for a layered comparison and facilitate a qualitative analysis of the construal of European integration in English academic discourse. However, further research is required to examine intercollocability between selected keywords to better comprehend recurrent and conventional ways of talking and thinking about European integration in different discursive settings. This methodology might be suitable for the linguistic analysis of nodal points in a variety of other relevant contexts, including but not limited to significantly larger corpora or corpora representing other genres of the same discourse.

REFERENCES

1. Baker, P. Using Corpora in Discourse Analysis. London: Continuum, 2006. 206 p.
2. Candlin, Ch. General editor's preface // The Construction of Professional Discourse ⁄ ed. by B. L. Gunnarsson, P. Linell and B. Nordberg. London; N.Y.: Routledge, 1997. P. viii-xiv.
3. European Integration Theory / ed. by A. Wiener and Th. Diez. 2nd ed. 2009. xvi, 295 p.
4. Forchtner, B. and Kølvraa Ch. Narrating a 'new Europe': From 'bitter past' to self-righteousness? // Discourse and Society, 2012. 23 (4): 377–400.
5. Fruttaldo, A. and Cambria M. Combining qualitative and quantitative approaches to discourse analysis: in conversation with Gerlinde Mautner and Alan Partington // Im@go, 2017. 9(VI): 284-304.
6. Krzyżanowski, M. The Discursive Construction of European Identities: A Multi-level Approach to Discourse and Identity in the Transforming European Union. Frankfurt/Main: Lang, 2010. 232 p.
7. Malavasi, D. and Mazzi, D. 2010. History v. marketing: keywords as a clue to disciplinary epistemology // Keyness in Texts / ed. by M. Bondi and M. Scott. Amsterdam: John Benjamins, 2010. P. 169-184.
8. Marchi, A., and Taylor, Ch. Establishing the EU: The representation of Europe in the press in 1993 and 2005 // Corpora: Pragmatics and Discourse. Language and computers. Vol. 68 / ed. by A. H. Jucker, D. Schreier, and M. Hundt. Amsterdam: Rodopi, 2009. P. 201-224.
9. Musolff, A. Metaphor and Political Discourse. Analogical Reasoning in Debates about Europe. Basingstoke: Palgrave, 2004. viii, 211 p.
10. Parodi, G. Variation across university genres in seven disciplines: a corpus-based study on academic written Spanish // International Journal of Corpus Linguistics, 2015. 20(4): 469-499.
11. Partington, A. Metaphors, motifs and similes across discourse types: Corpus-Assisted Discourse Studies (CADS) at work // Corpus-based Approaches to Metaphor and Metonymy /ed. by A. Stefanowitsch and S.Th. Gries. Berlin: Mouton de Gruyter, 2006. P. 267-304.
12. Pojanapunya, P. and Watson Todd, R. Log-likelihood and odds ratio: Keyness statistics for different purposes of keyword analysis // Corpus Linguistics and Linguistic Theory, 2016. 0(0). Published online.
13. Reyes, O. New Labour's politics of the hard-working family // Discourse Theory in European Politics: Identity, Policy and Governance / ed. by D. Howarth and J. Torfing. Basingstoke: Palgrave MacMillan, 2005. P. 231-254.
14. Scott, M. and Tribble, C. Textual Patterns: Keyword and Corpus Analysis in Language Education [Studies in Corpus Linguistics 22]. Amsterdam: John Benjamins, 2006. x, 203 p.
15. Stubbs, M. Three concepts of keywords // Keyness in Texts / ed. by M. Bondi and M. Scott. Amsterdam: John Benjamins, 2010. P. 21-42.
16. Swales, J. Research Genres: Exploration and Applications. Cambridge, UK: Cambridge University Press, 2004. 314 p.
17. Torging, J. Discourse Theory: Achievements, Arguments, and Challenges // Discourse Theory in European Politics: Identity, Policy and Governance / ed. by D. Howarth and J. Torfing. Basingstoke: Palgrave MacMillan, 2005. P. 1-32.
18. Waever, O. Discursive approaches // European Integration Theory / ed. by A. Wiener and Th. Diez. 2nd ed. 2009. P. 163-180.
19. Weiss, G. Searching for Europe: the problem of legitimisation and representation in recent political speeches on Europe // Journal Language & Politics. 2002. 1(1): 59-83.
20. Wiener, A. and Diez, Th. Preface // European Integration Theory / ed. by A. Wiener and Th. Diez. 2nd ed. 2009. P. vii-viii.
21. Wodak, R. Discourse and European integration // MIM Working Series, 2018. 18(1): 3-29.
22. Wodak, R. and Weiss, G. Visions, Ideologies and Utopias in the Discursive Construction of European Identities: Organising, Representing and Legitimising Europe // Communicating Ideologies: Language, Discourse and Social Practice / ed. by M. Pütz, J.A. Neff-van Aertselaer and T. A. van Dijk. Frankfurt: Lang, 2004. P. 225-252.
23. Yavorska H., Bohomolov O. An Uncertain Object of Desire: Europe in the Ukrainian Political Discourse. Kyiv: Publishing House of Dmytro Burago and A. Krymskyi Institute of Oriental Studies of the National Academy of Sciences of Ukraine, 2010. 136 p.

Эталонные ключевые слова в евроинтеграционном дискурсе

А. Л. Кривенко

Аннотация. В статье используется метод ключевых слов для идентификации узловых точек на пересечении уплотненных ассоциативных цепей, которые указывают на процессы воспроизведения и конструирования теорий европейской интеграции в письменном научном дискурсе. Разрабатывается методика выделения релевантных ключевых слов в сопоставимых специализированных подкорпусах текстов по теоретическим вопросам европейской интеграции, что позволяет снизить субъективность при отборе ключевых слов для дальнейшего анализа. Такая методика позволяет провести многоуровневое сравнение количественных данных, полученных в результате обработки определенного множества подкорпусов, и способствует дальнейшему качественному анализу толкований европейской интеграции в рамках различных дискурсивных практик.

Ключевые слова: евроинтеграционный дискурс, корпусновспомогательное дискурсоведение, специализированный корпус, узловые точки, анализ ключевых слов.