

A Community Detection Approach to Identify Hedging Language Patterns

Brett M. Drury^{1,2,*}, Samuel Morais-Drury³

¹ Liverpool Hope University, Liverpool,
United Kingdom

² University of Porto, Faculty of Engineering, Porto,
Portugal

³ Colégio Puríssimo Coração de Maria, Centro, Rio Claro,
Brazil

druryb@hope.ac.uk

Abstract. Hedging language is common in business communication. It conveys uncertainty, limits commitment, and bestows plausible deniability to the speaker. Hedging language is used frequently when economic actors discuss an organization's financial prospects publicly. Economic actors make statements to the mass media for differing reasons. So far, no research has detected common economic actors' hedging language in the mass media. This paper proposes a technique to discover distinct users of hedging language. The strategy uses a graph that contains job titles and hedging lexical bundle nodes. A community detection algorithm infers groups of job titles through their use of common hedging lexical bundles. The proposed method identified three distinct communities that had their own distinct hedging language. This article discusses the differences between the communities, as well as the link of hedging lexical bundles to sentiment and emotion.

Keywords. Graph, hedging language, network science, sociolinguistics.

1 Introduction

Business leaders manipulate their audiences through rhetorical strategies in their public statements. They must adopt this strategy. If they are truthful, their organization can face dramatic consequences [17].

However, business leaders cannot lie because misleading the market is a crime which will lead to a custodial sentence. So, business leaders use strategies that will mitigate the consequences of their public utterances. These strategies include framing [6] and hedging language [3].

These strategies give business leaders plausible deniability if their organization's performance disappoints the market. Public statements, however, contain a mix of speakers that use hedging language for legitimate purposes such as uncertainty about the future. Currently, there is no definitive list that categorizes economic actors solely by their use of hedging language.

This paper proposes an approach that combines job titles and hedging lexical bundles mapped to a graph and a community detection algorithm. This will help find groups of job titles with differing uses of hedging language. They will allow for custom strategies when inferring the future prospects of organizations through public communications by associated economic actors.

2 Literature Review

A motivation to analyse public statements of economic actors is to trade shares in their organizations [9].

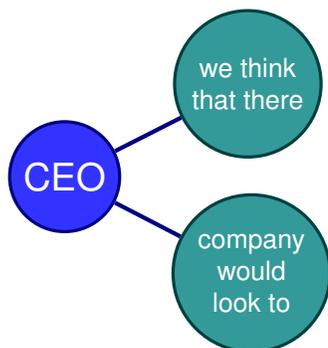


Fig. 1. Simple connection between job title node and lexical bundle nodes

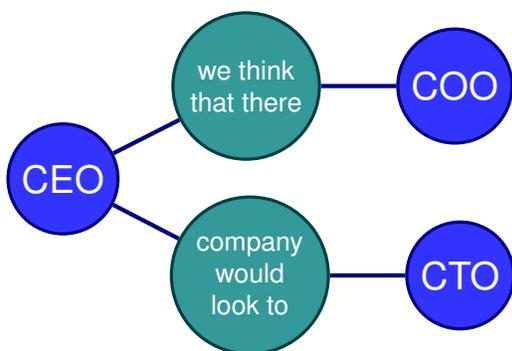


Fig. 2. Initial graph fragment

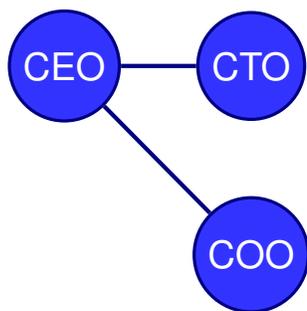


Fig. 3. Graph fragment after pruning

Business leaders have private information about their organization. This private information may leak into their public statements. These statements can show the future performance of their organization.

Traders who identify this information leak can outperform their competitors. This is not a trivial task because business leaders use hedging language to mask their true intentions.

Despite the obvious difficulties, there have been attempts to trade with public statements. The general sentiment of public statements [9] have been used to trade an index (NASDAQ), and [5] traded on business leaders' public statements when their utterances were similar to finance professionals. None of these approaches considered hedging language and relied upon sentiment. Lexical bundles are the most common word sequences.

They represent the vocabulary of a text collection [18]. Bundles can be any length. But, evidence suggests the best size is four words [16]. This approach has been used in many domains. These include Wikipedia [11], spam emails [14], and business communication [8]. The business communication research literature has demonstrated that there are significant differences between the vocabulary of individual economic actors [8].

Discourse networks are a method of mapping domain actors to specific statements or topics [13]. Networks in this case are simply nodes and edges, where the edges connect domain actors to specific statements or topics [13]. Distinct communities that are linked to a group of statements or topics can be identified through community detection algorithms [13]. This technique has been used to identify groups of opinion in the domains of minimal alcohol pricing [10], migration [19] and a sugar tax [2].

3 Methodology

The proposed method is a three-stage process that involves 1. building a graph from job titles and hedging lexical bundles 2. pruning a graph, and 3. applying a community detection to the pruned graph to discover communities of users of hedging language.

3.1 Graph Construction

The graph constructed in this step is a multigraph, therefore nodes can have more than one edge between them. The graph is constructed from the Minho Quotation Resource [7] which is a corpus of public statements by economic actors during the financial crisis.

Table 1. A full list of job titles per community

Community	Job Titles
Business Leaders	"Vice President", "Senior Executive", "SSS", "Chairman", "Spokesperson", "Deputy Director", "Special Envoy", "Solicitor", "Businessman", "Board Member", "Parliamentary Secretary", "Director Marketing", "Chief", "DPM", "Sales Analyst", "HAS", "Executive Editor", "Senior Adviser", "CAM", "CEO", "Energy Minister", "Co Head", "NSA", "Officer", "Lead Author", "Oil Minister", "DMD", "General Partner", "CSC", "Finance Minister", "CCO", "General Manager", "CSO", "Policy Director", "EVP", "Secretary", "Press Secretary", "Trustee", "Senior Manager", "Manager", "Executive", "Home Secretary", "Research Director", "Minister", "Owner", "President Ceo", "Lieutenant General", "Project Manager", "CVP", "Financial Secretary", "Founder", "Treasurer", "AAG", "Chair"
Finance Professionals	"Chief Economist", "Sergeant", "Credit Analyst", "Attorney", "CIE", "Labour Mp", "Author", "Lawyer", "Health", "Chief U S Economist", "Economist", "SIG", "SBS", "Defense Secretary", "Physician", "SMA", "Economics Professor", "JGS", "STM", "Former Ceo", "Political Scientist", "SIA", "Co Author", "Engineer", "Associate Professor", "Strategist", "Dealer", "Market Analyst", "Banking Analyst", "Governor", "Senior Analyst", "Co Chairman", "Correspondent", "Entrepreneur", "Commissioner", "Chief Scientist", "Business Secretary", "Consultant", "Oil Analyst"
Politicians	"Admiral", "Conservative Mp", "Adviser", "DDG", "ASA", "Counsel", "First Minister", "Secretary General", "National Officer", "Republican Leader", "Commander", "SCS", "Sales Manager", "Chancellor", "Senior Director", "CCS", "DFM", "Former Head", "Former President"

Table 2. The most frequent hedging lexical bundles by community

Business Leaders	Freq.	Finance Prof.	Freq.	Politicians	Freq.
we will continue to	627	i do n't think	476	as soon as possible	42
will be able to	620	will be able to	169	i do n't think	29
will allow us to	477	is likely to be	164	i think it 's	21
i do n't think	392	i think it 's	115	it would be a:	17
will enable us to	276	there will be a	101	to make sure that	17
i would like to	258	do n't think it	89	would have to be	15
we believe that the	226	it would be a	76	will have to be	13
would like to thank	220	are likely to be	72	would be able to	13
we will be able	219	do n't think we	69	it is clear that	12

The more lexical bundles that job titles have in common, the more likely they share a common hedging vocabulary. A simple example is shown in Figure 2, where three job titles are indirectly connected via two hedging lexical bundles. The Chief Executive Officer (CEO) and Chief Operating Officer (COO) are joined by the bundle, we think that there, and consequently the CEO and COO nodes have an edge. The Chief Technology Officer (CTO) and the CEO have an edge because of the bundle, company would look to.

After this step, there is a multigraph that only contains job title nodes.

3.3 Community Detection

The final step is to use a community detection algorithm, which in this case was greedy modularity communities [15] because it is a fast algorithm that is suitable for large graphs. The final graph has 638014 edges, and 298 nodes, consequently, this graph is suitable for this type

Table 3. Percentage hedging language type

Hedge Lang Type	Business Leaders	Finance Professionals	Politicians
Epistemic Verb	0.3	0.25	0.24
Epistemic Modal Verb	0.52	0.52	0.59
Approximations	0.01	0.01	0.01
Epistemic Adverb	0.04	0.05	0.05
Epistemic Adjective	0.12	0.16	0.12

Table 4. The most frequent use of epistemic verbs by community

Business Leaders	Percent	Finance Professionals	Percent	Politicians	Percent
Believe	0.28	Think	0.32	Think	0.31
Think	0.21	Believe	0.15	Believe	0.15
Expect	0.15	Expect	0.13	Expect	0.10
Feel	0.07	Hope	0.06	Hope	0.08
Understand	0.07	Understand	0.06	Understand	0.08

Table 5. The most frequent use of epistemic verbs by community

Business Leaders	Percent	Finance Professionals	Percent	Politicians	Percent
possible	0.31	likely	0.39	possible	0.31
likely	0.24	probably	0.16	likely	0.23
sure	0.16	possible	0.15	sure	0.16
probably	0.11	sure	0.10	probably	0.12
chance	0.08	unlikely	0.09	chance	0.09

Table 6. The most frequent use of epistemic modal verbs by community

Business Leaders	Percent	Finance Professionals	Percent	Politicians	Percent
would	0.40	would	0.37	would	0.44
can	0.25	could	0.19	can	0.17
could	0.14	can	0.18	could	0.16
should	0.11	should	0.12	should	0.14
may	0.06	may	0.10	may	0.06

of community detection algorithm. Communities with more than five job titles are determined to be distinct users of hedging language. At the end of this step, there were three communities. A simplified portion of the final graph can be found in Figure 4, where a node's community membership is indicated by a common colour. The code and data for this paper can be accessed from here².

²drive.google.com/file/d/1TuyfoPzJWtJTyS3N05k-7ZevmEDIMRoU/view?usp=sharing

4 Results

The proposed method discovered three distinct communities that broadly represented three types of speakers: Business Leaders, Politicians and Finance Professionals. The Business Leaders' Community contain job titles that are associated with leadership roles within private sector organizations and job titles such as Executive Chairman, VP, COO, and CEO.

The Politicians' Community has job titles that are associated with politicians who would comment upon public economic and financial issues. It contained job titles such as Premier, Attorney General and Shadow Chancellor.

The Finance Community comprises job titles associated with individuals whose employment would require them to make independent comments on economic and financial issues. The community contains job titles such as Currency Strategist, Senior Economist and Chief Economist. A selection of job titles per community can be found in Table 1. Intuitively, the communities have a semantic cohesiveness where each community have related job roles, which may dictate how the speaker will use hedging language.

5 Community Analysis

The individual communities should have distinct motivations and uses of hedging language. This section will discuss the differing characteristics of the use of hedging language by each community.

5.1 Frequent of Hedging Lexical Bundles

The first analysis is frequent lexical hedging bundles by community. The results are documented in Table 2. There was only one lexical bundle that was present in all communities, which is **i do n't think**, and a limited number of bundles that were present in two communities.

They are will be able to, i think it 's, and it would be a. A comparison of the most frequent hedging lexical bundles revealed differences in hedging language use between the communities.

Table 7. Comparison of hedging density by community

Business Leaders	Finance Professionals	Politicians
0.39	0.44	0.44

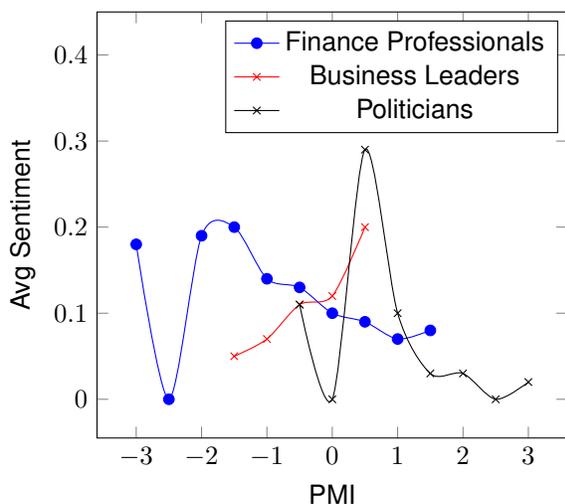


Fig. 5. Comparison of PMI against average sentiment

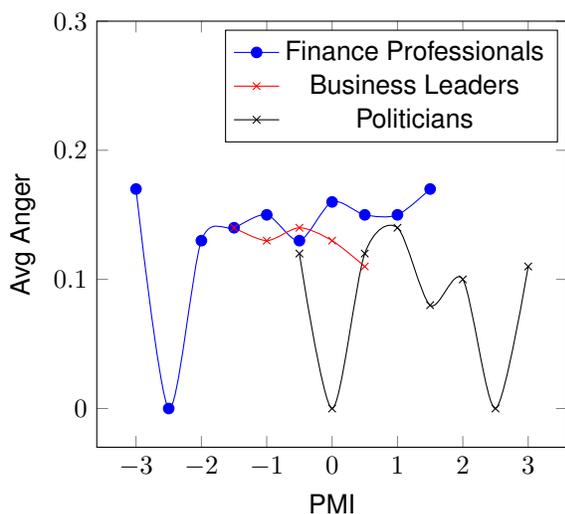


Fig. 6. Comparison of PMI against average anger

5.2 Frequency Analysis of Hedging Words

This article hypothesises that each community should have a distinct use of hedging language, as they share a common hedging vocabulary. A frequency analysis was used to determine the most popular hedging word types per community.

The analysis revealed Epistemic Modal Verbs, Epistemic Verbs, Approximations, Epistemic Adverbs and Epistemic Adjectives were the common types of hedging used by each group. The percentage breakdown of the hedging types can be found in Table 3, the most frequent use of each type per community is highlighted in bold.

There are relatively small differences between each group, with the business leaders using Epistemic Verbs relatively more often than other groups, and Finance Professionals using Epistemic Adjectives more than other groups. The Politicians used Epistemic Modal Verbs more often than the other groups. The differences in one type of hedging language per group are reflective of each group’s different roles in their communication with the mass media.

The Business Leaders group’s use of epistemic verbs is to express the speaker’s degree of certainty about a proposition. In the corpus, Business Leaders frequently use epistemic verbs in hedging statements to comment on future financial events, without commitment. A comparison of the most frequent epistemic verbs, and the percentage usage by each community, is shown in Table 4.

The most noticeable difference in the results is that the business leaders have a more frequent use of believe compared with other communities and a less frequent use of the verb think. The use of other verbs has a similar frequency across the communities. The more frequent use of the verb belief is less of a commitment to the truth of the statement than other types of hedging language.

It is important to note that the speaker from the business leaders group will have knowledge about the subject, and hedging language could be considered as a proxy of deception [3]. Epistemic adjectives form more of the hedging vocabulary of the Finance Professionals Community, than the other two groups.

They play the same role as epistemic verbs, as they allow the speaker to express a degree of belief in a statement. A comparison of the communities’ use of epistemic adjectives can be found in Table 5. The Financial Professionals Community uses the terms likely more often than the other groups, whereas the remaining community use the term

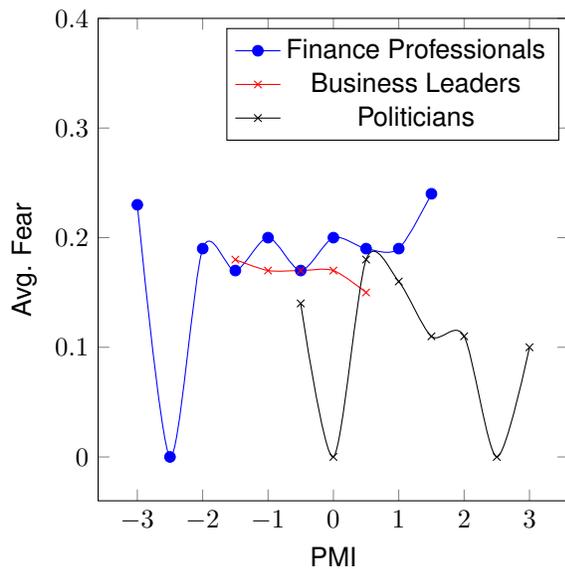


Fig. 7. Comparison of PMI against average fear

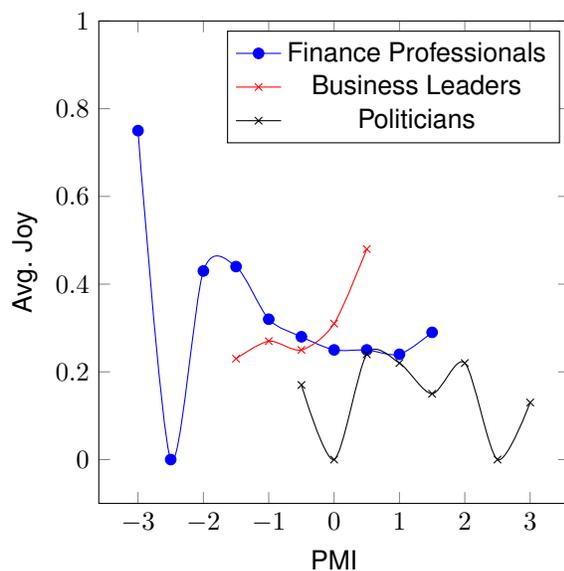


Fig. 8. Comparison of PMI against average joy

possible significantly more often than the Financial Professionals Community. In common with epistemic verbs, the use of epistemic adjectives by the Business Leaders Community is to have plausible deniability because the commitment to the statement that they are commenting on is weak by using terms such as possible.

The Business Leaders Community does not use the negative sense of epistemic adjectives frequently, whereas the Finance Professionals Community does, with the frequent use of terms such as unlikely. The commitment of the Financial Professionals Community to their statements is stronger than the Business Leaders Community, with frequent use of the likely.

The Politicians group use more epistemic modal verbs more often than the other groups. There was little variation in the most frequently used modal verbs, however, the Business Leaders Community used the stronger claim of can more than other groups, which is contrary to other to the analysis of the epistemic verbs and adjectives. The Politicians Community had a higher frequency of use of the modal verb would, which could be related to election promises, and contrary action to the current government's policies.

5.3 Community Use of Hedging Language

The communities' use of hedging language may vary because of the different motivations, consequently, the quantity of hedging may vary from community to community and the correlations with emotions or sentiment may also vary.

Appeals to emotion and positive sentiment are often used as rhetorical devices designed to manipulate audiences [4]. The analysis is designed to identify if there is a link between community specific hedging and positive emotion or sentiment, and how often each community uses hedging language.

The first analysis is to compare the percentage of sentences by each community that contain a hedging word. The results shown in Table 7 demonstrate that the Business Community uses hedging language at a lower rate than the other communities.

This is due to Business Leaders being unambiguous when they communicate positive news, as well as using scripted statements to the press [6]. A chi-squared analysis was applied to the variation in the frequency of sentences with hedging terms, which demonstrated that the variation between had a P-Value less than 0.0001.

This result infers the varying quantity of use of hedging language between the groups is influenced by a community's job roles. The remainder of the analysis is the correlation of the uniqueness of lexical bundles with a community and their correlation to sentiment or emotion. The association of a lexical bundle with a community is computed with Pointwise Mutual Information. The formula is shown in Equation 1, where $bund.$ represents a lexical bundle and $com.$ represents a community:

$$PMI(bund., com.) = \frac{P(bund., com.)}{P(bund.) P(com.)}. \quad (1)$$

The first experiment was to compare the average sentiment of the sentences that contain hedging lexical bundles against the PMI of the hedging lexical bundles in increments of 0.5. The results are in Figure 5, and it is clear from the figure that there is a correlation between sentiment and PMI of hedging lexical bundles for the Business Leaders Community.

This is confirmed by the Linear Regression result of 0.91 (R-Squared) for this community and 0.12 for the other communities. The correlation of the increasing positive sentiment with more specific hedging lexical bundles could be an indication of scripting by the Business Leaders Community, as positive sentiment is used in rhetorical communication strategies.

The communities demonstrate varying ranges of PMI with the Business Community having the smallest range, and Finance Professionals using the least community specific hedging lexical bundles, whereas the Politicians used the most specific hedging lexical bundles.

The next experiments compared the correlation of emotions against the PMI of lexical bundles. The first experiment is to compare anger against the PMI of lexical bundles. The results are in Figure 6 there is no correlation between anger and the PMI of the hedging lexical bundles, except for the Business Leaders Community where there is a very weak correlation between PMI and Anger.

The same experiment was repeated for fear, and the results can be found in Figure 7, and it is clear from Figure 7 that there is a weak correlation

between PMI and Fear for the Business Leader Community which has an R Squared score of 0.62. The final emotion that was analysed is joy, the results are shown in Figure 8. In common with the other results, there is no correlation between the Finance and Politicians Communities and the PMI of hedging lexical bundles. However, there is a correlation between the Business Leaders Community and the PMI of a lexical bundle.

5.4 Discussion

The analysis in this section demonstrates that there is relatively little difference between the Politicians and the Finance Professional Communities. However, the Business Leader Community demonstrate that they use hedging language for plausible deniability for positive statements as they use hedging terms that have a weak commitment to a proposition. The analysis of the Business Leaders Community also demonstrates that the more specific a hedging lexical bundle is to the community, then the more likely that the sentence that contains the bundle has a positive emotion or sentiment. This is a sign of scripting [1] where the statement is prepared in advance by a third party for the business leader and does not represent the view of the speaker.

5.5 Future Work

This paper proposes that the constraints of the economic actor's role are the primary determinant of their hedging language use, with factors like location or culture exerting a secondary influence. To verify this hypothesis and assess the generalizability of this technique, future research will focus on gathering data from sources in languages other than English.

6 Conclusion

This article presents a technique for grouping job titles through their use of hedging language. The technique produced three communities, and each language community had a different use of hedging language.

The Business Leaders Community seems to use hedging language as a form of plausible deniability, where positive statements can be communicated to the mass media without the consequences if the statement is false. The proposed technique allows for the grouping of speakers based on their use of a specific type of language. This technique will be used to develop custom sentiment classification for each group's statements so that it can be used in a trading strategy and take into account the specific use of hedging language to reduce the sentiment of a statement by an economic actor.

References

1. **Amernic, J. H., Craig, R. J. (2007).** Guidelines for CEO-speak: Editing the language of corporate leadership. *Strategy and leadership*, Vol. 35, No. 3, pp. 25–31. DOI: 10.1108/10878570710745802.
2. **Buckton, C. H., Fergie, G., Leifeld, P., Hilton, S. (2019).** A discourse network analysis of UK newspaper coverage of the “sugar tax” debate before and after the announcement of the soft drinks industry levy. *BMC Public Health*, Vol. 19, No. 1, pp. 1–14. DOI: 10.1186/s12889-019-6799-9.
3. **Burgoon, J., Mayew, W. J., Giboney, J. S., Elkins, A. C., Moffitt, K., Dorn, B., Byrd, M., Spitzley, L. (2016).** Which spoken language markers identify deception in high-stakes settings? Evidence from earnings conference calls. *Journal of Language and Social Psychology*, Vol. 35, No. 2, pp. 123–157. DOI: 10.1177/0261927X15586.
4. **Dowding, K. (2018).** Emotional appeals in politics and deliberation. *Critical Review of International Social and Political Philosophy*, Vol. 21, No. 2, pp. 242–260. DOI: 10.1080/13698230.2016.1196536.
5. **Drury, B., Dias, G., Torgo, L. (2011).** A contextual classification strategy for polarity classification of direct quotations from financial news. *International Conference On Recent Advances in Natural Language Processing*, Association for Computational Linguistics, pp. 434–440.
6. **Drury, B., Drury, S. M. (2021).** The identification of framing language in business leaders’ speech from the mass media. *Information Management and Big Data*, pp. 376–383. DOI: 10.1007/978-3-030-76228-5_27.
7. **Drury, B., Drury, S. M. (2021).** An update to the minho quotation resource. DOI: 10.48550/ARXIV.2104.06987.
8. **Drury, B., Drury, S. M. (2022).** Lexical bundle variation in business actors’ public communications. *International Conference on Text, Speech, and Dialogue*, Vol. 13502, pp. 339–351. DOI: 10.1007/978-3-031-16270-1_28.
9. **Drury, B. M., Almeida, J. J. (2012).** Predicting market direction from direct speech by business leaders. *1st Symposium on Languages, Applications and Technologies*, Vol. 21, pp. 163–172. DOI: 10.4230/OASlcs.SLATE.2012.163.
10. **Fergie, G., Leifeld, P., Hawkins, B., Hilton, S. (2019).** Mapping discourse coalitions in the minimum unit pricing for alcohol debate: A discourse network analysis of UK newspaper coverage. *Addiction*, Vol. 114, No. 4, pp. 741–753. DOI: 10.1111/add.14514.
11. **Hiltunen, T. (2018).** Lexical bundles in Wikipedia articles and related texts. *Applications of Pattern-driven Methods in Corpus Linguistics*, Vol. 82, pp. 189. DOI: 10.1075/scl.82.08hil.
12. **Islam, J., Xiao, L., Mercer, R. E. (2020).** A lexicon-based approach for detecting hedges in informal text. *Proceedings of the 12th Language Resources and Evaluation Conference, European Language Resources Association*, pp. 3109–3113.
13. **Leifeld, P. (2017).** Discourse network analysis: Policy debates as dynamic networks. *The Oxford Handbook of Political Networks*, pp. 301–326. DOI: 10.1093/oxfordhb/9780190228217.013.25.

14. **McVeigh, J. (2018).** Lexical bundles and repetition in email marketing texts. John Benjamins Publishing Company, pp. 213–250. DOI: 10.1075/scl.82.09mcv.
15. **Newman, M. E. J. (2006).** Modularity and community structure in networks. National Academy of Sciences, Vol. 103, No. 23, pp. 8577–8582. DOI: 10.1073/pnas.0601602103.
16. **Pinna, A., Brett, D. (2018).** Constance and variability: Using PoS-grams to find phraseologies in the language of newspapers. John Benjamins Publishing Company, pp. 107–130. DOI: 10.1075/scl.82.05pin.
17. **Ratner, G. (2008).** Gerald ratner: The rise and fall... and rise again. John Wiley and Sons.
18. **Schneider, G., Grigonytė, G. (2018).** From lexical bundles to surprisal and language models. Applications of Pattern-driven Methods in Corpus Linguistics, Vol. 82, pp. 15–56.
19. **Wallaschek, S. (2019).** Contested solidarity in the euro crisis and Europe's migration crisis: A discourse network analysis. Journal of European Public Policy, Vol. 27, No. 7, pp. 1034–1053. DOI: 10.1080/13501763.2019.1659844.

Article received on 10/04/2024; accepted on 12/07/2024.

**Corresponding author is Brett M. Drury.*