

# Russia and Ukraine through the Eyes of ParlaMint 4.0: A Collocational CADS Profile of Spanish and British Parliamentary Discourses

María Calzada Pérez

Departamento de Traducción y Comunicación, Universitat Jaume I (Castellón, Spain)  
calzada@uji.es

## Abstract

This article resorts to mixed methods to examine British and Spanish parliamentary discourse. The quantitative corpus-assisted (lexical priming) theory and data are complemented by the qualitative discourse historical approach. Two CLARIN ParlaMint corpora – ParlaMint-GB and ParlaMint-ES – are queried in the analysis, which focuses on English (“Russia” and “Ukraine”) and Spanish (“Rusia” and “Ucrania”) nodes and collocations. In sum, the analysis sketches a brief profile of each corpus. The British House of Commons is more homogenous, strongly associating “Russia” and “Ukraine” with their participation in the war. Furthermore, this chamber shows a greater interest in “Russia”. The Spanish Congreso de los Diputados indicates greater quantitative differences (heterogeneity). Here, “Russia” clearly transcends its role as a military contender and is also portrayed as an economic competitor for the West. Unlike in Britain, the Spanish lower house shows more mentions of “Ucrania”, which is assigned just one role – as an invasion victim. In conclusion, the productivity of corpus-assisted mixed methods is confirmed along with the precious value of the ParlaMint constellation.

**Keywords:** Parliamentary Discourse, ParlaMint, Russia/Ukraine.

## 1. Introduction

Parliaments are institutions of the utmost importance. Democratic systems count on them to safeguard political representation and accountability. They are not just a mirror on which societies look but also spaces where politicians propose, discuss, and justify their actions. Most importantly, they are responsible for drafting and passing the laws citizens abide by. Their function is, therefore, essential to uphold equality, transparency, and fairness.

It is no wonder they have already attracted attention from a wide variety of areas, notably political science (see, for instance, Box-Steffensmeier, Brady, and Collier, 2008; Hix, Noury, and Roland, 2006; Bütkofer and Hug, 2015) and sociology. Skubic and Fišer (2022) are particularly illuminating for a literature review on the latter discipline since they identify the most prominent topics discussed in sociology on parliamentary discourse. Furthermore, they list the prolific methods to do so, among which the gamut of (critical) discourse studies excel, informing over 60% of the sociological analyses reviewed. The authors (Skubic and Fišer, 2022: 82) advocate that “the goal of sociological research of parliamentary discourse is to analyze political discourse and language”. Hence, it is hardly surprising that they highlight the role of linguistics when approaching parliaments, and they recommend synergies with language-oriented studies.

Linguistics has also taken an interest in parliamentary/political discourse, as Calzada Pérez (2018) serves to testify. This work points to a growing pool of analyses approaching lower and upper houses from various prisms and targeting the micro- and macro-levels of parliamentary texts and contexts. Moreover, it confirms that, on this topic, linguistics also favors (critical) discourse studies.

When sociology and linguistics examine parliamentary interventions – as attested by both Skubic and Fišer (2022) and Calzada Pérez (2018) – they tend to draw on qualitative methodologies, with the highest potential for exposing descriptive results. However, they risk falling into subjectivism due to the small number of textual samples that are often analyzed.

A potential way to avoid subjectivism in parliament-related research is by advocating mixed methods, which boost qualitative results with quantitative data. Corpus-assisted studies (or CADS) do precisely this with “impressive results” (Garzone and Santulli, 2004: 353). With its name first coined by Partington (2004), CADS has been defined as “that set of studies into the form and/or function of language as communicative discourse which incorporates the use of computerized corpora in their analyses” (Partington, Duguid, and Taylor, 2013: 10). In other words, CADS uses corpus linguistics as a means to produce and dissect textual data for discourse studies.

Nevertheless, only a handful of analyses resort to CADS to examine parliamentary communication (e.g., Baker, 2006; 2010; Bayley, Bevitori, and Zoni, 2004; Bayley and San Vicente, 2004; Bevitori, 2004; Calzada Pérez, 2017; Calzada Pérez, 2017; Calzada Pérez, 2020; Dibattista, 2004; Garzone and Santulli, 2004; Vasta, 2004). This is partly because CADS depends on corpora, and researchers may find compilation and annotation somewhat cumbersome.

To aid experts in examining parliamentary discourse, in 2020, CLARIN vouched for the scholarly initiative led by Tomaž Erjavec, Maciej Ogrodniczuk and Petya Osenova, resulting in the ParlaMint project<sup>1</sup>. As stated on their website, at the time, ParlaMint-managed to muster the efforts of at least 17 groups of scholars, which compiled parliamentary corpora with debates from 2015 to 2021 from 17 different

1 <https://www.clarin.eu/parlamint>

countries, such as the British House of Commons and the Spanish Congreso de los Diputados (the two corpora analyzed in this article). Erjavec et al. (2023) describe the project’s rationale, the compilation and annotation stages, and the resulting corpora, which are “uniformly encoded, contain rich meta-data about 11 thousand speakers, and are linguistically annotated following the Universal Dependencies formalism and with named entities.” (Erjavec et al. 2023, 415). At this stage, the totality of the 17 corpora amounted to almost half a billion words, and each of them was split into two specific subcorpora: a reference compilation (with texts from 2015 to 30<sup>th</sup> January 2020) and a Covid-19 corpus (with texts from 31<sup>st</sup> January 2020). Covid-19 is, as seems clear, a focal point for project researchers.

A Parlament-II phase followed with data from 2022 and 2023. Subsequent phases are foreseen because scholars such as “sociologists are predominantly interested in current events, which means that it is of crucial importance for ParlaMint corpora to be updated on a regular basis” (Skubic and Fišer 2022, 89). ParlaMint-II has enlarged the time span of existing corpora, added new parliaments (there are now 29 parliaments from different countries and regions), upgraded the mark-up and annotation guidelines, tagged new metadata, and improved a common (Github-based) workflow. In practice, versions 3.0 and 4.0 were released in 2023, with yet another subcorpus under the label of “war”. Thus, ParlaMint II adds another focal point of analysis: parliamentary texts around the Russia – Ukraine war.

As a result, the ParlaMint constellation is a robust tool to look into parliamentary discourse from a (quantitative) corpus-driven standpoint or to back up (qualitative) discourse studies. It may add to the complexity of the field since experts can now dissect texts and contexts according to a range of parameters: speakers, affiliations, positions, and gender, among others. Most importantly, it is a powerful artefact to aid researchers in their comparative and chronological studies. Thanks to compilation and annotation uniformity, comparability and interoperability, it is now possible to go beyond the national level and contrast results between and among different parliaments. It is also possible to carry out Modern-Diachronic Corpus Discourse Studies (following Partington, Duguid, and Taylor, 2013).

Against this background, the present article looks into parliamentary discourse from a CADS perspective. After ParlaMint II, attention is devoted to the way Russia and Ukraine are represented in two of their (2015-2022) full corpora: ParlaMint-GB (with interventions from the British House of Commons) and ParlaMint-ES (with interventions from the Spanish Congreso de los Diputados). In other words, in this article, the quantity afforded by corpus linguistics is nuanced by the quality of discourse

studies. The former provides data and the notion of *lexical priming*. The latter contributes with the discourse historical approach (Wodak and Meyer, 2009). All this is explained further right below.

## 2. Priming Theory and Discourse Historical Approach

## 2.1 Priming Theory with Collocations

Corpus Linguistics is not just the source of quantitative data and corpus-based or corpus-driven methods (see McEnery and Hardie, 2012). It is also the realm that has seen the emergence of linguistic theories, among which *Priming*, it may be argued, is its most decisive one.

Priming theory is the work of Michael Hoey (2005: 8) (2005: 8), for whom “[a]s a word is acquired through encounters with it in speech and writing, it becomes cumulatively loaded with the contexts and co-texts in which it is encountered.”

In priming theory, concordances and collocations play an essential role. Concordances (also known as keywords in context, KWIC) are lines of text around a certain node, like the example below:

nes to Russia and its invasion of **Ukraine**. That is why we have brought it

Figure 1: a concordance line

Hoey (2013, 155) implies that, on the one hand, "the brain must be storing language in a manner analogous to (though obviously not identical to) the way a concordance represents language" and, on the other hand, that:

"when we encounter language we store it much as we receive it, at least some of the time, and that repeated encounters with a word (or syllable or group of words) in a particular textual and social context, and in association with a particular genre or domain, prime us to associate that word (or syllable or group of words) with that context and that genre or domain." (Hoey 2013: 155)

McEnery and Hardie (2012: 123) define a collocation as “a co-occurrence pattern that exists between two items that frequently occur in proximity to one another – but not necessarily adjacently or, indeed, in any fixed order”. It is these surrounding patterns (i.e. the co-text) that end up transferring a great deal of meaning to the central “node” in context.

Collocations are built upon concordances, which means that scholars must generate concordances first and then identify collocations, either manually (by counting and listing the words around the node) or automatically, using statistics measures. Some important measures for collocation generation are logDice, MI, MI3, T-score, Z-score, etc. For a particularly clear, in-depth explanation of corpus statistics, see Brezina (2018).

There are at least two ways scholars may examine cumulative exposure (hence lexical priming potential) to repetitive contextual and cotextual patterns (such as collocation): (a) by focusing on the primed items ("for example [...] all the lexical

primings associated with the word consequence," Hoey 2005: 14); and (b) by identifying relationship among lexical primings ("all the primings that contribute to the production of a sentence," Hoey 2005: 14). Calzada Pérez (2017) mentions a third path that Hoey seems to have overlooked: that of the prime *per se* (such as the word "consequence" in our previous example). Nevertheless, regardless of how collocation is approached, it is a prominent gateway into lexical priming. The present paper opts for the first alternative and proposes a collocational analysis as the bulk of its quantitative examination.

## 2.2 Discourse Historical Approach

The present study opts for the discourse historical approach (DHA), which, in principle, advocates a top-down analysis that starts with an exhaustive ethnographic examination of the historical and generic contexts in which the texts under discussion are produced.

Then, researchers turn to the actual texts and move from means and forms of realization through strategies to content, which they see as closely associated with the context already studied (Wodak et al., 1999: 36–42). Contents, strategies and means are three analytical dimensions that are 'closely interwoven' (Wodak et al., 1999: 30) and are particularly relevant to my work here. The content dimension is straightforward, pointing to the thematic areas of the objects of study. Means and forms of realization are also easy to comprehend since they refer to the different linguistic features (or textural traits) that make up texts. In fact, in the present article, means and forms are the collocational patterns surrounding the central nodes under study.

DHA's strategies, however, require further explanation and may be classified under several labels. For the purposes of this study, I highlight the operationality of the following two for this research:

1. Nomination: 'discursive construction of social actors, objects/phenomena/events and processes/actions' (Wodak and Meyer, 2009: 94). This strategy seems to take place within the area covered by Halliday's (1985) ideational meaning and, more specifically, concerning participants and processes. It is prominent in the present study.

2. Predication: 'discursive qualification of social actors, objects, phenomena, events/processes and actions (more or less positively or negatively)' (Wodak & Meyer, 2009: 94). Adjectives and other modifiers (such as appositions, relative clauses, and prepositional phrases) are the means to convey this strategy. Predication is also in the chambers under study, though not as frequently as the previous strategy.

In sum, this article proposes a DHA-inspired examination as part of the qualitative analysis. Nevertheless, here, the order of analysis is reversed

and proceeds from means and forms (in our case, collocations) to content through strategies. At the same time, and due to space constraints, the content–context connection is kept to the minimum and is left for further research.

## 3. Methodology

In agreement with the great interest ParlaMint assigns to the Russia-Ukraine war, this article aims to identify collocations associated with the main central nodes of "Russia"/"Ukraine" (in English) and "Rusia"/"Ucrania" (in Spanish) within the British and Spanish Chambers. This war is not solely of academic interest for ParlaMint but is one of the hottest issues in today's global world, attracting attention from an ample range of media and (economic, cultural, and societal) circles. In a way, it might be argued that it is one of those historical events that determine the standpoint of societies as a whole and individuals in particular.

To fulfil our goal, we queried ParlaMint-GB v.4.0 (with 2015 to 2022 interventions from Britain's lower chamber – the House of Commons) and ParlaMint-ES v.4.0 (with 2015 to 2022 interventions from Spain's lower chamber – the Congreso de los Diputados)(Erjavec, Kopp and Ogrodniczuk, et al.)<sup>2</sup>. As per lexical priming, we did this to discuss the cumulative meaning that is transferred from the context to the nodes in these parliamentary settings. Notice that we examined both ParlaMint-GB and ParlaMint-ES corpora in full in search of collocations rather than focus on the "war" subcorpus (containing material from 24<sup>th</sup> February 2022). This decision is explained by the fact that there has been a non-stop military conflict between the two countries from 12<sup>th</sup> April 2014 (with the war in Donbas) to now, hence almost perfectly overlapping ParlaMint's time span (2015–2022).

Collocations are generated with NoSketch Engine<sup>3</sup>, a free concordancer prepared to query all ParlaMint corpora in a comparable fashion. Three measures were used for collocation generation: LogDice, MI, and T-score. Collocations will be sorted in descending LogDice order. Using these three measures is not only a NoSketch Engine default but also a technique to find a suitable combination of frequent and strong collocations. Following Brezina (2018: 74), statistical details about collocation generation may be found in Table 1:

Statistics name	L and R span	Minimum Collocate Frequency (NC)	Filter
Log Dice MI T-score	-5 +5	5	lemma

Table 1: Collocation statistics.

In brief, the stages of analysis were as follows:

2 <http://hdl.handle.net/11356/1860>

3 <https://www.clarin.si/ske/#open>

- Computerized identification of all concordances of “Russia”/“Rusia”, “Ukraine”/“Ucrania” lemmas in ParlaMint-GB and ParlaMint-ES.
- Computerized generation of collocations of “Russia”/“Rusia”, “Ukraine”/“Ucrania” lemmas in ParlaMint-GB and ParlaMint-ES.
- Selection of the top 50 collocations of both “Russia”/“Rusia” and “Ukraine”/“Ucrania” lemmas in ParlaMint-GB and ParlaMint-ES.
- Quantitative discussion of top 50 collocations of both “Russia”/“Rusia” and “Ukraine”/“Ucrania” lemmas in ParlaMint-GB and ParlaMint-ES with special reference to implications drawn from lexical priming theory.
- Qualitative discussion of the top 50 collocations of both “Russia”/“Rusia” and “Ukraine”/“Ucrania” lemmas in ParlaMint-GB and ParlaMint-ES with special reference to implications drawn from DHA.

Notice that space constraints limit the extension and depth of the analysis described here. This is why only 50 collocations are considered. Further studies will go beyond the conclusions drawn here.

#### 4. Analysis. Russia, Ukraine; Ucrania, Rusia: Same Difference in Britain and Spain?

##### 4.1 Preliminary Data

First, below are some of the most basic quantitative data regarding the full size of both ParlaMint-GB and ParlaMint-ES.

ParlaMint-GB: includes speeches from the House of Commons (and House of Lords) from 2015-2022 (see Table 2).

Tokens	139,686,402
Words	124,744,599
Sentences	5,323,032
Paragraphs	1,406,962
Documents	670,912

Table 2: ParlaMint-GB in figures.

We have just used speeches from the House of Commons – Britain’s lower chamber – for collocation generation to make material comparable.

ParlaMint-ES, in full, contains speeches from Spain’s lower chamber – the Spanish Congreso de los Diputados – from January 2015 to 23<sup>rd</sup> February 2023 (see Table 3).

Tokens	22,118,291
Words	19,423,835
Sentences	770,424

Paragraphs	243,994
Documents	76,351

Table 3: ParlaMint-ES in figures.

We have only queried interventions from 2015 to 2022 for collocation generation to make material comparable.

Table 4 contains data about collocation generation in ParlaMint-GB:

	Lemma: Russia	Lemma: Ukraine
Number of collocations	1893	1514
Number of hits for lemma	7328	6091
Number of node/lemma hits per million	52.46	43.6
Percent of the corpus	0.005246 %	0.004360 %
Corpus size	139,686,402	139,686,402

Table 4: Collocations in ParlaMint-GB.

Table 5 contains data about collocation generation in ParlaMint-ES.

	Lemma: Rusia	Lemma: Ucrania
Number of collocates	192	399
Concordance size (number of lemma hits)	444	1181
Number of node/lemma hits per million	20.07	53.39
Percent of the whole corpus	0.002007%	0.005339%
Corpus size	22,118,291	22,118,291

Table 5: Collocations in ParlaMint-ES.

##### 4.2 Quantitative Analysis and Priming Theory

Tables 2-5 show that the ParlaMint-GB corpus (124,744,599 words) is much larger than the ParlaMint-ES corpus (19,423,835). This size divergence is due to the fact that sessions convened in the House of Commons are much more frequent and longer than those in the Congreso de los Diputados. In effect, this means that members of the British parliament are exposed to a greater amount of linguistic data than their Spanish counterparts in general. Lexical priming inputs are bound to be greater in the former than in the latter.

When analyzing collocations, and precisely because of the difference in the size of corpora, we must now refer to comparable figures – those pointing at the

number of times that nodes appear per million words or pmw. Otherwise, corpora cannot be compared on equal terms. In this case, British MPs are exposed to a greater amount of the “Russia” node (52.46 pmw) than Spanish MPs (20.07 pmw). This cumulative exposure to references implies that British MPs are bound to have a stronger (more vivid, more linguistically informed, more ingrained by frequency) image of “Russia” than the Spanish MPs. If we turn to Ukraine, we realize the situation is very different. British MPs are comparably less exposed to the “Ukraine” node (46.3 pmw) than Spanish deputies (53.39 pmw). On this occasion, the latter receive more cumulative exposure and are bound to have more ingrained perceptions in their minds.

Notice also that British parliamentarians are more exposed to mentions of “Russia” (52.46 pmw) than “Ukraine” (43.6 pmw). The difference is 8.86 points. Apart from the fact that it is quite the opposite in the Spanish Parliament, the cumulative exposure to the “Ucrania” node (53.39 pmw) more than doubles the exposure to the “Rusia” node (20.07 pmw). The gap in exposure between the two nodes in the Spanish chamber (33.32 pmw) is, thus, especially wide (in statistics, this is measured via effect size measures such as LogR: 1.91) and statistically significant (LL: 15.11;  $p<0.001$ ) vis-à-vis what happens in the House of Commons.<sup>4</sup>

When focusing on the number of collocates that accompany and prime the nodes, higher figures are observed in ParlaMint-GB than in ParlaMint-ES. The British chamber has 1893 collocates for “Russia” and 1514 for “Ukraine”. As is clear, the raw variety of potential lexical priming transfer is larger for the first node than for the second. This difference is statistically significant (LL:174.14;  $p<0.0001$ ). However, the size of this raw difference (known in statistics as effect size) is virtually non-existent (LogR:0.34). On the contrary, in the Congreso de los Diputados, potential lexical priming is more intense for “Ucrania” (with 399 different collocates) than for “Rusia” (199). In this case, the collocates of “Ucrania” are more than double those of “Rusia”. Resorting to statistics again, this difference is significant (LL:74.06;  $p<0.0001$ ) and with a large effect size (LogR: 1.06).

In sum, when it comes to “Russia/Rusia” and “Ukraine/Ucrania”, linguistic behavior quantitatively differs in both the British and Spanish chambers not only in the amount of exposure to the nodes pmw but also in the range size of collocates that are bound to impregnate these nodes. If we go beyond raw data and examine the statistics, the collocate span (or range size) difference is especially heterogeneous in the Spanish Chamber. This difference is statistically significant (LL:174.14;  $p<0.0001$ ) and of a great effect size (LogR: 1.06).

At this point, only hypotheses are possible. The wide gap detected between the nodes in the Spanish house and its greater heterogeneity in collocates

show less convergence in this chamber than in the British house. This recalls prior research (Calzada Pérez, 2023), which discusses other cases where the Spanish Congreso de los Diputados is shown to be more heterogeneous (and prone to contextual events) than the British House of Commons. Something like this may be happening here. Also worth noting is that the main node interest shifts from “Russia” (in ParlaMint-GB) to “Ucrania” (in ParlaMint-ES). The different mention of “Ukraine” (in ParlaMint-GB) and “Ucrania” (in ParlaMint-ES) is larger (heading towards twice the amount of difference with a LogR of 0.73) and undoubtedly significant (LL: 73.91).

### 4.3 Qualitative Analysis and DHA

For a qualitative analysis of collocations, we have to go beyond figures and examine them in a rather more manual fashion. Indeed, this has advantages as it allows researchers to go deeper into lexical priming (or potential meaning transfer from the collocates to the node). However, the main disadvantage of any manual work is that we need to downsize linguistic samples. For instance, it would be difficult for scholars to focus on 1893 different ParlaMint-GB collocates of “Russia”. It would be even less feasible to report on this extensive work in an article with the space limitations of the present one. This is why this section reports on the top 50 collocations of “Russia” and “Ukraine” (from ParlaMint-GB) and “Rusia” and “Ucrania” (from ParlaMint-ES). These collocations are grouped in Tables 9 and 10.

These tables arrange collocates in three categories for each “Russia”/“Ukraine” node: (a) common collocates for both nodes; (b) common collocates which appear in the top 50 rank in the case of one of the nodes but not the other; and (c) specific collocates for each node.

For example, “invasion” is a top 50 collocate of (and primes) both “Russia” and “Ukraine” in ParlaMint-GB, as seen in Table 6.

Node	Freq	Coll. freq.	T-score	MI	logDice
Russia	126	1742	11.21683	10.42916	8.83039
Ukraine	360	1742	18.96966	12.21048	10.5565

Tabla 6: “Invasion” as a collocate of “Russia” and “Ukraine.”

The term “China” is a top 50 collocate of “Russia” but appears in position 506 as a collocate of “Ukraine”. See statistics in Table 7.

Node	Freq	Coll. freq.	T-score	MI	logDice
Russia	346	10922	18.57027	9.23809	9.27902
Ukraine	8	10922	2.66005	4.0702	3.94565

Table 7: “China”, as collocate of “Russia” and “Ukraine.”

4 Statistics data are calculated using the <https://ucrel.lancs.ac.uk/lwizard.html>.

Finally, Table 8 shows specific collocates of “Russia” and “Ukraine.”

Coll.	Node	Fq	Coll. fq.	T-score	MI	logDice
Assad	Russia	346	109227	18.5702	9.23809	9.27902
Zelensky	Ukraine	8	10922	2.66005	4.0702	3.94565

Table 8: Specific collocates of “Russia” and “Ukraine.”

Table 9 registers the top 50 collocations of ParlaMint\_GB. Notice that collocates are sorted according to Log-Dice (the higher the Log-Dice, the higher the word appears in the table).

Collocation Type	Russia Collocations	Ukraine Collocations
Common	Ukraine Putin invasion Crimea aggression invade Russia NATO illegal Russian Belarus war eastern ally attack condemn President Military weapon incursion	invasion Russia Putin russian eastern aggression war invade Crimea incursion NATO military Ukraine illegal weapon President ally attack condemn Belarus
Common but in different ranks	China sanction Today Iran threat pose Syria annexation regime Security States against influence Turkey US gas action behaviour intelligence pressure annex India disinformation Germany	territorial Sovereignty integrity defend Georgia troops unprovoked Poland sovereign border humanitarian conflict brutal situation crisis scheme Russians stand brave
Totally Specific	Assad veto	Homes flee

resurgent Sanctions Korea Brazil	solidarity Ukrainian grain lethal refugee visa Zelensky Ukrainians aid
---	--

Table 9: Collocations in ParlaMint-GB.

Table 9 may be analyzed in line with DHA methodology: moving from means through strategies to content. It shows that, in ParlaMint-GB, “Russia” and “Ukraine” share 20 collocates within the top 50 rank. Most contribute to nomination strategies, which characterize participants, processes, and objects. See in alphabetical order:

- States or institutions: “Belarus”, “Crimea”, “Russia”, “Ukraine”, and “NATO”.
- Human participants: “ally” and “Putin.”
- Phenomena and processes: “aggression”, “attack”, “condemn”, “incursion”, “invade”, “invasion”, “military”, “war”, “weapon.”

By way of illustration (and for reasons of space), here are only two examples of common collocates of “Russia” and “Ukraine”.

- Even in Russia, Putin’s invasion is now having disastrous consequences. (HC20220616)
- As we have heard today, the destabilization resulting from Putin’s invasion of Ukraine continues, bringing with it humanitarian crises that go way beyond the region in which we see military action. (HC20220721)

There are also 3 predication-related collocates through which participants, processes and objects are characterized: “eastern”, “illegal”, and “Russian”. Below are some examples of “eastern”:

- Although it is important that we take Russian security concerns seriously, we must resist at all costs any attempts by Russia to re-imperialize eastern Europe. (HC20220117)
- The war in eastern Ukraine drags on; the Nord Stream pipeline has been shut down; flights are being cancelled left, right and centre; and Britain is facing an unprecedented heat wave as our climate changes in front of our very eyes. (HC20220718)

In short, common collocates tend to be directly associated with the war, as is particularly clear when the focus is set on phenomena and processes, which almost all are (near) synonyms or may be placed in the same semantic realm: aggression, attack, etc. Thus, through common collocates, “Russia” and “Ukraine” are primed to be understood as contenders in the military conflict.

Many conclusions emerge when the eyes are turned to those common collocates spaced out in the

ranking list. Two are especially relevant for the present paper. While “Russia” is primed by its connections to other countries, some of which are not necessarily allies of Great Britain (“China”, “Iran”, “Syria”, “Turkey”), “Ukraine” is particularly predicated with evaluative adjectives such as “brutal”, “brave”, “humanitarian”, “unprovoked”, resorting to a more affective discourse that places the node in a more friendly position.

- There is no doubt that revanchist Russia and Iran have grown closer under Putin’s leadership. (HC20220630)
- Putin’s war on Ukraine is brutal, illegal and a calculated attack on peace and stability in Europe. (HC20220224)

Though a handful, specific collocates portray a different image of both nodes. “Russia” is linked to what seems to be a lexical priming trend, through which it is connected to allies such as “[Bashar Al-] Assad”, “Korea”, or “Brazil”, in an “othering” technique, which ends up separating Russia from the West, in general, and Britain, in particular. In the meantime, “Ukraine” is primed in the opposite direction, and a different trend (among others) is spotted. This trend (see Table 9 above) shows the node as associated with Ukrainian refugees that flee from a lethal war and receive Britain’s aid and solidarity through the concession of visa(s) and the application of the Homes for Ukrainian scheme. As in the following example:

- This is a whole Government effort, as well as a UK-wide effort to support families and the Homes for Ukraine scheme. (HC20220620)

For its part, Table 10 registers the top 50 collocations of ParlaMint\_ES. Again, notice that collocates are sorted according to Log-Dice (the higher the Log-Dice, the higher the word appears in the table).

Collocation Type	Rusia Collocations	Ucrania Collocations
Common	Ucrania invasión invadir Rusia agresión Putin guerra OTAN provocado ucraniano conflicto frontera ataque Europa	invasión guerra Rusia Putin invadir agresión provocado conflicto ataque Ucrania ucraniano frontera Europa OTAN
Common but in diffent ranks	Gas amenaza tensión Estados depender	Consecuencia

	parte importar rechazar relación Unión afectar solamente Europea impacto	
Totally Specific	China sanción proveedor exportación Crimea India Turquía exportador dependencia pétroleo suministro agresor dependiente carbón procedente comprar natural Unidos energético 2020 demanda convertir	ruso tropa derivado enviar arma Moldavia bético envío Georgia agravado RUSA militar pueblo refugiado armamento material desestabilización Palestina primo integridad defensivo paz crisis resistencia Embajada Bielorrusia liberado agravar desplazado ayudar criminal terrible Taiwán brutal Minsk

Table 10: Collocations in ParlaMint-GB.

As Table 10 shows, ParlaMint-ES projects a very different image of the nodes. On this occasion, what is particularly striking is that there are many more specific collocates for each node (22 for “Russia” and 35 for “Ukraine”). Hence, while in the House of Commons (overlapping or spaced out) similarities are “the norm” when referring to node collocates, in the Congreso de los Diputados specificities dominate. Now, the number of collocations for each node differs strikingly, and the nature of such collocations is also idiosyncratic. This reaffirms the intuition/ hypothesis/ previous results that suggest that the House of Commons is more homogenous and stable than the Congreso de los Diputados.

The nodes “Rusia” and “Ucrania” share 14 collocates within the top 50 rank. Most contribute to nomination strategies, through which participants, processes

and objects are characterized. Among them (in alphabetical order):

- States or institutions: “Rusia” [Rusia], “Ucrania” [Ukraine], “Europa” [Europe], “OTAN” [NATO],
- Human participants: “Putin”
- Phenomena and processes: “agredir” [to carry out aggression], “ataque” [attack], “agresión” [aggression], “conflict” [conflict], “frontera” [border], invadir [“invade”], invasion [“invasion”], “guerra” [“war”], “weapon”.

By way of illustration (and for reasons of space), here are only two examples of common collocates of “Rusia” and “Ucrania”.

- El empobrecimiento de Ucrania, Europa y Rusia será la consecuencia de estas sanciones, como la propia Unión Europea ya está advirtiendo. (CD20220302)
- Hoy, en España y en Europa sufrimos economía de guerra porque a España y a Europa la guerra de Ucrania no nos es ajena. (CD20220309)

There is only 1 common predication-related collocate (the lemma “provocado” [“provoked”]), pointing at the reasons for the conflict. The examples below represent this predication: responsibility is assigned to Russia, while Ukraine is portrayed as the invasion victim. Alternatively, Spain is also seen as suffering the consequences of the war.

- El trasfondo de la subida de precios de la energía hay que buscarlo en la situación provocada de manera intencionada por Rusia para tensionar los mercados del gas y de la electricidad en la Unión Europea, con el único objetivo, señorías, de minar la recuperación económica europea. (20220316)
- Entonces reparé en el añadido del enunciado en el orden del día: para informar sobre las medidas económicas y sociales adoptadas por el Gobierno para dar respuesta a la crisis provocada por la guerra en Ucrania. (CD20221013)

The number of common collocates that rank far apart in the collocational list is now less frequent than in the case of the House of Commons. Space constraints lead us to mention this category in passing, pointing out that in the case of “Ucrania”, a particularly strong collocate is “consecuencias” [consequences].

Node	Freq	Coll. freq.	T-score	MI	logDice
Rusia	6	9308	2.37321	5.00503	4.33348
Ucrania	37	9308	6.00106	6.21814	6.85286

Table 11: “Consecuencia” as lemma collocate in ParlaMint-ES.

Like with the lemma “provocado”, the way “consecuencia” is used with “Ucrania” suggests that MPs are concerned about the impact of the war (not only on Ukraine itself) but also (especially?) on Spain.

- En definitiva, financiar políticas públicas para hacer frente a las consecuencias de la guerra de Ucrania y lograr un pacto de rentas. (CD20220913)

Specific collocates now abound and portray very different images of both nodes. With its foes and friends, “Russia” is primed as a major world economy, with a great potential impact upon the West. See the clearest collocational trend below:

- Geopolitical spaces: “Crimea”, “India”, “Turquía” [Turkey]
- Economic terms: “carbón” [coal], “comprar” [buy], “demanda” [demand], “dependencia” [dependency], “dependiente” [dependent], “energético” [energy], “exportación” [exports], “exportador” [exporter], “(gas) natural” [natural (gas)], “suministro” [supply], “petróleo” [oil], “proveedor” [supplier].

The following example provides an illustration:

- Usted ha convertido a Rusia en el tercer proveedor de gas en España. (CD20221221)

Through specific collocates, Ukraine, in turn, is reduced to its military role and linked to other, very concrete, geopolitical world regions with which the country is identified (in Spain). See the main trend below:

- Nomination collocates highlighting Ukraine’s role as war participant: “arma” [arm], “armamento” [weaponry], “desestabilización” [destabilization], “desplazado” [displaced], “liberado” [liberated], “militar” [military], “paz” [peace], “refugiado” [refugee], “resistencia” [resistance], “tropa” [troop].
- Adjectival collocates with an affective value: “brutal” [brutal], “defensivo” [defensive], “liberado” [liberated], and “terrible” [terrible].
- Nomination strategies placing Ukraine in relation to friends and enemies: “Bielorrusia” [Belarus], “Palestina” [Palestine], “Taiwán” [Taiwan].

In sum, through especially nomination and predication strategies, in ParlaMint-GB, “Russia” and “Ukraine” are associated with the war through common (semi-)common, and specific collocates. However, ParlaMint-ES has a very different portrayal of “Rusia” and “Ucrania.” The former is depicted as an important economic competitor, transcending its participation in the conflict. Othering strategies are spotted in the analysis (by association with allies that are enemies or adversaries of the West). The latter is reduced to its role as the invasion victim, and the Congreso de los Diputados takes sides with it not just through affective predication but also by sharing the consequences of such an invasion with Ucrania.

## 5. Conclusions

This paper examines British and Spanish parliamentary discourse around the nodes “Russia” and “Ukraine” (in English) and “Rusia” and “Ucrania” (in Spanish). To do so, quantitative CADS is complemented by qualitative DHA. The results of the

combination are certainly illuminating. Furthermore, CLARIN ParlaMint-GB and ParlaMint-ES are queried with free NoSketch Engine. After this study, it is advocated here that the ParlaMint constellation is a powerful tool for research into parliamentary discourse.

Concerning quantity, (some of) the lexical priming potential of both parliamentary chambers is revealed in the analysis. Quantitative raw data suggests that British MPs are more exposed to nodes and collocations. However, when looking into statistics, British deputies are seen to be particularly primed to the node “Russia”. By contrast, their Spanish counterparts show greater interest in “Ucrania”. The gap between exposure to both nodes is particularly wide in the Spanish Congreso de los Diputados, where “Ucrania” has double the number of hits than “Rusia”. In fact, this gap difference (or effect size) between the two chambers is large enough to be mentioned here and statistically significant. Also, the range of collocates is particularly heterogeneous (with greater effect sizes) and significant in the Spanish Congreso de los Diputados. According to prior studies (Calzada Pérez, 2023), heterogeneity is a “common” feature in the Spanish Parliament and often suggests that this chamber is more exposed to context than its British equivalent. This result adds to the conclusion drawn in past studies. Yet further research is required.

Regarding qualitative results, the nature of MPs lexical priming to nodes and (common, quasi-common and specific) collocates of “Russia”/“Rusia” and “Ukraine”/“Ucrania” differs in ParlaMint-GB and ParlaMint-ES drafting two different profiles for the nodes. British MPs are primed to see “Russia” and “Ukraine” in again a more homogenous manner, as participants in a war. Spanish MPs boost “Ukraine”’s victim role and sympathize with it. In the Congreso de los Diputados, Russia is seen as a (economic and fighting) contender whose activity may have “terrible” “consequences” (to use some of the collocates discussed above) not just for “Ucrania” but also for Spain and its allies.

## 6. Acknowledgments

This research was carried out within the framework of the project *Original, translated and interpreted representations of the refugee crisis: methodological triangulation within corpus-based discourse studies* (PID2019-108866RB-I00 / AEI / 10.13039/501100011033), funded by the Spanish Ministry of Science and Innovation.

ParlaMint-ES v. 4.0 has been compiled and marked up by María Calzada Pérez, Rubén de Líbano and Monica Albini with funding from CLARIN. Annotation and indispensable guidance and aid are due to Tomaž Erjavec and Matyáš Kopp.

Special thanks are due to Tomaž Erjavec, Maciej Ogodniczuk and Petya Osenova for their inspiration and most efficient guidance within the ParlaMint project.

Carmen Pérez Basanta read the first draft of this article and provided most valuable insights.

## 7. Bibliographical References

Baker, P. (2006). *Using Corpora in Discourse Analysis*. London; New York: Continuum.

———. (2010). *Sociolinguistics and Corpus Linguistics*. Edinburgh: Edinburgh University Press.

Bayley, P., Bevitori, C. and Zoni, E. (2004). Threat and Fear in Parliamentary Debates in Britain, Germany and Italy. In P. Bayley (Ed.), *Cross-Cultural Perspectives on Parliamentary Discourse*. Amsterdam; Philadelphia: John Benjamins, pp. 185–236.

Bayley, P. and San Vicente, F. (2004). Ways of Talking about Work in Parliamentary Discourse in Britain and Spain’. In P. Bayley (Ed.), *Cross-Cultural Perspectives on Parliamentary Discourse*. Amsterdam; Philadelphia: John Benjamins, pp. 237–69.

Bevitori, C. (2004). Negotiating Conflict: Interruptions in British and Italian Parliamentary Debates. In P. Bayley (Ed.), *Cross-Cultural Perspectives on Parliamentary Discourse*. Amsterdam; Philadelphia: John Benjamins, pp. 87–109.

Box-Steffensmeier, J.M., Brady, H.E. Collier, D. (Eds). (2008). *The Oxford Handbook of Political Methodology*. Oxford; New York: Oxford University Press.

Brezina, V. (2018). *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge: University Printing House [Kindle version].

Bütikofer, S. and Hug, S. (2015). Strategic Behaviour in Parliament. *The Journal of Legislative Studies*, 21(3): 295–322.

Calzada Pérez, M. (2017). Corpus-Based Methods for Comparative Translation and Interpreting Studies. *Translation and Interpreting Studies*, 12(2): 231–252.

———. (2017). Five Turns of the Screw: A CADS Analysis of the European Parliament. *Journal of Language and Politics*, 16(3): 412–33.

———. (2018). Researching the European Parliament with Corpus-Assisted Discourse Studies: From the Micro- and Macro-Levels of Text to the Macro-Context. *Revista Española de Lingüística Aplicada/ Spanish Journal of Applied Linguistics*, 30(2): 465–90.

———. (2020). A Corpus-Assisted SFL Approach to Individuation in the European Parliament: The Case of Sánchez Presedo’s Original and Translated Repertoires. *Meta: Journal Des Traducteurs/Meta: Translators’ Journal*, 65(1): 142–67.

———. (2023). The Representation of Migration in Parliamentary Settings: Critical Cross-Linguistics Corpus-Assisted Discourse Analyses. *Humanities and Social Sciences Communications*, 10(1): online. <https://doi.org/10.1057/s41599-023-02496-y>.

Dibattista, D. (2004). Legitimizing and Informative Discourse in the Kosovo Debates in the British House of Commons and the Italian Chamber of Deputies. In P. Bayley (Ed.), *Cross-Cultural Perspectives on Parliamentary Discourse*. Amsterdam; Philadelphia: John Benjamins, pp. 151–84.

Erjavec, T., Ogrodniczuk, M., Osenova, P. et al. (2023). The ParlaMint Corpora of Parliamentary Proceedings. *Language Resources and Evaluation*, 57(1): 415–48.

Garzone, G. and Santulli, F. (2004). What Can Corpus Do for Critical Discourse Analysis? In A. Partington, J. Morley and L. Haarman (Eds.), *Corpora and Discourse*, Bern: Peter Lang, pp. 351–68.

Hix, S., Noury, A. and Roland, G. (2006). Dimensions of Politics in the European Parliament. *American Journal of Political Science*, 50(2): 494–520.

Hoey, M. (2005). *Lexical Priming: A New Theory of Words and Language*. London; New York: Routledge.

———. 2013. Lexical Priming and Translation. In A. Kruger, K. Wallmach and J. Munday *Corpus-Based Translation Studies: Research and Applications*, London: Continuum, pp. 153–68.

McEnery, T. and Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge: CUP.

Partington, A. (2004). Corpora and Discourse: A Most Congruous Beast. In A. Partington, J. Morley and L. Haarman (Eds.), *Corpora and Discourse* Bern: Peter Lang, pp. 11–20.

Partington, A., Duguid, A. and Taylor, C. (Eds.), (2013). *Patterns and Meanings in Discourse: Theory and Practice in Corpus-Assisted Discourse Studies (CADS)*. Amsterdam: John Benjamins.

Skubic, J. and Fišer, D. (2022). Parliamentary Discourse Research in Sociology: Literature Review. In *Proceedings of the Workshop ParlaCLARIN III within the 13th Language Resources and Evaluation Conference*, edited by D. Fišer, M. Eskevich, J. Lenardic and F. de Jong, pages 81–91. Marseille, France: European Language Resources Association.

Van Eemeren, F.H. and Houtlosser, P. (2006). The Case of Pragma-Dialectics. In S. Parsons, N. Maudet, P. Moraitis, and I. Rahwan (Eds.), *Argumentation in Multi-Agent Systems*, Berlin; Heidelberg: Springer, pp. 1–28.

Vasta, N. (2004). Consent and Dissent in British and Italian Parliamentary Debates on the 1998 Gulf Crisis. In P. Bayley (Ed.), *Cross-Cultural Perspectives on Parliamentary Discourse*. Amsterdam; Philadelphia: John Benjamins, pp. 111–49.

Wodak, R., De Cillia, R., Reisigl, M. and Liebhart, K. (1999). *The Discursive Construction of National Identity*. Edinburgh: Edinburgh University Press.

Wodak, R. and Meyer, M. (2009). *Methods for Critical Discourse Analysis*. London; Thousand Oaks California: SAGE Publications Ltd.

## 8. Language Resource References

Erjavec, T., Kopp, M. and Ogrodniczuk, M., et al. (2023) Linguistically annotated multilingual comparable corpora of parliamentary debates ParlaMint.ana 4.0, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, <http://hdl.handle.net/11356/1860>.