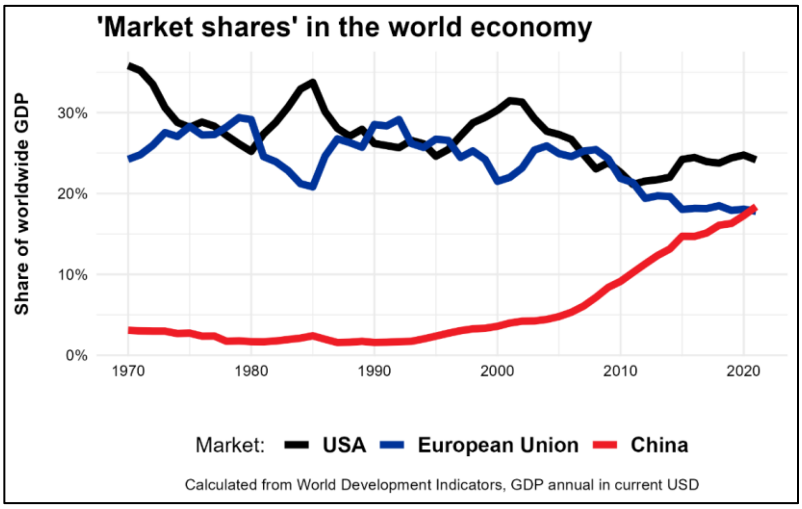
International recognition of European Union 'actorness': Language-based evidence from United Nations General Assembly speeches 1970-2020

Online Appendices

(Final version June 20, 2025)

A1: 'Market shares' of major economic powers



Calculated from World Development Indicators, GDP annual in current USD

Figure 1: Selected world GDP shares over time

A2: Dictionary of EU references

The following dictionary has been used to identify references to the EU or the EC in UNGD speeches. To ensure flexibility as well as to only match valid abbreviations, I employ regular expression in the flavour of the R programming language (using the stringr package, Wickham 2015):

- (European Union)
- (([^A-Za-z]|^)(EU)([^A-Za-z]|$))
- (([^A-Za-z]|^)(E\\.U\\.)([^A-Za-z]|$))
- (European Communit(y|ies))
- (([^A-Za-z]|^)(EC)([^A-Za-z]|$))
- (([^A-Za-z]|^)(E\\.C\\.)([^A-Za-z]|$))
- (European Economic Communit(y|ies))
- (([^A-Za-z]|^)(EEC)([^A-Za-z]|$))
- (([^A-Za-z]|^)(E\\.E\\.C\\.)([^A-Za-z]|$))

Implementation of the respective coding procedures can be inspected or repeated along the following scripts contained in the replication package:

- 0_DependencyParsing.R
- 3_CodeAgentsFromSemanticMotifs.R
- 7_HumanValidation_Actorness.R
- X_CodeAgentsFromSRL.R

A3: Dictionaries for references to benchmark ROs and IOs

<table>
<tr><td><i>African Union</i></td><td>(African Union)
(([^A-Za-z]|^)(AU)([^A-Za-z]|$))
(([^A-Za-z]|^)(A\\.U\\.)([^A-Za-z]|$))
(Organi(s|z)ation of African Unity)
(([^A-Za-z]|^)(OAU)([^A-Za-z]|$))
(([^A-Za-z]|^)(O\\.A\\.U\\.)([^A-Za-z]|$))</td></tr>
<tr><td><i>Andean Community</i></td><td>(Andean Community)
(([^A-Za-z]|^)(CAN)([^A-Za-z]|$))
(([^A-Za-z]|^)(C\\.A\\.N\\.)([^A-Za-z]|$))
(Andean Pact)"</td></tr>
<tr><td><i>ASEAN</i></td><td>(Association of Southeast Asian Nations)
(([^A-Za-z]|^)(ASEAN)([^A-Za-z]|$))
(([^A-Za-z]|^)(A\\.S\\.E\\.A\\.N\\.)([^A-Za-z]|$))</td></tr>
<tr><td><i>Caribbean Community</i></td><td>(Caribbean Community)
(Caricom)
(CARICOM)
(([^A-Za-z]|^)(CC)([^A-Za-z]|$))
(([^A-Za-z]|^)(C\\.C\\.)([^A-Za-z]|$))</td></tr>
<tr><td><i>OAS</i></td><td>(Organi(z|s)ation of American States)
(([^A-Za-z]|^)(OAS)([^A-Za-z]|$)),
(([^A-Za-z]|^)(O\\.A\\.S\\.)([^A-Za-z]|$))</td></tr>
<tr><td><i>IMF</i></td><td>(International Monetary Fund)
(([^A-Za-z]|^)(IMF)([^A-Za-z]|$))
(([^A-Za-z]|^)(I\\.M\\.F\\.)([^A-Za-z]|$))</td></tr>
<tr><td><i>World Bank</i></td><td>(World Bank)</td></tr>
<tr><td><i>WTO</i></td><td>(World Trade Organi(z|s)ation)
(([^A-Za-z]|^)(WTO)([^A-Za-z]|$))
(([^A-Za-z]|^)(W\\.T\\.O\\.)([^A-Za-z]|$))</td></tr>
<tr><td><i>NATO</i></td><td>(North Atlantic Treaty Organi(z|s)ation)
(([^A-Za-z]|^)(NATO)([^A-Za-z]|$))
(([^A-Za-z]|^)(N\\.A\\.T\\.O\\.)([^A-Za-z]|$))</td></tr>
<tr><td><i>OSCE</i></td><td>(Organi(z|s)ation for Security and Co-operation in Europe)
(([^A-Za-z]|^)(OSCE)([^A-Za-z]|$))
(([^A-Za-z]|^)(O\\.S\\.C\\.E\\.)([^A-Za-z]|$))</td></tr>
</table>

Table 1: Dictionaries for identifying references to different regional and international organisations

A4: Developing and validating a measure for (EU) actorness recognition

The present study requires a scalable measurement of whether international political speeches explicitly recognize 'actorness' of the EU – that is, the 'capacity to behave actively and deliberately in relation to other actors in the international system' (Sjöstedt 1977, 16). This appendix first introduces *four text-as-data approaches to capture EU actorness*, benchmarks them against an original data set of respectively human-coded sentences from UNGD speeches, to ultimately justify the measure used in the main analyses.[1]

The initial and most simple approach to measure actorness of an entity (in our case primarily the EU) in texts is assessing whether it is mentioned at all (*approach 1*). This essentially assumes that any explicit EU reference implies that the respective speaker recognizes the EU as an entity that has a capacity to act. In text-as-data terms, this assumption would allow us sticking to a simple dictionary-based analysis along the different ways the EU can be referred to (cf. appendix A2 above).

However, this approach might be noisy or even biased in the light of our conceptual ambition. Linguistic and sociological discussions of measuring 'agency' in texts (see esp. Franzosi, De Fazio, and Vicari 2012; Knight 2022) clearly defy the idea that the recognition of an entity's mere existence equals acknowledging its capability to actually do something on its own. Qualitatively, this also seems to hold in the current application: cursorily reading random samples of EU references in UNGD speeches quickly shows instances in which the EU is just mentioned as some passive entity. For example, the EU is often just listed as one of many multilateral institutions in existence, speakers only note their states' membership in the EU, or the EU is solely used as a geographical reference point ('compared to the EU average'). While such statements do imply that the EU exists, they do not necessarily suggest that the respective speaker wants or does say that the EU is an

[1] All R scripts, functions, and validation data presented in this appendix are also contained in the replication package provided via https://dvn.iq.harvard.edu/dvn/dv/internationalinteractions. Please see the documentation files and contact the author in case of questions.

entity that can act by itself. Just counting EU references therefore entails a notable risk of 'false positives' in measuring the concept of 'actorness' that we are interested in here.

Therefore, I aim to extract conceptually more specific information from the different ways the EU is referenced in UNGD speeches. The core measurement idea that I pursue is that *a speaker mentioning the EU recognizes its actorness only if and when he or she explicitly links the EU to some kind of past, current, or future action*. Natural language offers a myriad of ways to articulate such entity-action links, but all of them are in principle contained in the syntactic relations encoded in the grammar of a statement.[2] Thus, I extend the presence of an EU reference with additional information on the syntactic role that this entity plays in relation to explicit actions expressed in a piece of text.

For this task, *dependency parsers* are extremely useful (Atteveldt et al. 2017; Stuhler 2022). Exploiting word order and the 'part-of-speech' functions of individual words (such as 'noun', 'verb', 'adjective', etc.), these NLP-tools extract the syntactic structure of sentences by establishing and labelling relationships between 'head' words and the words modifying them – e.g. by indicating that a particular noun acts as the nominal subject of a verb in a sentence.[3] The resulting structure of a sentence is then stored in a respective dependency tree (examples in Figure 2 below). In principle, such syntactic networks encode semantically richer information on the roles that specific entities such as the EU play in any given sentence. What we need are rules to reliably extract instances in which the EU is explicitly linked to any kind of action and test their validity for capturing the concept of actorness in our specific application.

The most straightforward of such syntactic rules to conclude that a sentence implies a recognition of EU actorness is relying on its role as an active or passive subject in that sentence (*approach 2*). If

[2] Note that this idea of relational content analysis has also been very prominent in classical human-coded approaches in the political and communication sciences, such as the core sentence approach (e.g. Kleinnijenhuis, De Ridder, and Rietberg 1997; Kriesi et al. 2006) or claims analysis (Koopmans and Statham 1999).

[3] In the present application, I resort to the parsers of the industry-leading SpaCy python library (Honnibal and Montani 2020) as wrapped in the spacyr R-package (Benoit and Matsuo 2020). For querying and visualising the resulting dependency trees I resort to the semgram (Stuhler 2022) and rsyntax (Welbers and Van Atteveldt 2022) packages. For implementation, see the scripts "2_SemanticMotifExtraction.R" and "1_ExtractActorness.R" in the replication archive.

and when the EU syntactically operates as the subject of a verb, this unequivocally indicates that the sentence states that the EU is or has been doing something (see examples 1 and 2 in Figure 2). To identify such instances in the dependency trees of all UNGD sentences, I resort to the corresponding eight extraction rules on 'action motifs' developed by Oscar Stuhler (2022: see esp. pp. 1597-1601) as implemented in his spearheading *semgram* R package.[4] These extraction rules cover a wide range of instances in which nouns or proper nouns act as subjects of a verb, including auxiliary verbs, direct conjunctions, or passive "by" constructions.

While this approach should significantly reduce the risk of false positives in detecting actorness recognition in comparison to mere mentions of the EU, it entails a risk of false negatives in the case of more complex syntactic patterns in UNGD speeches. For example, cursory reading of example sentences shows that the EU is often presented as acting together with rather large lists of other actors, syntactically creating long conjunction chains that put the EU at quite some distance to the actual verb in the dependency tree of a sentence.

I therefore also test the validity of a semantic role labelling algorithm (*approach 3*). Unlike extracting a set of fixed set of syntactic patterns, such models try to identify the latent predicate-argument structures in sentences along a machine-learning approach which is trained on large collections of linguistically annotated texts (for an encompassing introduction, see Jurafsky and Martin 2009: Chapter 19). SRL algorithms also build on dependency trees, first detect all verbs (predicates), to then classify all words linked to them (arguments) into predefined sets of semantic roles. While there is no conclusive or generally accepted set of such roles in linguistics, one typically found role is that of an 'agent', meaning the 'volitional causer of an event' (ibid.), which comes close to our conception of 'actorness' here.

With identifying this particular role, the SRL-approach thus promises a more complete perspective on sentences' semantics even in the presence of the high syntactic complexity that is typical for

[4] The eight extraction rules for Stuhler's action motifs are also fully documented in the script "1_ExtractActorness.R" script in the replication package.

diplomatic language. Thus I test the validity of the SRL-module in the AllenNLP language models as provided and implemented in the RELATIO library provided by Ash and colleagues (2022)[5] and code EU actorness whenever the EU occurs in the agent fields extracted for any given sentence in the UNGD speeches (again using the dictionary specified above).

However, SRL-models are computationally expensive and they also come with significant and hardly controllable risks of measurement biases in the light of our ambition to measure ‘actorness’. SRL-algorithms are, to the best of my knowledge, only available as proprietary software and key elements of the machine classification remain opaque. For example, choices on the text spans covered during training and classification, the set of considered roles including their potential overlap or exhaustiveness, as well as restrictions of individual semantic roles to specific verb lists may theoretically drastically affect the results but are beyond the researcher’s control.

Finally, we also have to note that political and especially diplomatic language is often characterised by a very condensed and often rather nominal style (Biber, Conrad, and Reppen 1998; Rauh 2023). For our purpose to capture instances of EU actorness recognition, this kind of language creates another notable risk of ‘false negatives’: actions may often not be expressed as verbs but rather come in the form of nominalizations or adverbial noun phrases. Such syntactic patterns clearly signal actorness for human recipients, but they deviate from the verb-focussed extraction logics that Stuhler’s action motifs or the SRL approach build on. For example, a statement like ‘The EU-led support was crucial’ clearly indicates that the speaker wants to say that EU was leading and supporting something, while this not encoded in a subject-verb syntax (here: support-is) that the other two approaches would extract.

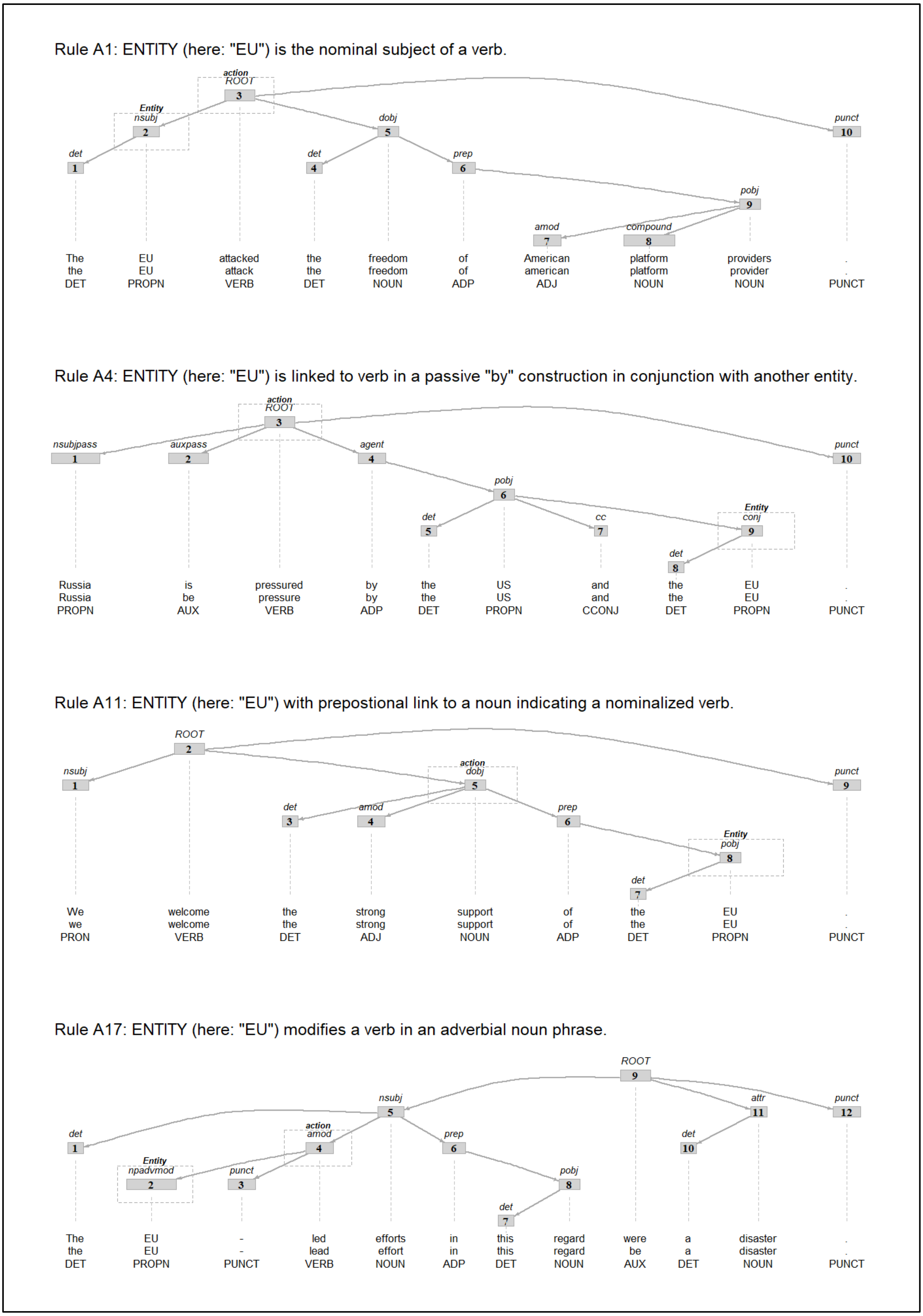
Thus, as a final approach to be validated (*approach 4*), I expand Stuhler’s action motifs with another set of eight fixed extraction rules from sentences’ dependency trees (building on the rsyntax package, Welbers and Van Atteveldt 2022). My additional rules capture all instances in which an

[5] https://github.com/relatio-nlp/relatio (last accessed: 16.02.2023). The script ‘X_SemanticRoleLabelling.R’ in the replication package illustrates application to the UNGD corpus data but requires a separate Python environment will all dependencies installed on your system (see the archive’s README file for further detail).

entity such as the EU modifies a verb in an adverbial noun phrase (e.g. 'the EU-mediated ceasefire') or where it is linked with a noun that represents a nominalized verb, either as a direct compound ('EU support'), as a prepositional object ('agreement between the EU and the US'), or through a possessive relationship ('the EU's refusal'). The two lower panels of Figure 2 provide exemplary dependency trees and highlights my corresponding extractions.

While this offers a much broader coverage than Stuhler's action motifs alone and while the rules are fully transparent and replicable, also here theoretical misclassification risks exist. On the one hand, very complex syntactic patterns with long conjunction chains might be missed. My extraction rules cover up to three layers of such conjunctions (e.g. 'the agreement of the US, Russia, and the EU') but may produce 'false negatives' for longer chains. On the other hand, the identification of nominalizations cannot be rendered completely reliable in the face of the various forms of nominalizations that the English language offers. My rules identify nominalized verbs works along a regular expression comprising typical endings of nominalizations (e.g., '-tion', '-ment', '-ance', etc.) as well as a long list of so-called zero derivation nominalizations where a verb and nominalized action are written identically (e.g. 'approach', 'ban', 'demand', 'support', etc.). While this is transparently documented[6] and can be adapted in future applications, some theoretical risk of false positives or false negatives remain.

[6] The respective functions are contained in the script '1_ExtractActorness.R' in the replication package.



Rule A1: ENTITY (here: "EU") is the nominal subject of a verb.

Rule A4: ENTITY (here: "EU") is linked to verb in a passive "by" construction in conjunction with another entity.

Rule A11: ENTITY (here: "EU") with prepostional link to a noun indicating a nominalized verb.

Rule A17: ENTITY (here: "EU") modifies a verb in an adverbial noun phrase.

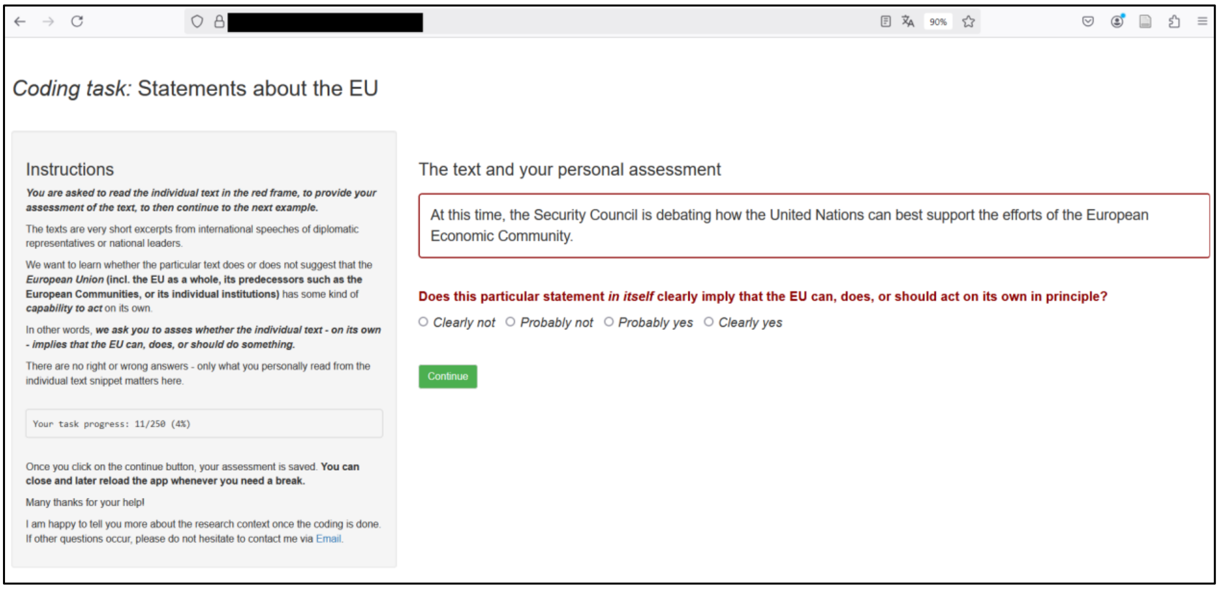
Figure 2: Dependency trees of example sentences illustrating different extraction rules for EU actorness.

Summarising this theoretical discussion in Table 2, we now have four text-as-data approaches to automatically detect the recognition of EU actorness in political speech. These approaches vary in their complexity and, importantly, in potential risks of measurement error in terms of over- or underestimating the 'true' prevalence of EU actorness recognition.

Approach	Actorness indicator	Advantages	Risks
EU mentions	EU occurs in text	- Straightforward implementation	- Potential 'false positives': EU as passive entity or recipient of an action only
Semgram action motifs	EU is active or passive subject of a verb	- EU clearly linked to action - Fixed, transparent, and replicable extraction rules	- Dep. parsing required - Potential 'false negatives' in case of complex syntactics or nominalizations
Semantic role labelling (SRL)	EU occurs in 'agent' field of any SRL-solution by sentence	- Broader conception of agent/action structures - May handle complex syntactic structures better	- Computationally expensive - Proprietary software, opaque classification - Potential 'false negatives', esp. in case of nominalizations
Semgram action motifs expanded	EU is active or passive subject of a verb, modifies a verb in an adverbial noun phrase, or is direct compound of, prepositionally linked to, or in possessive relationship with a nominalized verb	- Broder set of syntactic EU/action links; potentially fewer misclassifications - Fixed, transparent, and replicable extraction rules	- Dep. parsing required - Misidentification of nominalized verbs

Table 2: Overview of EU actorness classification approaches

Whether these theoretical advantages and disadvantages matter in our concrete application can be assessed empirically only if we benchmark each of the classifications against human interpretation of political statements about the EU. Therefore, I drew a stratified *random sample of 750 sentences* mentioning the EU from the UNGD corpus. Then I asked three human coders – graduate political science students with no prior information on the study or the specific measurements to be tested – to read each of these sentences and to assess EU actorness recognition by asking them whether the sentences "imply that the EU can, does, or should act on its own" on a 4-point Likert-scale (implicitly forcing a choice while easing the burden for coders). This coding exercise was implemented in a web-based R-shiny app that is also included in the replication package (together with the resulting coded data). Figure 3 provides a screenshot of this as seen by the coders, including all instructions that they have received.



[redacted]

Coding task: Statements about the EU

Instructions

You are asked to read the individual text in the red frame, to provide your assessment of the text, to then continue to the next example.

The texts are very short excerpts from international speeches of diplomatic representatives or national leaders.

We want to learn whether the particular text does or does not suggest that the ***European Union* (incl. the EU as a whole, its predecessors such as the European Communities, or its individual institutions)** has some kind of ***capability to act*** on its own.

In other words, ***we ask you to asses whether the individual text - on its own - implies that the EU can, does, or should do something.***

There are no right or wrong answers - only what you personally read from the individual text snippet matters here.

```
Your task progress: 11/250 (4%)
```

Once you click on the continue button, your assessment is saved. **You can close and later reload the app whenever you need a break.**

Many thanks for your help!

I am happy to tell you more about the research context once the coding is done. If other questions occur, please do not hesitate to contact me via Email.

The text and your personal assessment

At this time, the Security Council is debating how the United Nations can best support the efforts of the European Economic Community.

Does this particular statement *in itself* clearly imply that the EU can, does, or should act on its own in principle?

○ *Clearly not* ○ *Probably not* ○ *Probably yes* ○ *Clearly yes*

Figure 3: Screenshot of online app for human coding of EU actorness in UNGD sentences.

To assess how difficult this context-free task is for humans, one hundred of these sentences were rated by all three coders. Taking only the overall tendency of coders' assessment into account (sentence implies EU actorness or not), Figure 4 summarises their agreement. In this small sample, intercoder agreement varies between 58 and 78 percent, suggesting that assessing EU actorness in individual sentences is hardly a trivial task. Qualitative inspection of the overlap sample (provided in the replication archive) shows that humans disagree especially on sentences where a speaker claims to speak 'on behalf of' the EU, where the agent of the sentence are 'EU member states' (instances intentionally not covered by my auto-classifications), or where some coders seemed to have applied some contextual knowledge about EU actions which were not explicitly mentioned in the text (e.g. in instances discussing world trade or the Russian Annexation of the Crimea peninsula). How do the four automated classification approach fare against this partially noisy human benchmark?

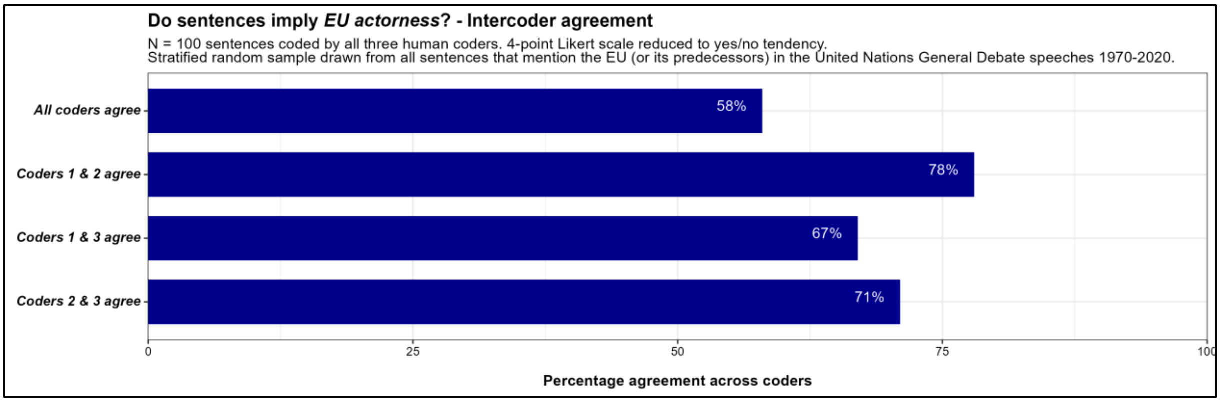
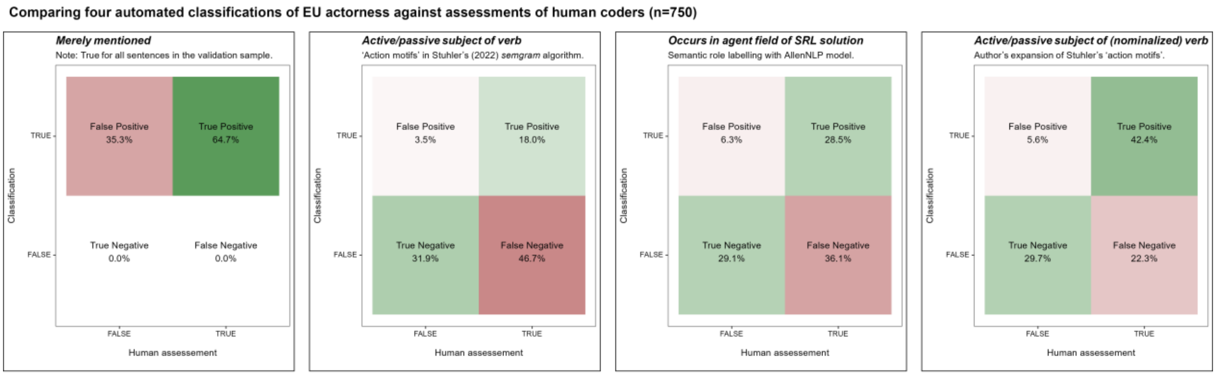


Figure 4: Intercoder reliability for EU actorness

Figure **5** compares the output of each automated classification approach against the human tendency to either see some or no EU actorness across the full validation sample of 750 sentences. It visualizes the respective truth tables and the standard performance metrics for machine-classification tasks that can be derived from them.

Accuracy simply provides the share of all correctly classified sentences when taking the human perspective as the benchmark. *Recall,* sometimes also called sensitivity, focusses on how well a model avoids 'false negatives' by measuring the share of true positives that the machine retrieved among all instances that humans have coded as positive. *Precision*, in contrast, focuses on avoiding false positives by capturing the share of true positives among all cases that the respective model has classified as positive. As highlighted in the theoretical discussion above already, both of these directed metrics often stand in tension to each other as optimizing one can lead to a decline in the other. The *F1 score* therefore also provides a balanced perspective along the harmonic mean of recall and precision.



Accuracy	Recall	Precision	F1.Score
64.67	NA	NA	NA

Accuracy	Recall	Precision	F1.Score
49.87	0.28	0.84	0.42

Accuracy	Recall	Precision	F1.Score
57.6	0.44	0.82	0.57

Accuracy	Recall	Precision	F1.Score
72.13	0.66	0.88	0.75

***Figure 5*:** Performance of different text-based classification methods for EU actorness

As suspected above, the leftmost panel of Figure 5 initially shows that merely taking any EU mention as evidence for EU actorness recognition would lead to a serious overestimation. Even though all sentences in the validation sample contained a reference to the EU or its predecessors, the coders saw evidence for some capability to act on part of the EU only in about 65% of them. In other words, we find a notable false positive rate of around 35%. Given that the sample did not contain any negatives in the light of this classification method, the other performance metrics cannot be meaningfully interpreted here.

The second panel in of Figure 5 assess EU actorness more conservatively along Stuhler's action motifs – meaning that it takes the fact that the EU appears as an active or passive subject of a verb in the respective sentence as the decisive criterion. We see that this simple syntactic addition reduces the false positive rate drastically when comparing the results to a classification along mere EU mentions. But as suspected above, this comes at the high price of producing many false negatives. In other words, this approach misses many instances in which humans saw EU actorness implied by the sentence. It thus lacks sensitivity (recall) and thus underestimates the concept of interest quite strongly.

The commercial and computationally more expensive SRL approach (panel three in Figure 5) reduces this false negative rate by around 10 percentage points in the overall validation sample and introduces only few more false positives. Even though it still depends on syntactic links between the entity of interest and any verb in the sentence, it increases overall accuracy markedly to 58%

percent of agreement with human coders across the 750 sentences. Yet and still, underestimation is still an issue as around 36% of the sample are sentences in which humans did see EU actorness and this particular classification method did not.

Finally, my combination of Stuhler's verb-focussed action motifs with additional rules capturing directed syntactic links of the EU with nominalizations or in adverbial noun phrases (panel four in Figure 5) provides clearly the best relative balance of precision and recall with the highest overall F1 score of .75. Underestimation is arguably still an issue, but it is almost fourteen percentage points lower than the SRL method as the next best competitor. And one has to note that the overall accuracy of this model (72.1%) comes close to the highest level of agreement that we have observed for human coders in the intercoder-reliability check in Figure 4 above (78%). Accordingly, also the laborious task of letting humans code all instances of EU references (or that of other countries or international organisations used as benchmarks in the main text) is not very likely to lead to a more accurate measurement.

Against these validation results, I conclude that *the combination of syntactically linking the EU to actions represented as verbs, adverbial noun-phrases, or nominalized verbs is the most reliable and valid way to measure EU actorness in international political speeches at scale.*

A5: Developing/validating the embedding-based measure for issue contexts

Testing the theoretical arguments developed in the main text requires a reliable and valid measure to capture three broad issue contexts – trade & economy; liberal democracy; security – in individual sentences (the coding unit at which actorness recognition is measured, see above). Conceptually, these issue contexts may plausibly overlap or co-occur within individual sentences, while their relative emphasis could vary in a continuous manner. Consider, three fictious example statements meant to illustrate these conceptual points:

- *'EU support for trade liberalization is crucial for ensuring the security but also liberal democracy in Ukraine'.*
- *'EU support for trade liberalization is crucial for ensuring liberal democracy but especially for the security that the military troops and operations in Ukraine aim to achieve.'*
- *'EU support for trade liberalization is crucial for fostering economic growth and viable business in Ukraine.'*

Arguably, for human readers knowing the meaning of the words used in these statements, it is rather clear that the first example mixes all three issue contexts at roughly equal shares, that the second one emphasizes all three as well but puts greater weight on security issues, while the last example operates solely in a trade & economy context. This is the kind of variation that our text-based measure should ideally capture.

To uncover latent themes in text, current NLP applications in the political sciences often resort to topic models, building on or expanding the Latent Dirichlet Allocation approach (for a seminal introduction, see Blei 2012). Topic models assume that each text is a mixture of topics and that each topic is characterized by a specific distribution of words. Through an iterative process, the model assigns words to topics probabilistically based on their co-occurrence patterns across the full corpus, aiming to maximize the coherence of word groupings within each topic. The output includes a distribution of topics over each document and a distribution of words over each topic (allowing for 'mixed membership' in both instances). While this output matches the conceptual demands specified above in principle, some severe caveats for our specific application here must be noted.

On the one hand, topic models are essentially inductive – in three respects. First researchers need to interpret the meaning of resulting topics ex post, which is typically done along the frequency and specificity of the words within each of them. Yet, this contrasts our clear theoretical interest in the three pre-defined issue areas. Second, the topic model results – in particular the granularity of the resulting topics and the exact distribution of word weights across them – are extremely sensitive to the researcher's choice about the number of topics to optimize for in the first place (typically called the k-parameter). While measures for a statistically optimal fit of topic models exist, there is no guarantee that the resulting distributions are semantically and/or theoretically meaningful, especially if the concrete contents of otherwise stable semantic categories change over time (consider terms like 'military' and 'terrorism' within the broader semantic category of 'security' as an example). Third, topic models are built for optimizing the latent structure across the full corpus exhaustively. Here, however, I do not assume that the three issue areas describe everything that can be said in the UNGD in full – to the contrary, in many sentences probably none of the three issue contexts will be invoked at all. But I also want to avoid that theoretically irrelevant topics absorb variation in word choice that might be relevant for any of the three contexts I am interested in. Thus, the inductive logic of topic models clashes with my particular research interest.

To be sure, there are advanced topic model specification that are able to deal with some of these issues individually. Seeded topic models (e.g. Watanabe and Baturo 2023) are useful to nudge the clustering process to pre-defined categories of interest while dynamic topic models allow variability of word weights for topic predictions over time (e.g. Greene and Cross 2015). However, they are still highly sensitive to the choice on the overall number of topics, are hard to combine, and remain hard to validate ex post.

Moreover, and particularly important here, topic models struggle with the rather short coding units that I am focusing on here. Individual sentences provide only very limited context for the word co-occurrence patterns that the algorithms build on, leading to sparse and fragmented word and topic distributions which may undermine interpretability and accuracy of the analysis further.

Therefore, I built my measurement here on the key idea that any word occurring in the natural English language has more or has less affinity to the three broad issue areas that I want to capture in continuous terms. I thus approach this as a question of semantic similarity captured in word embedding algorithms and the resulting word vector models (Pennington, Socher, and Manning 2014; Spirling and Rodríguez 2022). These models encapsulate the core idea of distributional semantics according to which 'you shall know a word by the company it keeps' (Firth 1957). Put simply, they are trained by moving a fixed window through very large and, in terms of the overall English language, representative text corpora and storing word-to-word co-occurrences in each snapshot. The resulting word-to-word co-occurrence matrix is then reduced by PCA-like procedures to a lower dimensional space in which words that regularly co-occur with similar neighbours also receive similar vector values. Expressed differently, words that often share certain contexts live closer together on the respective dimensions of the vector space.

Building on this powerful idea of capturing meaning across many latent dimensions, I first select five key terms that should be semantically very close to the three target concepts I want to measure. Then I used an open-source and very large pre-trained word vector model – the GloVe vectors trained on the whole English Wikipedia and the CommonCrawl corpus offering a vocabulary of 400k words located across 300 dimensions each (Pennington, Socher, and Manning 2014) – to extract the Cosine similarities of all other words to the average vector of the five respectively selected seed terms. These similarities give us a quantitative indicator of how closely related each word in the vocabulary is to the concepts of interest in semantic terms.

The table below illustrates the procedure by specifying the five respective seed words for each issue context of interest to then list the top-20 most semantically similar terms learned from the broad word vector model representing the overall English language (recall that all 400k words in the vocabulary receive such a score).

	Economy & Trade		Liberal Democracy		Security	
	Word	**Simil. to target vector**	**Word**	**Simil. to target vector**	**Word**	**Simil. to target vector**
Selected seed words defining the target vector	*trade*	0,78	*rights*	0,85	*military*	0,75
	economy	0,77	*freedom*	0,79	*security*	0,75
	market	0,76	*human*	0,77	*war*	0,74
	business	0,76	*democracy*	0,76	*terrorism*	0,73
	commerce	0,72	*law*	0,64	*peace*	0,71
Semantically similar words retrieved from the word vector model (top-20)	industry	0,72	freedoms	0,64	forces	0,68
	economic	0,7	laws	0,59	conflict	0,68
	markets	0,69	liberties	0,59	terror	0,65
	sector	0,66	constitutional	0,58	iraq	0,63
	growth	0,64	respect	0,58	troops	0,62
	businesses	0,63	advocates	0,57	terrorist	0,61
	exports	0,62	legal	0,56	efforts	0,6
	export	0,61	political	0,56	threat	0,6
	financial	0,61	equality	0,56	government	0,59
	investment	0,6	citizens	0,56	afghanistan	0,59
	global	0,6	civil	0,55	force	0,59
	consumer	0,59	activists	0,54	armed	0,59
	retail	0,59	rule	0,54	army	0,58
	sectors	0,58	fundamental	0,53	terrorists	0,58
	companies	0,58	protection	0,53	attacks	0,58
	economies	0,58	advocacy	0,52	weapons	0,57
	prices	0,57	principles	0,52	saying	0,57
	sales	0,57	government	0,52	civilian	0,57
	finance	0,56	groups	0,52	leaders	0,56
	firms	0,56	dignity	0,52	fighting	0,56

Table 3: Word weights to measure issue context similarity learned from GloVe embedding model

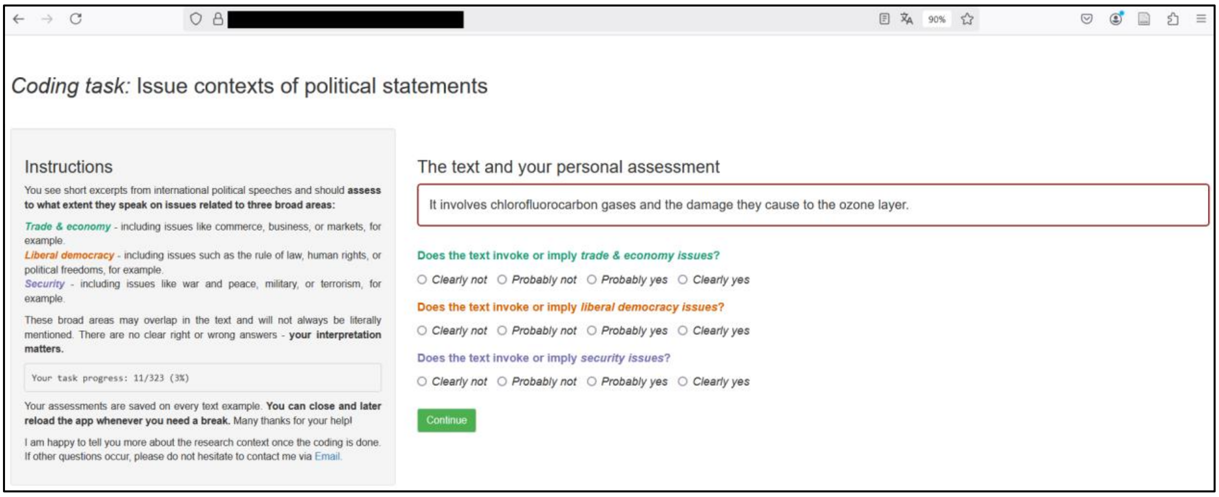
The words we 'learn' to be particularly relevant from this approach initially show high face validity in the light of three broad issue contexts that we are interested in. Words like 'firms', 'investment' or 'export' load high on semantic similarity to the trade & economy vector. Words like 'troops', 'forces' or 'attacks' load high on the security vector. Similar to the intuition of topic models, some words like 'government', for example, load on several issue contexts which we should expect against the assumption that the issue may overlap in natural language.

But again, a key difference to topic models is that we are not bound by words' actual co-occurrence in the short coding units within the corpus itself, but build on their semantic similarity or interchangeability in the overall English language. This is furthermore not blurred by their potential closeness to other latent issue areas that are not of interest here which would be punished in a topic model algorithm. These features allow us to score the short statements of interest in a targeted, deductive manner even if specific words are rare in the corpus or constrained to individual time periods only.

To put this measurement into action, each of the almost one million UNGD sentences was accordingly scored by summing the average similarity of all the words they contain (excluding English stopwords, numbers and punctuation) – with one aggregated value for each of the three issue contexts of interest. The procedure is implemented in the script '3_EmbeddingSimilarities.R' in the replication package (which also offers the raw GloVe vectors in rds format, but note that large memory requirements and run times in the documentation of the replication archive). Qualitative examples are provided in Table1 of the main text and more can be directly extracted from the 'SemanticSimils.rds' data in the replication package.

Yet, a broader and more systematic benchmarking against human interpretation is required to really validate whether this method of capturing the three issue contexts works in this aggregated manner. Thus, I drew a random sample from all sentences in the UNGD corpus, stratified such that the full range of similarity values for all three issue contexts and their possible combinations (along quintiles) in the UNGD corpus were covered.

The resulting 974 sentences were then read by three human coders – graduate political science students with no prior information on the study or the specific measurements to be tested. For each of the sentences in the validation sample they had to answer whether the respective sentences 'invoke or imply [trade & economy|liberal democracy|security] issues' on a 4-point Likert-scale ('clearly not', 'probably not', 'probably yes', 'clearly yes'). This coding exercise was implemented in a web-based R-shiny app that is also included in the replication package (together with the resulting coded data). Figure 6 provides a screenshot of this app as seen by the coders, including all instructions that they have received.



[redacted]

Coding task: Issue contexts of political statements

Instructions

You see short excerpts from international political speeches and should **assess to what extent they speak on issues related to three broad areas:**

Trade & economy - including issues like commerce, business, or markets, for example.
Liberal democracy - including issues such as the rule of law, human rights, or political freedoms, for example.
Security - including issues like war and peace, military, or terrorism, for example.

These broad areas may overlap in the text and will not always be literally mentioned. There are no clear right or wrong answers - **your interpretation matters.**

```
Your task progress: 11/323 (3%)
```

Your assessments are saved on every text example. **You can close and later reload the app whenever you need a break.** Many thanks for your help!

I am happy to tell you more about the research context once the coding is done. If other questions occur, please do not hesitate to contact me via Email.

The text and your personal assessment

It involves chlorofluorocarbon gases and the damage they cause to the ozone layer.

Does the text invoke or imply *trade & economy issues*?

○ *Clearly not* ○ *Probably not* ○ *Probably yes* ○ *Clearly yes*

Does the text invoke or imply *liberal democracy issues*?

○ *Clearly not* ○ *Probably not* ○ *Probably yes* ○ *Clearly yes*

Does the text invoke or imply *security issues*?

○ *Clearly not* ○ *Probably not* ○ *Probably yes* ○ *Clearly yes*

Figure 6: Screenshot of online app for human coding of issue contexts invoked in UNGD sentences.

The thus received human ratings were then compared against the numerical values generated from the embedding-based similarity weights generated along the procedure outlined above. Figure 7 summarizes the results by showing the mean values and confidence intervals of our issue context scores across the four values that the coders picked for any of the three specific issue contexts and any of the 974 sentences in the sample.

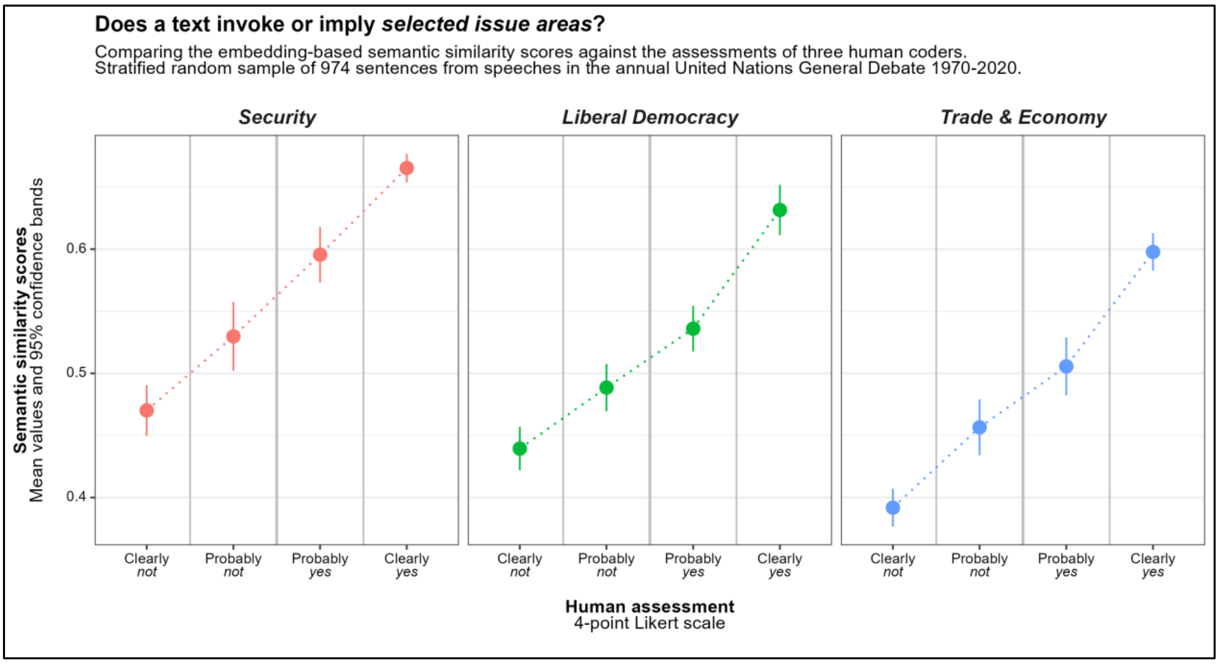


Figure 7: Human ratings of issue context prevalence against embedding-based scores.

The resulting patterns strongly support the measurement approach proposed in this paper. We see that the retrieved scores increase continuously and almost linearly with the human assessments of whether the respective issue context was invoked by the respective sentence in question. In statistical terms, moreover, the retrieved scores discriminate robustly between the four individual categories that human coders could choose from, thus also replicating human assessment in a rather granular manner. *This shows that the chosen method offers a valid measurement for the three issue contexts of interest in sentences drawn from UNGD speeches.*

Yet, the results also highlight some notable characteristics of the scales on which the retrieved scores can vary. Given that any word will most likely co-occur with any word in the large training data of the underlying word vector model at least a few times, the resulting cosine similarities will mostly end up on the positive range of the theoretically possible Cosine similarity scale (-1 to +1). Moreover, Figure 7 suggests that the three issue contexts seem to come with slightly different intercepts in the validation data prepared for the human coders. Substantively, it is not surprising that security issues but also democracy talk (cf. Stephen 2015) have slightly higher baseline likelihood in UN debates than trade & economy issues.

Both characteristics of the scales warrant caution in interpreting them in absolute terms, an issue that is often also overlooked in other word-frequency based methods (cf. Rauh 2018). In the analyses of the main manuscript I therefore only use standardized values in the descriptive analysis or capture relative change as in the linear probability models. This ensures that we can reliably and validly make comparative statements on issue context prevalence as required by the theoretically derived hypotheses.

A6: EU actorness recognition – counts per annual UNGD

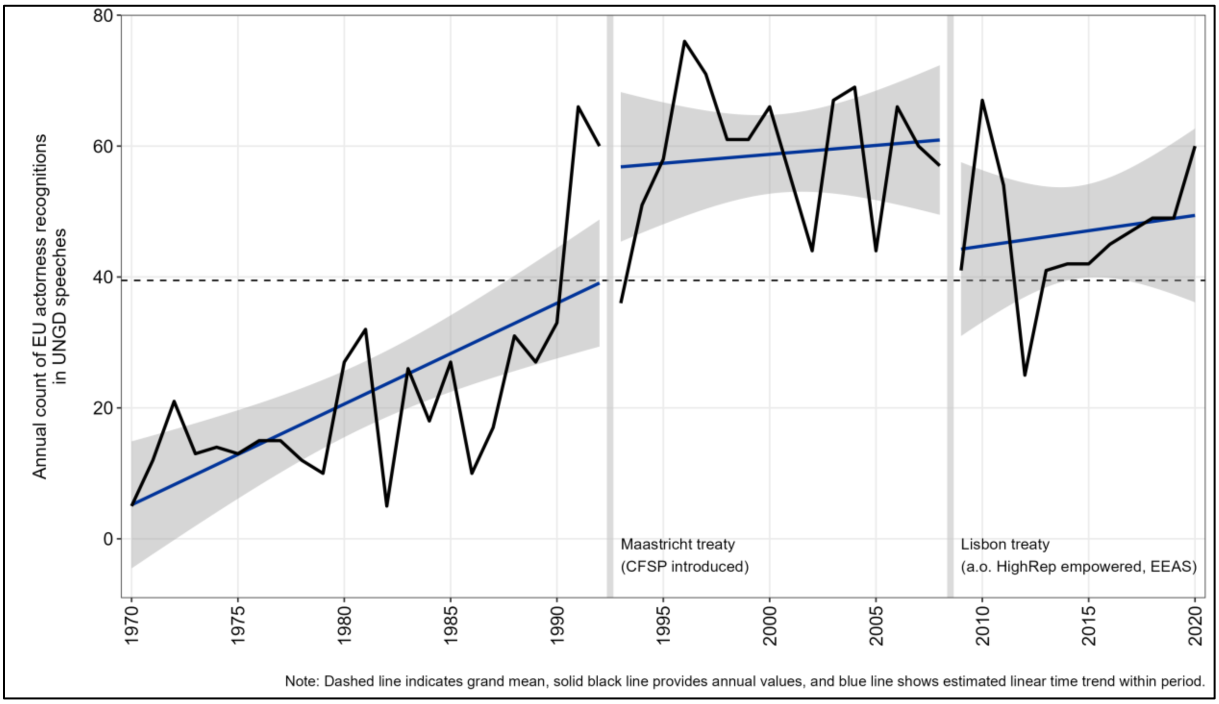
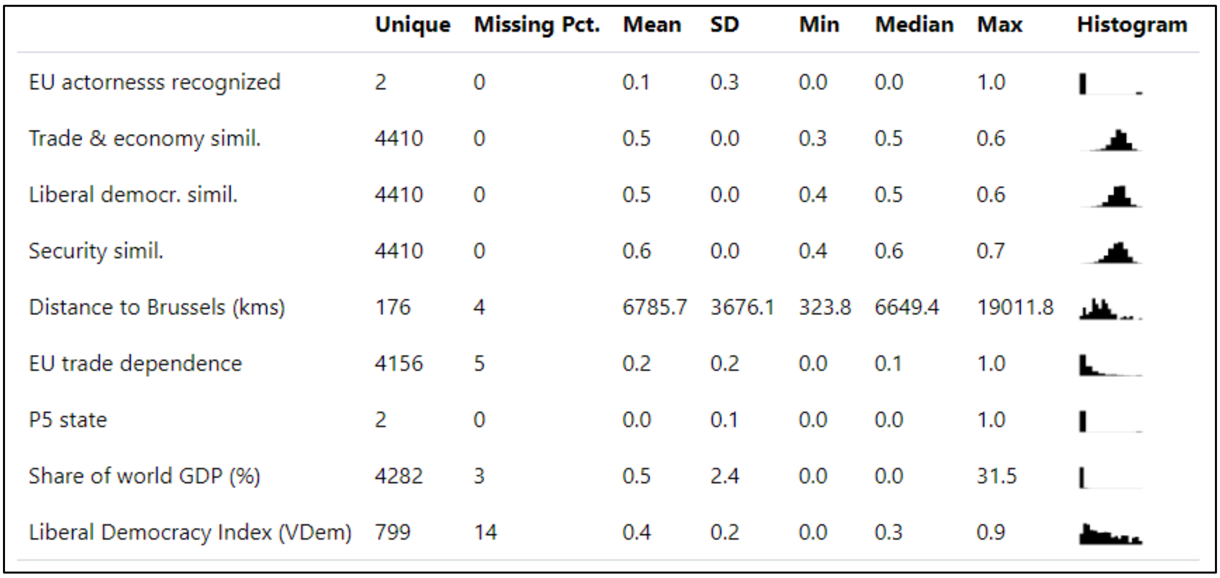


Figure 8: Assessing prevalence of EU actorness recognition in UNGD debates along counts of respective instances.

A7: Data and variables used in the linear probability model

Table 4 provides an overview of all country-level data available for the 4,407 UNGD speeches by country representatives between 1993 and 2020.



	Unique	Missing Pct.	Mean	SD	Min	Median	Max	Histogram
EU actornesss recognized	2	0	0.1	0.3	0.0	0.0	1.0	
Trade & economy simil.	4410	0	0.5	0.0	0.3	0.5	0.6	
Liberal democr. simil.	4410	0	0.5	0.0	0.4	0.5	0.6	
Security simil.	4410	0	0.6	0.0	0.4	0.6	0.7	
Distance to Brussels (kms)	176	4	6785.7	3676.1	323.8	6649.4	19011.8	
EU trade dependence	4156	5	0.2	0.2	0.0	0.1	1.0	
P5 state	2	0	0.0	0.1	0.0	0.0	1.0	
Share of world GDP (%)	4282	3	0.5	2.4	0.0	0.0	31.5	
Liberal Democracy Index (VDem)	799	14	0.4	0.2	0.0	0.3	0.9	

Table 4: Description of variables available for multivariate estimation

We have to note that some of the data for the independent variables described in section 3 of the main text are not available for every country/year combination. The LPM presented in the main text uses only the 3,638 complete cases (= 83%). In other words, cases with missing data were dropped by listwise deletion and are briefly listed here.

First, we lack distance data for Congo (Kinshasa), Liechtenstein, Monaco, Montenegro, the Palestinian Territories, Romania, Serbia, South Sudan, Timor-Leste, and Vatican City.

Second, for some countries and individual years (often: wars) no trade data is available. This concerns at different points in time (between one and 17 years): Afghanistan, Bosnia & Herzegovina, Croatia, Eritrea, Estonia, Israel, Latvia, Liberia, Lithuania, Marshall Islands, Micronesia (Federated States of), Moldova, Monaco, Nauru, Palau, San Marino, São Tomé & Príncipe, Slovenia, Somalia, South Sudan, Turkmenistan, Vatican City, Venezuela, and Yemen. For a few countries trade data is missing throughout the 28-year period: Andorra, Liechtenstein, Monaco, North Korea.

The biggest factor, however, is the lacking coverage of the V-Dem liberal democracy index particularly for smaller states that are partially or completely lacking here (accounting for about 14% of the raw speech observations): Andorra, Antigua & Barbuda, Bahamas, Bahrain, Belize, Brunei, Dominica, Grenada, Kiribati, Liechtenstein, Marshall Islands, Micronesia (Federated States of), Monaco, Nauru, Palau, Palestinian Territories, Samoa, San Marino, Serbia, St. Kitts & Nevis, St. Lucia, St. Vincent & Grenadines, Tonga, Tuvalu, Vatican City, Yemen.

While our estimation data is thus probably somewhat biased against war-torn and especially against small island states, these lists do not raise suspicion on biases for or against any of the tested hypotheses.

Note, finally, the extremely right-skewed distribution on the GDP world share measures. To avoid turning this effectively into a US/China dummy and to avoid biased standard errors, this variable enters the model as a log transformation.

A8: Alternative model specifications and error correction approaches

Model #	1	2	3	4	5	6	7	8	9	10	11
Estimator	**OLS**	**OLS**	**OLS**	**OLS**	**OLS**	**Logit**	**Logit**	**Logit**	**Logit**	**Logit**	**Logit (Firth correct.)**
Std. errors	*IID*	*Heterosked. -robust*	*Clustered by: year*	*Clustered by: iso3*	*Clustered by: iso3 & year*	*IID*	*Heterosked. -robust*	*Clustered by: year*	*Clustered by: iso3*	*Clustered by: iso3 & year*	*IID*
Issue: Trade & Economy	**0.02*****	**0.02*****	**0.02*****	**0.02*****	**0.02*****	**0.29*****	**0.29*****	**0.29*****	**0.29*****	**0.29*****	**0.29*****
	-0.01	-0.01	0	-0.01	-0.01	-0.07	-0.07	-0.06	-0.08	-0.07	-0.07
Issue: Liberal democracy	**-0.02****	**-0.02****	**-0.02****	**-0.02***	**-0.02***	**-0.30****	**-0.30****	**-0.30****	**-0.30***	**-0.30***	**-0.30****
	-0.01	-0.01	-0.01	-0.01	-0.01	-0.1	-0.1	-0.1	-0.13	-0.13	-0.1
Issue: Security	**0.01+**	**0.01+**	**0.01+**	**0.01**	**0.01**	**0.19+**	**0.19+**	**0.19+**	**0.19**	**0.19**	**0.18+**
	-0.01	-0.01	-0.01	-0.01	-0.01	-0.1	-0.11	-0.11	-0.14	-0.14	-0.1
Distance to Brussels	**-0.03*****	**-0.03*****	**-0.03*****	**-0.03+**	**-0.03+**	**-0.37*****	**-0.37*****	**-0.37*****	**-0.37+**	**-0.37+**	**-0.36*****
	-0.01	-0.01	-0.01	-0.01	-0.01	-0.08	-0.1	-0.1	-0.19	-0.2	-0.08
EU trade dependence	**0.04*****	**0.04*****	**0.04*****	**0.04****	**0.04****	**0.30*****	**0.30*****	**0.30*****	**0.30****	**0.30****	**0.30*****
	-0.01	-0.01	-0.01	-0.01	-0.01	-0.06	-0.06	-0.06	-0.1	-0.1	-0.06
P5 State	**0**	**0**	**0**	**0**	**0**	**0.07**	**0.07**	**0.07**	**0.07**	**0.07**	**0.07**
	-0.01	-0.01	0	-0.01	-0.01	-0.06	-0.06	-0.05	-0.06	-0.06	-0.06
GDP world share (log.)	**-0.01**	**-0.01**	**-0.01**	**-0.01**	**-0.01**	**-0.1**	**-0.1**	**-0.10+**	**-0.1**	**-0.1**	**-0.09**
	-0.01	0	0	-0.01	-0.01	-0.07	-0.06	-0.05	-0.12	-0.11	-0.07
V-Dem Lib. Democracy	**0.02*****	**0.02*****	**0.02****	**0.02***	**0.02***	**0.22*****	**0.22*****	**0.22****	**0.22***	**0.22***	**0.22*****
	-0.01	-0.01	-0.01	-0.01	-0.01	-0.06	-0.06	-0.08	-0.09	-0.11	-0.06
Num. Obs.	3638	3638	3638	3638	3638	3638	3638	3638	3638	3638	3638
R2	0.051	0.051	0.051	0.051	0.051	0.071	0.071	0.071	0.071	0.071	0.062
R2 Adj.	0.049	0.049	0.049	0.049	0.049	0.064	0.064	0.064	0.064	0.064	
AIC	1227.2	1227.2	1227.2	1227.2	1227.2	2142.6	2142.6	2142.6	2142.6	2142.6	
BIC	1283	1283	1283	1283	1283	2198.3	2198.3	2198.3	2198.3	2198.3	
RMSE	0.29	0.29	0.29	0.29	0.29	0.28	0.28	0.28	0.28	0.28	0.28
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001, standard errors below coefficients											

Table 5: Regression models across different model specifications
(model 1 is visualised in the main text)

Appendix references

Ash, Elliott, Germain Gauthier, and Philine Widmer. 2022. "Text Semantics Capture Political and Economic Narratives." *arXiv:2108.01720 [Econ, q-Fin]*, February. http://arxiv.org/abs/2108.01720.

Atteveldt, Wouter van, Tamir Sheafer, Shaul R. Shenhav, and Yair Fogel-Dror. 2017. "Clause Analysis: Using Syntactic Information to Automatically Extract Source, Subject, and Predicate from Texts with an Application to the 2008–2009 Gaza War." *Political Analysis* 25 (2): 207–222. doi:10.1017/pan.2016.12.

Benoit, Kenneth, and Akitaka Matsuo. 2020. "Spacyr: Wrapper to the 'spaCy' 'NLP' Library. R Package Version 1.2.1." https://cran.r-project.org/package=spacyr.

Biber, Douglas, Susan Conrad, and Randi Reppen. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge ; New York: Cambridge University Press.

Blei, David. 2012. "Probabilistic Topic Models." *Commun. ACM* 55 (4): 77–84. doi:10.1145/2133806.2133826.

Firth, J. R. 1957. "A Synopsis of Linguistic Theory 1930-55." *Studies in Linguistic Analysis* 1952–59. Oxford: The Philological Society: 1–32.

Franzosi, Roberto, Gianluca De Fazio, and Stefania Vicari. 2012. "Ways of Measuring Agency: An Application of Quantitative Narrative Analysis to Lynchings in Georgia (1875–1930)." *Sociological Methodology* 42 (1). SAGE Publications Inc: 1–42. doi:10.1177/0081175012462370.

Greene, Derek, and James Cross. 2015. "Unveiling the Political Agenda of the European Parliament Plenary: A Topical Analysis," June. http://arxiv.org/abs/1505.07302.

Honnibal, Matthew, and Ines Montani. 2020. "spaCy: Industrial-Strength Natural Language Processing." https://spacy.io/.

Jurafsky, Dan, and James H. Martin. 2009. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall.

Kleinnijenhuis, J, JA De Ridder, and EM Rietberg. 1997. "Reasoning in Economic Discourse: An Application of the Network Approach to the Dutch Press." In *Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences from Texts and Transcripts.*, edited by CW Roberts. Mahwah, NJ: Erlbaum. http://sf4.ub.fu-berlin.de/F?func=file&file_name=find-b&local_base=fub01.

Knight, Carly. 2022. "When Corporations Are People: Agent Talk and the Development of Organizational Actorhood, 1890–1934." *Sociological Methods & Research* 51 (4): 1634–1680.

Koopmans, Ruud, and Paul Statham. 1999. "Political Claims Analysis: Integrating Protest Event and Political Discourse Approaches." *Mobilization: An International Quarterly* 4 (2): 203–221. doi:doi:10.17813/maiq.4.2.d759337060716756.

Kriesi, Hanspeter, Edgar Grande, Romain Lachat, Martin Dolezal, Simon Bornschier, and Timotheos Frey. 2006. "Globalization and the Transformation of the National Political Space: Six European Countries Compared." *European Journal of Political Research* 45 (6): 921–956. doi:10.1111/j.1475-6765.2006.00644.x.

Pennington, Jeffrey, Richard Socher, and Christopher D. Manning. 2014. "GloVe: Global Vectors for Word Representation." In *Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. http://www.aclweb.org/anthology/D14-1162.

Rauh, Christian. 2018. "Validating a Sentiment Dictionary for German Political Language—a Workbench Note." *Journal of Information Technology & Politics* 15 (4): 319–343.

Rauh, Christian. 2023. "Clear Messages to the European Public? The Language of European Commission Press Releases 1985–2020." *Journal of European Integration* 45 (4). Routledge: 683–701. doi:10.1080/07036337.2022.2134860.

Sjöstedt, Gunnar. 1977. *The External Role of the European Community*. Swedish Studies in International Relations. Farnborough, Hampshire: Saxon House.

Spirling, Arthur, and Pedro L. Rodríguez. 2022. "Word Embeddings What Works, What Doesn't, and How to Tell the Difference for Applied Research." *Journal of Politics* 84 (1): 101–115.

Stephen, Matthew D. 2015. "'Can You Pass the Salt?' The Legitimacy of International Institutions and Indirect Speech." *European Journal of International Relations* 21 (4). SAGE Publications Ltd: 768–792. doi:10.1177/1354066114563417.

Stuhler, Oscar. 2022. "Who Does What to Whom? Making Text Parsers Work for Sociological Inquiry." *Sociological Methods & Research* 51 (4). SAGE Publications Inc: 1580–1633. doi:10.1177/00491241221099551.

Watanabe, Kohei, and Alexander Baturo. 2023. "Seeded Sequential LDA: A Semi-Supervised Algorithm for Topic-Specific Analysis of Sentences." *Social Science Computer Review*, May. SAGE Publications Inc, 08944393231178605. doi:10.1177/08944393231178605.

Welbers, Kasper, and Wouter Van Atteveldt. 2022. "Rsyntax: Extract Semantic Relations from Text by Querying and Reshaping Syntax." https://cran.r-project.org/package=rsyntax.

Wickham, Hadley. 2015. "Stringr: Simple, Consistent Wrappers for Common String Operations. R Package Version 1.0.0." https://cran.r-project.org/package=stringr.