Argument-Structure Constructions in Organization Names in the English Eurolect: The Case of [ORG + V + that + SC]

Fernando Sánchez Rodas  
Research Institute of Multilingual Language Technologies, University of Malaga, SPAIN

Gloria Corpas Pastor  
Research Institute of Multilingual Language Technologies, University of Malaga, SPAIN
Abstract

Construction Grammar (Hoffmann & Trousdale 2013) has received very little attention in onomastics, let alone corpus-based approaches, as corpora are just starting to be applied to the empirical study of names (Motschenbacher 2020). This study employs Named Entity Recognition plus verbal pattern extraction in an intermodal corpus (Bernardini 2016) of EU discourse, or Eurolect (Sandrelli 2018). The methodological aim is to mine English argument-structure constructions (Goldberg 1995) with subordinate clauses introduced by that and organization names in the subject slot ([ORG + V + that + SC]). First, the personification recognition method of Dorst, Mulder, and Steen (2011) is applied to quantitatively prove the strong relationship between the extracted argument-structure constructions and personification metaphors in EU discourse. Second, the constructions and their form-meaning pairings are described, both per subcorpus and globally. Results show that, at a macro- and meso-level of schematicity, the [ORG + V + that + SC] construction transversally symbolizes personification as an understanding scheme for institutional relations, constructing organization names with semantically human verbs of belief, speech, and thought. At a microscopic level, however, some constructions occur exclusively in one of the four subcorpora (non-translated, translated, non-interpreted, and interpreted English), meaning that they could be covering specific mediating functions through their name-verb slot choices.

Keywords: argument-structure constructions, Construction Grammar, Eurolect, intermodal corpora, Named Entity Recognition, organization names

Introduction

The term “Construction Grammar” (CxG) refers to a family of closely related linguistic approaches which contrast with Chomskyan views on language and idiomaticity (Hoffmann & Trousdale 2013). Contrary to mainstream Generative Grammar, in CxG constructions are not the result of a limited set of transformations or derivations, but symbolic units which are linked to each other and constitute complex networks (Goldberg 2006, 1995). Corpus-based constructional approaches have recently shown great potential for the description of both non-mediated and mediated discourse. This is especially true for idiomatic constructions (for example verbal collocations with “disease”/enfermedad (Corpas Pastor 2015), the construction [V + PP + de miedo] in translated Spanish (Corpas Pastor 2017), idioms expressing insanity in English and Spanish (Corpas Pastor 2021a), and resultative idioms expressing a change of mind in English (Corpas Pastor 2022). However, specific applications to highly semasiological disciplines such as terminology or onomastics are almost non-existent in any mediation mode or language (Faber 2012). This comes somewhat as a surprise, given the interest that preponderant construction grammarians have recently shown in onomastics in the form of some co-edited volumes (cf. Ainiala & Ostman 2017).

One of the linguistic phenomena in which the intersection of CxG and onomastics could deploy its full potential is that of ontological metaphor, and more specifically, personification. Personifications are metaphorical constructions with a classical tradition in various fields (linguistics, rhetoric, literature, the arts, etc.), including Deconstruction Theory (De Man 1986). The development of cognitive science also contributed to the revisiting of personification, as it began to be considered one of the most basic ontological metaphors (Kövecses 2002; MacKay 1986; Lakoff & Johnson 1980). Little empirical work has been done on the manifestations of personification in discourse, and it remains unclear how personifications can be reliably identified and analyzed (Dorst 2011). However, its connections with the study of names are clear, especially in the case of personifications-with-metonymy. These are violations of the selectional restrictions of the basic sense of a verb caused by the replacement of a human agent or patient with a metonymically related non-human agent or patient, which frequently happens to be a name, such as in The CND office telephoned to ask her for voluntary evening help (Dorst et al. 2011) or Paris and Washington are having a spat (Evans & Green 2006). Among them, personifications of organization names\(^2\) (European Union, FBI, Microsoft...) are an attractive object of study, as they offer information about components of meaning that go beyond grammar and approach the encyclopedic knowledge of speakers (Viimaranta & Mustajoki 2020).

Our study covers a specific case of argument-structure construction with organization names, [ORG + V + that + SC], in a corpus of EU English, or Eurolect (Sandrelli 2018). In this institutional variety, such a
construction can be found in a wide range of mediation modes, as some examples from our corpus show (see the Methodology section):

- Non-translated English Eurolect (*Bundestag decided that harmonisation was to be sought at EU level*)
- Non-interpreted English Eurolect (*the Commission considers that national authorities are better placed to assess the factual situation and intervene if necessary*)
- Interpreted English Eurolect (*the European Commission believes that effective application of EU environmental law and the obligation as an outcome of the relevant two European Directives should contribute to resolving this complex and critical situation*)

The construction is filled with subject organization names (ORG) followed by simple verbs (V) and the complementizer *that* introducing a subordinate clause (SC). Our main aim is two-fold: (1) to demonstrate how the combination of Construction Grammar with state-of-the-art language technologies (that is, Named Entity Recognition [NER] plus automatic verbal pattern extraction) opens the door to obtaining considerable amounts of raw metaphorical data from specialized corpora in a fast and efficient way; and (2) to analyze the form-meaning pairings of the metaphorical constructions (personifications) obtained this way, with a view to uncovering name-verb interactions, schematization, and networks among constructions in the various subcorpora at hand.

The remainder of this paper is organized as follows. Methodological aspects and datasets for the analysis are provided in the methodology section. Then, a two-step data analysis which involves metaphor identification and constreal account of personification and schematicity is described. Finally, the main results of our analysis are presented and discussed, along with concluding remarks and future lines of research.

**Methodology**

The methodology of this paper relies heavily on corpus-based Named Entity Recognition, or NER (Jacquet et al. 2019), combined with automatic verbal pattern extraction (Corpas Pastor 2020). Previous research using similar NLP techniques in English-Spanish European Parliament corpora discovered a tendency for name-verb patterns such as *Commission considers* or *la Comisión entiende* (Sánchez Rodas 2022b; Corpas Pastor & Sánchez Rodas in press). These basic term-embedding collocations form the skeleton of legal rules by providing action and enabling terms to enter relations, but most times they are not comprehensive enough to symbolize personification, which tends to be conveyed by more complex constructions (e.g., *the Commission continues monitoring the implementation of EU environmental policy*). For this reason, this paper will study the argument-structure construction [ORG + V + *that* + SC] in the English Eurolect.

Paraphrasing Goldberg (1995: 224–225), personifications can be defined as argument-structure constructions in which the association with humanly relevant scenes is achieved through a non-human subject or object: something transferring something to someone, something causing something to move or to change state, something experiencing something, something undergoing a change of state or location, and so on.

**Choice of Corpus Data and Corpus Management System**

Our corpus-based approach is in the wake of the recently envisioned possibilities of introducing corpora in onomastics (Motschenbacher 2020) and employs PETIMOD 3.0, an intermodal corpus of the Committee on Petitions (PETI) of the European Parliament (EP). PETIMOD 3.0 is the last enlargement of a collection containing eight subcorpora of all the possible variants of mediated and non-mediated discourse (non-translated/translated/translated/interpreted) in the English and Spanish Eurolect (Corpas Pastor & Sánchez Rodas 2022; Sánchez Rodas 2022a, 2022b, 2021). The oral subcorpora contain committee interventions from Members of the European Parliament (MEPs) and invited speakers; written subcorpora contain mostly petitions, SIR documents (Summary, Information, and Recommendation), draft agendas, minutes, letters, and opinions. Our study will focus only on the English subcorpora because of space limitations (see Table 1). In the table below, the following abbreviations are used: N-T_EN (Non-Translated English); N-I_EN (Non-Interpreted English); T_EN (Translated English); and I_EN (Interpreted English).
**Table 1**: Size of PETIMOD 3.0 (English) per Document Number, Types, and Tokens

<table>
<thead>
<tr>
<th>Subcorpora</th>
<th>Documents</th>
<th>Types</th>
<th>Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-T_EN</td>
<td>880</td>
<td>29,273</td>
<td>1,159,248</td>
</tr>
<tr>
<td>N-I_EN</td>
<td>155</td>
<td>2,931</td>
<td>23,879</td>
</tr>
<tr>
<td>T_EN</td>
<td>7</td>
<td>1,820</td>
<td>7,530</td>
</tr>
<tr>
<td>I_EN</td>
<td>137</td>
<td>3,222</td>
<td>33,891</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td><strong>1,179</strong></td>
<td><strong>37,246</strong></td>
<td><strong>1,224,548</strong></td>
</tr>
</tbody>
</table>

On the one hand, the distinction between non-translated and non-interpreted texts responds to the empirically confirmed linguistic traits that translations and interpretations into any world language present when compared to texts originally produced in that same language, and which have been termed translation and interpretation universals: simplification, explicitation, convergence, normalization, and transference (Fumagalli 1999; Baker 1993). On the other hand, the mediated texts in this corpus are simply labelled “translated” and “interpreted” because they pertain to the prototypical categories (that is, written translations rendered asynchronically and simultaneous interpretations performed in both cases by humans). However, intermodal corpora could theoretically include and compare any translation and interpreting sub-mode (for example, written translations vs. sight translations, consecutive interpretations vs. simultaneous interpretations, machine translations vs. human translations).

First, the four subcorpora shown in Table 1 were uploaded to the VIP system, an open-source, purpose-built integrated platform of various functionalities, tools, and resources intended to enhance interpreters’ use of technology (Corpas Pastor 2021b, 2020). Then, a list of patterns identified as candidates for argument-structure constructions was automatically obtained using the pattern extraction functionality. This functionality integrates NER, offering the possibility of looking for structures with named entities (“ENT”) followed by verbs (“V”), and even allowing for the search to be restricted to a specific named entity tag, such as “PER” (personal names), “GPE” (countries, cities, states), and “LAW” (named documents made into laws). We introduce “ORG”, which includes names of companies, agencies, and institutions. Finally, the system does not allow the user to search for syntactic arguments like “SC” (subordinate clause), which would be closer to common CxG notation. To solve this, we introduce the complementizer *that* as a prototypical point of access to a subordinate clause in argument-structure constructions (Figure 1).
Figure 1: Example of an ORG + V + that Pattern Search in VIP (source: http://www.lexytrad.es/en/resources/vip/). The corpora displayed on the left are our own.

Named Entity Recognition with SpaCy and DeepPavlov

To measure the performance of the two models of VIP for pattern extraction (SpaCy and DeepPavlov), the ORG + V + that patterns of each corpus were exported to Excel files and compared in terms of precision, recall, and F-1 score. Precision errors include wrongly identified, classified, and/or transcribed named entities (for example, the Rules of Application provides that, the European Court of Justice declared that) plus wrongly identified verbs, mostly ambiguous -ing forms (for example, INTA saying that). The remaining patterns were manually recognized with the help of that-as-complementizer bigram lists, analogical to the wordlists previously employed for simple named entity (NE) recall (Corpas Pastor & Sánchez Rodas 2022). Four bigram lists (one per subcorpus) were generated in VIP, exported to an Excel file, and filtered by V + that patterns (for example, anticipated that) and PREP + that patterns that could be a hint of phrasal verbs (for example, down that as part of notes down that, out that as part of pointed out that), whose concordances were searched for. Tables 2 and 3 compare the metrics of each model for all subcorpora.

Table 2: SpaCy precision, recall, and F-1 scores for [ORG + V + that]

<table>
<thead>
<tr>
<th>COUNTS</th>
<th>N-T_EN</th>
<th>N-I_EN</th>
<th>T_EN</th>
<th>I_EN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automatically recognized</td>
<td>229</td>
<td>3</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Errors</td>
<td>23</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Relevant</td>
<td>206</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td><strong>Precision</strong></td>
<td><strong>0.90</strong></td>
<td><strong>0.67</strong></td>
<td><strong>0</strong></td>
<td><strong>1</strong></td>
</tr>
<tr>
<td>Manually recognized</td>
<td>129</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Total relevant</td>
<td>336</td>
<td>2</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td><strong>Recall</strong></td>
<td><strong>0.61</strong></td>
<td><strong>1</strong></td>
<td><strong>0</strong></td>
<td><strong>0.67</strong></td>
</tr>
<tr>
<td><strong>F-1 score</strong></td>
<td><strong>0.73</strong></td>
<td><strong>0.80</strong></td>
<td><strong>0</strong></td>
<td><strong>0.80</strong></td>
</tr>
</tbody>
</table>
Table 3: DeepPavlov precision, recall, and F-1 scores for [ORG + V + that]

<table>
<thead>
<tr>
<th>COUNTS</th>
<th>N-T_EN</th>
<th>N-I_EN</th>
<th>T_EN</th>
<th>I_EN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automatically recognized</td>
<td>303</td>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Errors</td>
<td>16</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Relevant</td>
<td>287</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td><strong>Precision</strong></td>
<td><strong>0.95</strong></td>
<td><strong>0.67</strong></td>
<td><strong>0</strong></td>
<td><strong>0.67</strong></td>
</tr>
<tr>
<td>Manually recognized</td>
<td>49</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Total relevant</td>
<td>336</td>
<td>2</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td><strong>Recall</strong></td>
<td><strong>0.85</strong></td>
<td>1</td>
<td>0</td>
<td><strong>0.67</strong></td>
</tr>
<tr>
<td><strong>F-1 score</strong></td>
<td><strong>0.90</strong></td>
<td><strong>0.80</strong></td>
<td>0</td>
<td><strong>0.67</strong></td>
</tr>
</tbody>
</table>

The average F-1 score for ORG + V + that extraction in English is slightly higher (+0.01) for DeepPavlov, as it performs better in the N-T_EN subcorpus (0.58 vs. 0.59 average). By contrast, no instances were found in T_EN, either with SpaCy or DeepPavlov. This could have to do with the fact that most of its source languages are constructionally distant from English (for example, Bulgarian, Greek, Lithuanian, Romanian). The difference between the number of constructs in N-T_EN and the other subcorpora (N-I_EN, T_EN, I_EN) could be motivated by the higher number of documents, but it could also mean that the [ORG + V + that + SC] argument-structure construction is strongly linked to non-translated discourse and is much less frequent in mediated and/or oral texts. Although differences in the performance of both models are very small, we use these reference values to undertake the analysis of the DeepPavlov results.

Analysis

Our analysis of the data has been performed in two phases. First, we apply metaphor identification methods based on relevant literature (Steen et al. 2010) to recognize personifications among the ORG + V + that patterns extracted in Section 2.2. Our primary goal is to establish a solid link between argument-structure constructions with names and cognitive metaphor. After this step, a proper constructional analysis is performed, bearing in mind the effects of personification on the characteristics of constructions (mostly on schematicity).

Personification Rates in [ORG + V + that + SC] Constructions

To quantify personifications in the construction [ORG + V + that + SC], we follow the methodology deployed by Dorst et al. (2011), replacing the Macmillan English Dictionary for Advanced Learners with the current online version of the Macmillan Dictionary (https://www.macmillandictionary.com/), a more up-to-date and productive resource which is also corpus-based. We take the ORG + V + that patterns extracted with DeepPavlov as a starting point, look for the associated [ORG + V + that + SC] construction in the corpus concordances, and manually discard those samples in which the constructed verb does not have a human meaning according to a suitable dictionary definition and/or the context. The humanness of the verb can be deduced from one or more signal elements from the definitions (Dorst et al. 2011): nouns and pronouns (“people”, “someone”), body parts (“thumb”, “finger”), verbs and verbal collocations (“to make a choice”, “to express something in speech or writing”), and sentence examples in italics (I recommend that you buy a more powerful computer). Although out of scope for this study, a degree of personification was also perceived, based on the definition order in the dictionary entry; the earlier human signals are detected, the more it seems to reinforce the metaphor, as it could mean that the human sense of the verb is more prototypical, and the subsequent definitions represent metaphorical extensions.7

Table 4: Personification percentages for the [ORG + V + that + SC] construction

<table>
<thead>
<tr>
<th>PETIMOD 3.0</th>
<th>N-T_EN</th>
<th>N-I_EN</th>
<th>T_EN</th>
<th>I_EN</th>
</tr>
</thead>
<tbody>
<tr>
<td>ans-names.pitt.edu</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ISSN: 0027-7738 (print) 1756-2279 (web)  Vol. 71 No. 2, Spring 2023  DOI 10.5195/names.2023.2534
Relevant constructs | 336 | 2 | 0 | 3
Personifications | 335 | 2 | 0 | 3
Percentage (%) | 99.6 | 100 | 0 | 100

The percentages in Table 4 show that practically all the extracted [ORG + V + that + SC] relevant constructs (99.7% of the total) are personifications. These percentages support the hypothesis that in EU discourse, the cognitive metaphor of personification is symbolized in measurable, specific argument-structure constructions nesting certain categories of names, such as ORG. Below we present a few alphabetically sorted examples of [ORG + V + that + SC] constructions which seem especially humanizing on the basis of the main verb and the content of the subordinate clause:

- **[ORG + ANTICIPATE + that + SC]:** The Commission anticipated that further contacts were most likely still to be made before a definitive conclusion on the compatibility of the rules could be reached.
- **[ORG + EXPLAIN + that + SC]:** The Court of Justice explained that, in order to qualify as a ‘trader’, the person concerned must be acting ‘for purposes relating to his trade, business, craft or profession’ or in the name of or on behalf of a trader.
- **[ORG + HOPE + that + SC]:** The Commission hopes that this information will be of assistance in informing the students about the actions taken to protect the environment and educational programmes initiated at EU level.
- **[ORG + RULE + that + SC]:** The European Court of Human Rights (ECtHR) ruled that the United Kingdom was in breach of Article 3 of Protocol No 1 of the European Convention on Human Rights in relation to prisoner voting rights.

The only example which did not qualify for personification was the pattern Commission + is + that, which was part of the following sentence:

At this moment in time, the information from the Spanish Government to the European Coordinator for the Mediterranean Corridor and to the Commission is that all the high speed lines listed above will be realised on time.

Because Commission is not the subject of the complex sentence, the pattern does not instantiate here an actual [ORG + V + that + SC] argument-structure construction. Besides, the verb to be is considered semantically too broad for signaling personification. In any case, it is highly interesting to observe that there is still a nuance of humanization in the use of prepositions in subject complements, as if the organizations were individuals exchanging facts in a conversation (the information from the Spanish Government to the European Coordinator for the Mediterranean Corridor and to the Commission).

### Constructional Analysis

The analysis of the constructions in this paper is first divided by subcorpus (non-translated, non-interpreted, and interpreted), showing the corresponding slot fillers for each of them. Then, it presents some wrap-up results built through generalization and schematization. To this end, we use a list of notation schemes and resources from the literature on constructional analysis:

- a. Horizontal charts (Tables 5, 6, and 7) have individual counts for the name and verbal slot fillers in each PETIMOD 3.0. subcorpus, as seen in Corpas Pastor (2015).
- b. The informal form-meaning constructional description use bidirectional arrows and International Phonetic Alphabet (IPA) transcriptions of Hoffmann (2017)
- c. Verbal definitions for describing the meaning poles are from the *Macmillan Online Dictionary*
d. Non-compositional meaning aspects are in curly brackets supported by concordances from the corpus (Corpas Pastor 2021a)

e. The terminology on schematicity distribution is from Traugott (2008: 27): constructs as tokens/data-points \((\text{the Commission concludes that the issue raised by the petitioner does not constitute a breach of EU law})\), micro-constructions as individual construction types \(([\text{Commission} + \text{CONCLUDE} + \text{that} + \text{SC}])\), meso-constructions as similar constructions \(([\text{nCOMMISSION} + \text{CONCLUDE} + \text{that} + \text{SC}])\), and macro-constructions as the highest level of schematicity \(([\text{nCOMMISSION} + \text{V.REASONING} + \text{that} + \text{SC}])\). According to Traugott, similar strings can ”participate in a number of different constructions, where syntax, meaning, and pragmatic function can be construed at various degrees of granularity, ranging from general classes to individual idiosyncratic combinations”.

Note, however, that Traugott uses the term ”macro-construction” to refer to the classic labels found in the literature (for example, transitive, resultative), whereas here it is mostly used to refer to Transversal constructions (highly schematic argument-structure constructions that appear in non-mediated [non-translated, non-interpreted] and mediated [interpreted] subcorpora at the same time) [see Section 4].

Right below each horizontal table, detailed graphs (Figures 2, 3, and 4) show the concrete forms and numbers of micro-constructions. In the non-translated subcorpus, only the ten most frequent constructions are shown for the sake of conciseness (Figure 2). The constructional notation system is based on Wasserscheidt (2019): \([\ ]\) = linguistic unit (construction); \(\_\) = slot; \(\_\_\_\text{SUBSCRIPT}\) = semantic restriction of a slot; CAPITALS = morphological paradigm; italics = fixed form.

**Non-translated English (N-T\_EN)**

<table>
<thead>
<tr>
<th>SLOT FILLERS (N-T_EN)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ORG</strong></td>
<td>Commission (271), Court (18), CJEU (14), the Commission services (5), the Court of Justice (3), Generalitat (2), the Court of Justice of the European Union (CJEU) (2), the European Court of Justice (2), Bundestag (1), Council (1), Council of Europe (1), ECB (1), EFSA (1), ESMA (1), European Court of Human Rights (ECtHR) (1), Ombudsman (1), SCHER (1), SCHEER (1), STECF (1), the Commission (1), the Commission’s services (1), the Committee on Plants, Animals, Food and Feed (1), the Court of Justice of the European Union (1), the Lyon administrative tribunal (1), the Munich Regional Court (1), VGH (1)</td>
</tr>
<tr>
<td><strong>V</strong></td>
<td>consider (87), note (54), conclude (38), recall (17), hold (16), believe (12), confirm (8), understand (8), observe (7), point out (7), stress (7), acknowledge (6), find (6), rule (5), suggest (5), underline (4), emphasise (3), expect (3), explain (3), reiterate (3), state (3), agree (2), assume (2), claim (2), inform (2), highlight (2), propose (2), reply (2), add (1), announce (1), anticipate (1), ascertain (1), clarify (1), declare (1), decide (1), ensure (1), estimate (1), hope (1), insist (1), recognise (1), recommend (1), report (1), request (1), trust (1)</td>
</tr>
</tbody>
</table>
Nine of the most frequent micro-constructions (221 out of 335 constructs) in the non-translated English subcorpus include the name Commission, which gives an idea of how relevant this body for the PETI Committee is. A petition may be sent to the Commission for preliminary investigation, with a view to assessing the subject matter of the petition in relation to relevant legislation and policies (Committee on Petitions 2018). The most common micro-construction is [Commission + CONSIDER + that + SC] (85 out of 335). It can be described informally as: FORM: [kəˈmɪʃən CONSIDER ðæt + SC] \(\leftrightarrow\) MEANING: ‘the Commission has a particular opinion about someone or something’ (see example 1):

(1) the Commission considers that the means of redress available under Spanish law are the most adequate mechanism to deal satisfactorily with the issues at stake.

The second most frequent micro-construction, [Commission + NOTE + that + SC], can be described like: FORM: [kəˈmɪʃən NOTE ðæt + SC] \(\leftrightarrow\) MEANING: ‘the Commission notices or realizes something {formal}’ (see example 2):

(2) The Commission notes that the petitioner has appealed various wolf hunting decisions before national Courts.

It can be observed that there is a considerable downsized of constructs from the first to the second micro-construction (85 for [Commission + CONSIDER + that + SC], 51 for [Commission + NOTE + that + SC]). This could mean that the most important function of the Commission in this context is to give a final opinion which could help the Petitions Committee to decide on a petition’s admissibility, whereas [Commission + NOTE + that + SC] signifies reasoning processes which are collateral to such consideration. Examples 3 and 4, preceded by the discourse marker “in the first place” and the “conclusion” title respectively, seem to confirm this intuition:

(3) In the first place, the Commission notes [that] the allegation brought forward by the petitioners that an additional update of the EIA Report would have been necessary, in particular to give
Conclusion The Commission considers that the competent national authorities are best placed to intervene in this instance and the petitioner is invited to make use of the options available under national law.

In third place, the construction [Commission + CONCLUDE + that + SC] (22 constructs) could be paired as: FORM: [kəˈmɪʃən CONCLUDE ðæt + SC] ←→ MEANING: ‘the Commission decides that something is true after looking at all the evidence it has’. It also appears to be closely related with textual sequencing and positioning in different petition sections. It is written right after discourse markers (“Given the above”) and often introduces the concluding section of the institutional documents (see Example 5):

(5) Conclusion Given the above, the Commission concludes that, while the suggestion of the petitioner could seem a good way to further reduce errors and criminal acts, further action by the Commission is not foreseen in this matter.

The fact that the micro-constructions [Commission + NOTE + that + SC] and [Commission + CONCLUDE + that + SC] are both preceded by a discourse marker signals the existence of more complex (and schematic) constructions like [___DM, Commission + V + that + SC], in which a previous discursive slot would support or validate the reasoning processes of the Commission. In any case, microscopic semantic differences between verbal slot-fillers are still noticeable. For example, [Commission + CONCLUDE + that + SC] is much less frequent than [Commission + CONSIDER + that + SC], which implies that it is important for the Committee to describe the Commission as an opinionative body all throughout the texts, and that its opinions are regarded globally.

Moving to a different named entity, the micro-construction [CJEU + HOLD + that + SC] (7 constructs) could be described as: FORM: [sɪdʒɪˈjuːjʊ CONCLUDE ðæt + SC] ←→ MEANING: ‘the Court of Justice of the European Union says that something is true {slightly formal}’. Previously collocated prepositional phrases introduced by in (in particular, in Case C-618/10…) hint again at the existence of more complex constructions. This finding, combined with the low frequency of the construction, can indicate that judicial decisions are not that relevant in the work of the Commission and the Committee. Jurisprudence seems relegated to topics with a serious judicial history at a national or European level prior to petition (see example 6):

(6) In Case C-618/10 Banco Español de Crédito, the CJEU held that the applicable Spanish payment order procedure was incompatible with Directive 93/13 / EEC in conjunction with the principle of effectiveness, since it did not provide for an ex officio control of unfair contract terms by the competent court.

Finally, it is worth comparing example 6 with the micro-construction [Court + HOLD + that + SC] (5 constructs). It carries a slightly different meaning: FORM: [sɪdʒɪˈjuːjʊ CONCLUDE ðæt + SC] ←→ MEANING: ‘the Court of Justice of the European Union says that something is true {formal} {referential}'. Onomastic variation is not arbitrary, as it upgrades the register and is employed anaphorically when the form Court of Justice of the European Union is written before. This could imply the existence of further meso-constructions (___N.COURT + HOLD + that + SC) in which name choice makes a difference in the pragmatics of the meaning pole, but not in the semantics, as these are still carried by the verb (see example 7):

(7) It is worth referring in this regard to the Court of Justice of the European Union judgment of 14 February 2019 [ . . . ] The Court held that the differential treatment resulted from the date of recruitment since that date determined whether the former or the new rules (on salary scale and classification) were applicable.
Non-interpreted English (N-I_EN)

Table 6: Onomastic and verbal slot fillers for [ORG + V + that + SC] in N-I_EN

<table>
<thead>
<tr>
<th>SLOT FILLERS (N-I_EN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORG</td>
</tr>
<tr>
<td>V</td>
</tr>
</tbody>
</table>

Figure 3: [ORG + V + that + SC] micro-constructions in N-I_EN

Even though there are only two instances of the analyzed construction in non-interpreted English, they point towards shared schematicity, rhetoricity, and modality contact. The micro-construction [Commission + CONSIDER + that + SC], which was the most frequent in non-translated English, appears again in this corpus. This could be a shared, rhetorical way of conceptualizing the institution as a reasoning body, represented by the highly schematic construction [\_\_N.COMMISSION + \_\_V.REASONING + that + SC] (see example 8, where double slashes represent sentence-like segments in oral transcription):

(8) // it should also be noted that this is an individual project in relation to which the Commission considers that national authorities are better placed to assess the factual situation and intervene if necessary //

The second construction, [Catalan Supreme Court + MENTION + that + SC], can also be linked to similar micro-constructions in the non-translated corpus (for example, [Court + CLARIFY + that + SC], [the European Court of Justice + DECLARE + that + SC]), and thus to the macro-construction [\_\_N.COURT + \_\_V.SPEECH + that + SC]. However, both its onomastic and verbal fillers are not to be found in the other subcorpora. Again, it can be deduced that filler choices cause changes in the semantics and pragmatics of the constructional meaning pole. The Spanish institution Catalan Supreme Court is a concrete, remarkable choice which could imply a referential/intertextual function, whereas MENTION could be an oral preference. The construction can be then described as: FORM: \[\text{Catalan Supreme Court} + \text{MENTION} + \text{that} + \text{SC}\] \(\leftrightarrow\) MEANING: ‘the Catalan Court refers to something in a written document without giving many details {referential} {oral}’ (see example 9):

(9) // so taking into account that the Catalan Supreme Court mentioned that they were... that possibly there was a breach of EU law and that the internal procedures are yet not finalised in
the Supreme Court I think it's not yet the moment to close that file and we need to keep it open //

Interpreted English (I_EN)

Table 7: Onomastic and verbal slot fillers for [ORG + V + that + SC] in I_EN

<table>
<thead>
<tr>
<th>SLOT FILLERS (I_EN)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ORG</strong></td>
</tr>
<tr>
<td>Commission (1), the European Commission (2)</td>
</tr>
<tr>
<td><strong>V</strong></td>
</tr>
<tr>
<td>believe (1), reply (1), say (1)</td>
</tr>
</tbody>
</table>

Figure 4: [ORG + V + that + SC] micro-constructions in I_EN

In interpreted English, the construction [European Commission + BELIEVE + that + SC] represents explicitation, a typical feature of mediated discourse, which tends to spell out words and meanings (example 10). The European Commission is a multi-word named entity which varies to the Commission or Commission in the non-mediated corpora, where its belonging to the EU apparatus is taken for granted. Apart from explicitation, simplification (another trait of mediated discourse) is observed in the filling of [European Commission + SAY + that + SC] in example 11. The short verb say does not appear in non-mediated personifications, where it is replaced by longer, more formal choices (acknowledge, declare, etc.). [European Commission + SAY + that + SC] is at the same time convergent with [Commission + REPLY + that + SC], holding the mediated discourse together and making it more uniform than non-mediated (example 12). They converge in the macro-construction [__N,COMMISSION + __V,SPEECH + that + SC] by means of semantic coincidence (the Commission is represented as a speaking entity in both cases).

(10) // the European Commission believes that effective application of EU environmental law and the obligation as an outcome of the relevant two European Directives should contribute to resolving this complex and critical situation //
(11) // the European Commission said that in 2014 when the licenses were given they claimed that there was a public consultation //
(12) // it was the case in the gold mining extraction facility in Skouries and both times the Commission replied that this is a national issue ... that this is an issue that should be tackled at the level of national courts //
General Results and Discussion

Our detailed analysis at the micro-constructional level has revealed through generalization a set of highly schematic argument-structure constructions which can be termed transversal. Transversal constructions are filled with variants of organization names in the noun slot (__n.commission, __n.court), and also with semantically human verbs (__v.belief, __v.reasoning, __v.speech). Although such name-verb interactions and semantic labelling was already described for ORG+V collocations in smaller corpora (Sánchez Rodas 2022b; Corpas Pastor & Sánchez Rodas in press), this is the first study which makes this idea extendable to argument-structure constructions and explores all variations of mediated and non-mediated discourse in the English Eurolect. Transversality can be taken as a clear sign of the prototypicality of personification in EU texts; it is not merely a stylistic device, but a form of constructing meaning and conveying the different functions and/or degree of involvement that institutions have in relation to their fellow organs of the European body. The differences in numbers between corpora (suggested in Table 8 below) do not contradict this idea, nor what is found in the literature, as they can be read as a sign of simplification, a feature inherent to mediated and/or oral texts.

Table 8: [ORG + V + that + SC] transversal constructions in PETIMOD 3.0. (English)

<table>
<thead>
<tr>
<th>TRANSVERSAL CONSTRUCTIONS</th>
<th>N-T_EN</th>
<th>T_EN</th>
<th>N-I_EN</th>
<th>I_EN</th>
</tr>
</thead>
<tbody>
<tr>
<td>[__n.commission + __v.reasoning + that + SC]</td>
<td>139</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>[__n.commission + __v.speech + that + SC]</td>
<td>112</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>[__n.court + __v.speech + that + SC]</td>
<td>19</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Additionally, at a lower position in the cline of schematization, we can find a complementary phenomenon symbolized by what we have called mediating constructions (Table 9). Mediating constructions are exclusive to one of the four subcorpora, and apart from informing on the functions of each institution, they add non-compositional, grammaticalized pragmatic nuances through their name-verb slot choices. Most examples seem to be related to translation and interpreting universals (see [European Commission + SAY + that + SC], which is {convergent, explicating, and simplifying} all at once), but in the non-interpreted corpus, we can also find one construction [Catalan Supreme Court + MENTION + that + SC] in which the name choice suggests the pragmatic necessity of integrating a Spanish judicial authority in a native English text. In fact, recent studies (Sánchez Rodas 2022a; Sánchez Rodas submitted) suggest that the more onymic constructional analysis gets, the easier it becomes to distinguish mediating pragmatics (for example, translations like Basque Department of Health, borrowings like Sauvons nos Enfants Japon, language blending like the ‘Petón do Lobo’ Environmental Association,) from non-mediating pragmatics (for example, fixed-form traditional organizations like British Airlines), and explain their individual reasons.

Although more data are still needed, all these findings open the way to uncover the constructional rules of a possible grammar of translation and interpreting, at least in terms of constructions with names in EU texts. Changes in individual named entities that from the point of view of the translation process were previously described as shifts (Corpas Pastor and Sánchez Rodas 2022) could actually indicate contrasts between constructions with different pragmatic nuances in their form-meaning poles, each one serving distinct mediating (negotiating, conflict-resolving, etc.) functions in their corresponding subcorpora. Moreover, examples like [Catalan Supreme Court + MENTION + that + SC] suggest the possibility that mediating constructions may actually not be exclusive to mediated texts. They could be related to the necessity of creating a consensus inherent to the construction of any institutional text, be it translated/interpreted or not. Their meaning pole could also instantiate additional pragmatic functions of varying complexity, such as intertextuality, terminological coherence across both non-mediated and mediated texts, or extralinguistic, political functions still to be unearthed.
Table 9: [ORG + V + that + SC] mediating constructions in PETIMOD 3.0. (English)

<table>
<thead>
<tr>
<th>Argument-Structure Constructions</th>
<th>N-T_EN</th>
<th>T_EN</th>
<th>N-I_EN</th>
<th>I_EN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Converging</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Commission + REPLY + that + SC]</td>
<td>N-T_EN</td>
<td>T_EN</td>
<td>N-I_EN</td>
<td>I_EN</td>
</tr>
<tr>
<td>Explicitating</td>
<td>N-T_EN</td>
<td>T_EN</td>
<td>N-I_EN</td>
<td>I_EN</td>
</tr>
<tr>
<td>Simplifying</td>
<td>N-T_EN</td>
<td>T_EN</td>
<td>N-I_EN</td>
<td>I_EN</td>
</tr>
<tr>
<td>Integrating</td>
<td>N-T_EN</td>
<td>T_EN</td>
<td>N-I_EN</td>
<td>I_EN</td>
</tr>
</tbody>
</table>

Conclusion

Our study has highlighted the importance of pushing forward the study of names (particularly NEs) from an isolated to a corpus-based, constructional approach. We have focused on argument-structure constructions with organization names in a corpus of institutional texts, namely EU petitions and related documents. Using the most updated technologies to extract them (corpus-based verbal pattern extraction combined with NLP-based named entity recognition), we have uncovered a network of argument-structure constructions with individual forms and meanings across four mediated and non-mediated subcorpora. On the one hand, such constructions conform a transversal scheme by which they tell the recipients about the relations and functions of each EU institution in a humanizing way, as if they would “consider”, “say”, or “hope” things, to cite some examples. On the other hand, at a more microscopic level, some appear exclusively in one of the subcorpora, as if their slot choices would serve specific negotiating functions related to the pragmatics of drafting institutional texts, be they mediated or not.

Although further quantities of data (that is, different argument-structure constructions or name types) and additional interlinguistic studies are still needed to determine the scope of such findings, we believe that this is a promising start in the description of a grammar of EU texts, at least in what refers to names. What is even more important is that such constructional description seems to be extendable to mediated texts, which could make the study of translationese and interpretese extremely fine-grained, changing the point of view from exclusively quantitative or qualitative approaches to a comprehensive stance including concepts of memetics, complexity, schematicity, networks among constructions in different corpora, and so on. Inarguably, the inclusion of names in the study of constructions becomes steadily unavoidable, especially in institutional translation and interpreting, given the different degrees of semantic accessibility their choice provides.

Endnotes

1 In the context of empirical translation studies, Kotze (2019: 346) states that one or more texts can be classified as mediated “in the sense that a prior text delimits and shapes the[ir] production”. This seems too broad a conceptualization for translation studies, and even for linguistics, as it could virtually comprehend all existing texts. We are more aligned with Stecconi (2009: 262–263), who considers mediation an existential condition of translation which differentiates it from other semiotic acts: “[…] it is logically impossible to label as translation a text that is not perceived as speaking on behalf of another – i.e. that does not mediate between source and target environments”. If we also consider the fact that mediated discourse can be written (translation) or spoken (interpretation), any given text in any language could be then classified into one of four basic mediation modes: non-translated, non-interpreted, translated, or interpreted.

2 The International Council of Onomastic Sciences (ICOS 2022) includes these names under the umbrella term “chrematonym”, which is the “name of a politico-economic or commercial or cultural institution or thing; a catch-all category”. Strictly speaking, they are a subtype of chrematonyms which sometimes receive special denominations in the literature, such as “institutiononyms” (Knappová 2017; Lücking et al. 2016; Majtán 1989), or simply, “organization names” (Sjöblom et al. 2012; Markert & Nissim 2008).
An intermodal corpus is “a resource that makes available for comparison samples of texts translated [and/or interpreted] in different modes” (Bernardini 2016: 129).

SpaCy and DeepPavlov are free open-source libraries in Python. The two pre-trained SpaCy-based models in VIP differ in the degree of granularity of the NER annotation scheme between languages (16 tags in English versus 4 tags in Spanish). DeepPavlov comes with a set of predefined components for solving NLP related problems and tasks, such as NER. For instance, it recognizes 18 NE tags in each language and uses classification models which are implemented as a number of different neural networks, such as BERT (Devlin et al. 2018) or sklearn models.

Transcription is a matter of relevance when identifying argument-structure constructions, as anomalies could impede the correct retrieval of complete constructs in corpus concordances. Most instances removed for this reason included joint footnote numbers (for example, the European Court of Justice considered that). Criteria for other types of evaluation, such as multi-word entity recognition scoring, might consider this a minor obstacle and correct it on the go, favoring the overall precision of the recognition system (for example, Italian Competition and Market Authority might be changed to Italian Competition and Market Authority and considered relevant).

The combination of bigram lists with concordance search was an effective solution for recall (Consider the SpaCy evaluation in N-T_EN, where 38.5 % of the total relevant constructs were recognized in this manner.). In fact, concordances also helped revise some precision results in which out-of-context pattern extraction had been misleading (for example, EC stipulates that vs. Article 8 of Directive 2004/38 / EC stipulates that, for periods of residence longer than 3 months...).

It was also observed that many of the consulted definitions directly signaled personification in their sentence examples (for example, the school agreed to send the students on the course) and that the dictionary structure relies on constructivism to a great extent. This can be seen in capitalized labels like INTRANSITIVE/TRANSITIVE or NOT USUALLY PROGRESSIVE included before verbal entries, or in sub-entries in bold letters, such as agree (that).

Acknowledgements

We would like to thank Professor I. M. Nick, the Editor-in-Chief of NAMES; and the anonymous reviewers for their invaluable help and constructive comments.

Funding

This work was supported by the Spanish Ministry of Universities (former Ministry of Science, Innovation, and Universities) under Grant FPU18/05803. It has also been carried out in the framework of various research projects on language technologies applied to translation and interpretation: VIP II (PID2020-112818GB-I00), Proof-of-concept VIP (PDC2021-121220-I00), and RECOVER (ProyExcel_00540).

Disclosure statement

No financial interest or benefit has arisen from this research.
References


Notes on Contributors

**Fernando Sánchez Rodas** is a Trainee Researcher with a grant awarded by the Spanish Ministry of Universities (formerly Ministry of Science, Innovation, and Universities). Currently, he is writing his PhD thesis on onomastic constructions in the EU multilingual discourse under the direction of Professor Gloria Corpas Pastor and is collaborating on the research projects of the Lexytrad group from the University of Malaga. His interests include corpus-based onomastics applied to translation and interpreting, with a special focus on christonomy.

**Gloria Corpas Pastor** is Visiting Professor in Translation Technology at the Research Institute in Information and Language Processing (RIILP) of the University of Wolverhampton, UK (since 2007); Professor in Translation and Interpreting at the University of Malaga, Spain (since 2008); honorary Adjunct Professor at Xi’an Jiaotong-Liverpool University, China (since 2020); and Director of the Research Institute in Multilingual Language Technologies (IUITLM) of the University of Malaga (since 2019). Published and cited extensively, her research covers translation and interpreting technologies, corpus linguistics, and contrastive and computational phraseology, among other topics.

**Correspondence to:** Professor Gloria Corpas Pastor, Research Institute in Multilingual Language Technologies, Malaga, Spain. Email: gcorpas@uma.es