Analyzing Environmental Discourse through Construction-Based Pattern Extraction

Elisa Chierchiello[1], **Eliana Di Palma**[1], **Ludovica Pannitto**[2], **Cristina Bosco**[1]
[1] University of Torino
[2] University of Bologna
{elisa.chierchiello, eliana.dipalma, cristina.bosco}@unito.it, ludovica.pannitto@unibo.it

Abstract

Environmental issues are at the centre of a debate currently taking place across all communication channels. This paper provides an analysis of texts in which these issues are discussed, with the novelty of applying a methodology that enables the extraction and comparison of different narratives and points of view. The texts used in this study are the English *Living Planet Reports* published biennially by the WWF from 2014 to 2024. The methodology is based on the extraction of constructions – patterns collected in the English constructicon CASA – which allow us to identify differences in the presentation of the issues discussed in the analysed texts. Our results show that this methodology can be very helpful in the comparative analysis of texts to reveal different perspectives, for example, to observe diachronic variations.

Keywords: environmental discourse, construction grammar, Grew, Universal Dependencies, corpus linguistics

1. Introduction

Environmental issues are increasingly debated by various public and private entities, as well as citizens, in all countries and across multiple communication channels and media. The complexity inherent to most of these issues, such as pollution, loss of biodiversity, and global warming, makes especially hard to discuss them or to recognize and compare the diverse opinions expressed when they are discussed. Different perspectives can express opinions related to varying urgencies and perceptions, with different impact on the reader and listener. For this motivation, not only the life sciences, but also disciplines such as computational linguistics are expected to address the challenges posed by environmental discourse by providing new approaches for their analysis.

In recent years, constructionist approaches have been increasingly applied across a wide range of subfields. In particular, Construction Grammar (CxG, Fillmore et al. 1988; Goldberg 2013) frameworks have proven effective in the study of information structure and discourse patterning (e.g., Lambrecht 1994; Östman 2005), as well as in research on context and pragmatics (e.g., Jing-Schmidt 2017). They have also been successfully employed in dialogue and conversation analysis (e.g., Linell 2009), and in investigations of register and genre (e.g., Fischer and Aarestrup 2021; Antonopoulou and Nikiforidou 2011; Nikiforidou and Fischer 2015).

A key principle of CxG is that every human language can be understood as a richly structured, semantically motivated, and constantly evolving inventory of symbolic units known as *constructions*. Each construction consists of a pairing between a specific form and a corresponding meaning. These form–meaning pairings vary widely in both size and level of abstraction: they range from simple morphemes such as *book* to larger phrasal units like *you can't teach an old dog new tricks*, and from fully fixed expressions such as *break the ice* to highly schematic patterns like $SUBJ_{agent}$ V $OBJ_{recipient}$ OBJ_{theme} (i.e., the English ditransitive construction, as in *She gave him a gift*).

In this paper, we apply a novel methodology for the automatic extraction of constructions to texts discussing environmental issues. The central hypothesis is that constructions can provide valuable insights into how environmental topics are framed. As a preliminary step, the study focuses on developing this extraction method, which is innovative in both environmental linguistics and the computational study of construction grammar. It then presents an analysis of constructions identified in the WWF corpus, offering initial evidence of the method's potential for examining environmental narratives and discourse.

As far as the texts, we use the collection of WWF's reports called *Living Planet Reports* (WWF-LPR) which are published every two years on the website of the international organization. As a case study we use the English data published from 2014 to 2024. These reports have been used in previous studies (Bosco et al., 2025; Chierchiello et al., 2025) and made available in Universal Dependencies (UD, De Marneffe et al. 2021) format in the multilingual treebank for environment TreEn (Pagano et al., 2025). UD is a cross-linguistically consistent framework for morpho-syntactic annotation that represents sentences as dependency trees, providing standardized information about grammatical rela-

Proceedings of the 2nd Workshop on Ecology, Environment, and Natural Language Processing (NLP4Ecology) @ LREC 2026, pages 87–104
May 12, 2026

tions and lexical features.

Taking advantage of the available morphological and syntactic annotations, our approach is grounded in the Construction Grammar framework. Considering that the environmental discourse is rich in recurrent constructions that carry stance, causality, and evaluation, we developed scripts to extract all the available constructions from the collected texts benefiting from their availability in UD format. Nevertheless, constructicons (Lyngfelt et al., 2018) are not directly searchable in UD corpora due to the lack of a shared UD-level formalization. To address this limitation, we focus on the English CASA (Herbst and Hoffmann, 2018) constructions inventory and we operationalize the defined constructions as `Grew` graph patterns (Guillaume, 2019) over UD parses.

The main **contributions** of this study are:

- A reproducible method to operationalize CASA constructions for the analysis of discourse;
- A validated constructional inventory that makes it possible to systematically study how sustainability is expressed in language;
- A corpus-based analysis of WWF's reports illustrating how constructional patterns shape environmental narratives.

2. Related Work

2.1. Computational Construction Grammar

A key theoretical aspect of CxG is the idea of language as a network of constructions: a *constructicon*. In recent years, however, another type of constructicon has been coming to the fore: an archive of construction descriptions, a sort of construction dictionary, halfway between CxG and lexicography (Lyngfelt et al., 2018). The idea was first introduced by Fillmore (2008), who initiated the development of an English constructicon (Fillmore et al., 2012; Lee-Goldman and Petruck, 2018) as a complement to the English FrameNet. Since this first step for English (followed by more recent resources such as *The English Constructicon* Perek and Patten 2019 and CASA, employed in this work, Herbst and Hoffmann 2018), much work has been done in the creation of constructicons for different languages: Brazilian Portuguese (Laviola et al., 2017), Swedish (Gruzitis et al., 2015), German (Ziem and Boas, 2017), Russian (Janda et al., 2018), and Japanese (Ohara, 2013), and more currently being developed.

Constructicons are systematic collections of language-specific constructions, in which the syntactic form and the corresponding meaning are reported for each construction. Unfortunately different notations and labels were used in different constructicons that make difficult to generalize constructions both intralinguistically and interlinguistically. Another challenge is the automatic extraction of constructions from texts. Similarly to our approach, (Lorenzi et al., 2023) conducted experiments on datasets annotated with UD, aiming to understand the impact of UD and frame information on the labeling of constructions elements, focusing on clausal constructions in Brazilian Portuguese (Torrent et al., 2018). Using both LSTM and Transformer-based models, they show that UD annotations and frame information can boost performance. However, their work analyzed only clausal constructions, whereas our methodology refers to all types of constructions, enabling us to base discourse analysis on all the extracted constructions.

2.2. `Grew`

`Grew`[1] (Guillaume, 2019) is a graph rewriting tool designed to manipulate and query linguistic representations encoded as graph structures. Through the `Grew` query language, users can search for occurrences of specific graph patterns within treebanks by specifying constraints on the nodes and relations involved in the query. Unlike simple keyword search, `Grew` supports graph-level queries, allowing users to precisely describe the linguistic structures they wish to retrieve from a corpus.

Both Weissweiler et al. 2024 and Pisciotta et al. 2025; Pannitto et al. 2024 introduced `Grew` queries as possible formalizations of construction-like objects.

2.3. NLP for Environmental Discourse

The application of NLP to environmental discourse has been mainly centered on issues related to climate change, while Sentiment Analysis has been the most used technique. According to a systematic review (Ibrohim et al., 2023) studies that have dealt with environmental issues using NLP are few and based on very coarse-grained polarity analysis, even for English. In a similar vein, Stede and Patz 2021 examine the application of Sentiment Analysis to climate change discourse, pointing out that different groups (the general public, policymakers, and scientists) use different genres, registers, and terminologies. Other studies applied Sentiment Analysis to topics such as smart cities, urban policies, and energy debates on social media platforms (Du et al., 2020).

Environmental NLP has steadily emerged as a research field, with workshops, shared tasks, and

---

[1] `https://grew.fr/`

domain-related language models such as ClimateBERT (Webersinke et al., 2022). However, model-driven and topic-focused approaches are used without much emphasis on the systematic study of linguistic constructions. Released corpora include an Italian resource for environmental analysis (Grasso et al., 2024a), some multilingual efforts (Bosco et al., 2023), and diachronic climate-focused corpora for English (Grasso et al., 2024b). Among the very few annotated resources the InsightsNet Climate Change Corpus (ICCC, Bartsch et al. 2023) has to be mentioned, which provides metadata, morphological, syntactic, and named entities, all automatically annotated. The only reported treebank about environmental issues is currently TreEn (Pagano et al., 2025) which is released in UD format.

Building on a corpus derived from WWF's reports, which spans a broader range of environmental issues instead of simplifying discourse analysis to sentiment polarity classification, we take a constructionist approach to identify recurring patterns of linguistic expressions that constitute environmental discourse.

3. Data: WWF's *Living Planet Reports*

The corpus used for our research is based on the WWF-LPR[2], a set of policy reports on biodiversity loss, ecosystem degradation, climate change, resource management, and sustainability transition that have been widely distributed across the globe in a large number of languages also mirroring cultural differences. The WWF-LPR presents an environmental story that encompasses ecological, economic, and socio-political elements, that makes them a strategic tool in the shaping of global environmental communication. As mentioned above, this corpus is part of TreEn (Pagano et al., 2025), a multilingual treebank project focused on environmental discourse and annotated within the UD framework. However, in the current contribution, we exclusively focus on the English subcorpus of this resource, and we take into account only data generated in the time span 2014-2024. The resulting dataset comprises six report editions, totaling 64,345 tokens and 2,713 sentences. Texts are automatically annotated and manually checked to conform to UD guidelines, which allows fine-grained morpho-syntactic querying. This layer of structured annotation enables us to query constructional patterns via `Grew` queries and examine how environmental narratives are linguistically encoded.

Through the lens of a UD-annotated environmental corpus, our goal is to look beyond the surface level of sentiment or stance classification and instead explore the structural patterns by which environmental meaning and argumentation are realized.

4. Method: From CASA to UD-based `Grew` Queries

In this study, the term "construction" does not refer to an abstract syntactic unit in the traditional grammatical sense, but to a formally operationalized pattern derived from CASA and implemented as a UD-based graph query in `Grew`. This section outlines how qualitative constructional descriptions were mapped onto dependency-based graph patterns.

4.1. Selection of Constructions

The constructions were taken directly from the CASA Constructicon[3] (Herbst and Hoffmann, 2018), which was chosen because of its wide coverage and availability as a structured inventory in a machine-readable format. Moreover, the qualitative descriptions of constructions can be easily transformed into dependency-based patterns. The inventory used is the revised version from December 2025. We mapped each construction's slot onto UD relations and pertinent morphosyntactic properties after looking at its syntactic schema and semantic function. As a result, constructional descriptions and graph-based representations were explicitly aligned. The result is a CASA inventory that has been UD-operationalized, allowing constructs to be automatically extracted using graph rewriting techniques.

4.2. Operationalization Principles

The operationalization of the CASA constructions into UD-based graph patterns followed a set of consistent principles.

First, the syntactic head of a construction corresponds to the UD head of a phrase, since we adopted the same UD head-selection criteria.

Second, whenever possible constructional slots were mapped to dependency relations. Other morphosyntactic constraints (e.g., `VerbForm=Inf`, `Degree=Cmp`, clause type) were encoded as feature-based restrictions, and argument roles like *subject*, *object*, *complement*, or *modifier* were translated into their corresponding UD relations. Also, we gave structural arrangements precedence over linear order and instead of capturing surface sequences, the patterns record hierarchical relationships since UD encodes syntactic structure regardless of word order.

Lexical restrictions, lastly, were only incorporated

---

[2] `https://livingplanet.panda.org/`

[3] `https://constructicon.de/`

when they were specifically mentioned in the original composition (e.g., fixed coordinators like *and*, certain auxiliary verbs). Discourse-pragmatic subtleties and intonation, which are not technically stored in UD, were not operationalized.

4.3. Query Design in `Grew`

Core relations. The queries were written applying a limited but well chosen set of core UD dependent relations that emerge across constructions. We relied on relations like `root`, `subj`, `obj`, `iobj`, `xcomp`, `ccomp`, `advcl`, `acl`, `nmod`, `obl`, `amod`, `compound`, `conj`, `cc`, `mark`, and `case`. The main idea was to give each construction a structure that is easy to recognise, regardless of word order or language.

Constraints. Structural relations were merged with morphosyntactic restrictions expressed over UD features, such as `upos`, `lemma`, `VerbForm`, `Tense`, `Degree`, and `PronType`. The construction's grammatical profile is refined by feature constraints, whereas dependency relations capture its basic structure. These restrictions enable us to distinguish, for instance, between declarative and interrogative sentence types, comparative forms and positive adjectives, and infinitive and finite clause complements.

Coverage vs precision trade-offs. Designing the queries required a constant balance between recall and precision. Optional elements (e.g., modifiers), attachment ambiguities, and cross-linguistic variation can easily lead to under-generation (missing valid instances) or over-generation (including structurally similar but constructionally unrelated cases). To solve this, we gave priority to the fundamental structural arrangement that characterizes every structure, treating auxiliary components as optional whenever feasible. Especially in big, diverse corpora, this approach tolerates syntactic heterogeneity while maintaining the constructional backbone.
Figure 1 illustrates the operationalization of construction [66], *indefinite_plural*. In UD terms, the construction is identified as a plural noun that does not license a determiner dependency. The absence of a `det` relation therefore serves as the formal constraint distinguishing indefinite plural noun phrases from definite or determiner-marked ones.

Indefinite Plural NP Construction
NOUN Number = Plur
Forbidden dependency:
$\nexists\, D, \texttt{det(NOUN, D)}$

Figure 1: Indefinite Plural NP Construction, CASA ID: 66

The corresponding `Grew` query implementing this constraint is shown below:

```
pattern {
  N [upos="NOUN", Number="Plur"];
}
without {
  D [];
  N -[det]-> D;
}
```

The query, performed on a set of UD parsed sentences, returns a list of hits each composed by the sentence ID an a list of labeled nodes matching the ones declared in the `Grew` pattern.

5. Construction Inventory and Documentation

In CASA, each construction is identified by an ID and the corresponding documentation can be accessed via the CASA Constructicon website.[4] In the remainder of this paper, CASA IDs will be reported in square brackets.

5.1. Construction Families Covered

The construction inventory examined in this study includes recurrent syntactic patterns relevant to the encoding of environmental narratives in policy discourse. We selected construction families covering aggregation, comparison, modification, coordination, and argument structure, which are central to the rhetorical strategies of this genre.
In particular, we examine 133 constructions, among:

- **Noun Phrase constructions**, like pre- and post-modification structures, which play an important role in the compact formulation of environmental ideas (e.g., *"environmentally friendly", "causes of unsustainable development"*).

---

[4] `https://constructicon.de/constructions/[ID]`, where [ID] is the construction identifier.

- **Connection constructions**, like [146] additive, [147] alternative, and [145] asyndetic coordination patterns, often employed to group drivers, impacts, and policy responses (e.g., *"positive , equitable, and sustainable"*).
- **Comparative constructions**, including [116] morphological comparatives and also *more/-less ... than* patterns, encoding escalation, urgency, and evaluation (e.g., *"higher than it would be"*).
- **Argument-structure and causative patterns**, which are part of the representation of agency and responsibility in environmental processes (e.g. *"world must move towards a just transition"*).

5.2. Release Format

The release consists of the outputs of the `Grew` queries applied to the WWF-LPR corpus.

The outputs are provided as structured tables (in spreadsheet format), where each row corresponds to a matched instance and includes information on the construction label, sentence identifier, matched token span, head token, and relevant slot assignments.

File naming follows the original CASA construction identifiers, ensuring traceability between the construction inventory and the extracted instances.

The dataset is archived on ZENODO[5] and is available at: https://doi.org/10.5281/zenodo.19256171.

6. Corpus-based Evaluation

All the experiments are based on the English section of the WWF-LPR corpus as released in TREEN (Pagano et al., 2025).

Each construction is implemented as one or more `Grew` queries, according to the operationalization principles in Section 4. Queries were run independently over the entire corpus. For constructions operationalised using multiple query strategies (such as comparative alternations), matches returned by different queries were combined at the extraction stage under a single construction label. All extraction outputs were exported as structured tables including: (i) construction label, (ii) sentence identifier, (iii) matched token span, (iv) head token, (v) relevant slot assignments.

6.1. Quantitative Measures

To describe how constructional patterns are distributed across the corpus, we combine absolute and normalized frequency measures.

Frequency measures. For each construction, we calculated its raw frequency (total number of hits returned by the query in the corpus) and its normalized frequency, expressed as occurrences per 10,000 tokens, to enable comparison across subcorpora of different sizes.

$$\text{NormFreq}_{c,y} = \frac{N_{c,y}}{T_y} \times 10{,}000 \qquad (1)$$

where $N_{c,y}$ is the number of matches for construction c in year y, and T_y is the total number of tokens in the corresponding subcorpus.

Sentence coverage. We measured the number of unique sentences containing at least one occurrence of a given construction. This indicates how pervasive the construction is in discourse, regardless of how many times it appears in a sentence.

$$\text{Coverage}_{c,y} = \frac{S_{c,y}}{S_y} \qquad (2)$$

where $S_{c,y}$ is the number of sentences in year y containing at least one instance of construction c, and S_y is the total number of sentences in the same subcorpus.

Construction density. We also calculated the average number of constructions per sentence to estimate the structural complexity of environmental narrative passages.

$$\text{Density}_{y} = \frac{M_y}{S_y} \qquad (3)$$

where M_y is the total number of construction matches in year y, and S_y is the total number of sentences in the same subcorpus.

6.2. Manual Validation Protocol

To assess the reliability of the operationalization process, we conducted a manual evaluation of the constructions extracted by the automated procedure. All constructions included in the inventory were subject to validation. The aim of this evaluation is not to provide exhaustive annotation of all extracted instances, but rather to estimate extraction precision while capturing variability across construction types.

For each construction, up to five instances per year were randomly sampled from the `Grew` output, resulting in a maximum of 30 instances per construction across the six reports (2014–2024). For constructions with fewer than five instances available in a given year, all available instances were included, leading to a smaller overall sample size.

---

[5] https://zenodo.org/

Sampling was stratified by year and performed over the full set of extracted instances for each construction, ensuring coverage across all subcorpora. Instances were randomly selected from the pool of matches returned by the queries without additional filtering.

Each sampled instance was independently annotated by two annotators, who judged whether the extracted instance corresponded to a valid realization of the target construction. In cases of disagreement, a third annotator adjudicated the final decision.

Full validation results for all evaluated constructions are reported in Appendix A.

Precision estimation. Extraction precision was calculated as the proportion of correctly validated instances over the total number of annotated instances for each construction:

$$\text{Precision} = \frac{C}{N} \tag{4}$$

where C is the number of validated instances and N is the total number of annotated instances for that construction.

Uncertainty estimation. Because the evaluation is based on a limited sample (up to 30 instances per construction), the reported precision values represent estimates rather than full-corpus performance measures. To quantify sampling variability, 95% confidence intervals were computed under a binomial approximation.

The standard error (SE) of the precision estimate is defined as:

$$SE = \sqrt{\frac{p(1-p)}{N}} \tag{5}$$

where p denotes the observed precision. The 95% confidence interval is then obtained as:

$$CI_{95\%} = p \pm 1.96 \cdot SE \tag{6}$$

We emphasize that the evaluation is intended as an initial validation of the operationalization approach rather than a comprehensive assessment of extraction performance. The results nevertheless provide useful indications of the strengths and limitations of query-based construction extraction.

7. Results

This section presents the quantitative and qualitative evaluation of the results of the UD-based operationalization of CASA constructions applied to the WWF-LPR.

7.1. Extraction Precision

In Table 1, we present the precision values obtained from the manual assessment of the sampled constructions. In general, the extraction process demonstrates high reliability for most constructions. For many patterns, we observe high precision values, often around 1.00, for the sampled instances, such as [69] *definite_plural_np*, [68] *definite_singular*, [136] *progressive*, [92] *genitive_np*. A large number of the grammatical patterns which are more central to the descriptive structure of the WWF-LPR show high precision values, e.g. [120] *declarative_statement*, [151] *adjunct*, and [109] *valency_postmodifier*. The observed precision values vary considerably across constructions, reflecting differences in structural ambiguity and the degree of formal constraint in their operationalization. Constructions anchored in stable syntactic dependencies tend to achieve higher precision, whereas more schematic patterns, which rely on broader matching conditions or lexical variation, are more prone to noise. In particular, lower precision values are observed for constructions involving named entities and surface patterns associated with semantically ambiguous configurations, where the interpretation depends more heavily on lexical choice or discourse context. These results highlight the trade-off between coverage and precision in query-based extraction, and underscore the importance of construction-specific tuning.

7.2. Overall Extraction Statistics

Table 2 presents the sentence-level coverage of the selected constructions across the WWF-LPR. In addition to sentence-level coverage, normalized frequency confirms the same tendency. Across all editions, [106] *premodifier_of_noun* emerges as the most pervasive pattern, with coverage values consistently above 0.85 across all years. This stability highlights the strongly nominal and information-dense character of the genre. The consistently high presence of [109] *valency_postmodifier* and related postmodification patterns [107, 108, 113] suggests that noun phrases are not only frequent, but structurally complex, often combining pre- and post-modification within the same syntactic unit.

A gradual increase in the use of [66] *indefinite_plural* and [67] *mass_np* constructions also suggests a tendency towards abstraction and systemicity. These constructions have a tendency to refer to collective entities and unbounded conceptual domains, which again emphasizes the macro-level perspective of the reports.

Verbal patterns like [120] *declarative_statement* and [3, 4] *monotransitive* are relatively stable across years, showing that even though the reports are assertive and informationally dense, they are not nar-

Table 1: Extract of precision results for selected constructions (eq. 4). N is the number of annotated instances and C the number of instances judged correct. Confidence intervals are computed under a binomial approximation (95%).

Group	**Construction**	C	N	**Precision**	**95% CI**
Best	definite_plural_np	30	30	1.000	[0.884–1.000]
	declarative_statement	29	30	0.967	[0.903–1.000]
	adjunct_base	28	30	0.933	[0.844–1.000]
Intermediate	additive_coordination_base	21	30	0.700	[0.536–0.864]
	monotransitive_clausal_objects	20	30	0.667	[0.498–0.836]
	to_recipient_goal	14	26	0.538	[0.346–0.730]
Lowest	full_name	2	18	0.111	[0.000–0.256]
	on_focus_area	11	30	0.367	[0.194–0.540]
	first_name_np	0	30	0.000	[0.000–0.116]

Table 2: Sentence coverage (eq. 2) of selected constructions in the WWF-LPR (2014–2024).

Construction	**2014**	**2016**	**2018**	**2020**	**2022**	**2024**	**avg**
premodifier_of_noun	0.848	0.937	0.865	0.885	0.903	0.917	0.893
indefinite_plural	0.673	0.786	0.741	0.716	0.794	0.831	0.757
mass_np	0.502	0.615	0.490	0.657	0.597	0.693	0.592
valency_postmodifier	0.591	0.651	0.721	0.662	0.723	0.661	0.668
declarative_statement	0.494	0.442	0.410	0.397	0.511	0.468	0.454
monotransitive	0.401	0.341	0.347	0.319	0.409	0.421	0.373

rative in nature. Rather, the discourse is organized around general statements, systemic descriptions, and evidence-based claims.

Apart from the high-frequency constructions, a significant part of the data set has zeros or near-zeros coverage across the board for all years. This is not necessarily meaningless data; it can represent genre constraints as well as methodological challenges in operationalizing CASA constructions in UD-based queries. Furthermore, the limited size of the corpus under examination may have influenced the results of the extraction.

First, interactional constructions such as [122] *yes/no questions*, [123, 124] *wh-questions*, and so on are absent from most reports. This is to be expected, given the institutional and expository nature of the WWF-LPR, which prefers informational statements to interaction.

Second, stance markers (e.g., hypothetical or predictive forms) occur sporadically in the corpus (one or two occurrences in particular years), which suggests that epistemic hedging and conditional framing are not important in this genre and are probably restricted to local argumentation. Thirdly, first-person constructions are used sparingly overall, especially first-person singular constructions. Generally, the reports refrain from using first-person constructions in favor of impersonal constructions, nominalizations, and passive constructions. When first-person plural constructions are used, they often have institutional rather than personal connotations.

Table 3: Construction density (eq. 3, matches per sentence) in the WWF-LPR.

Year	M_y	**Tokens**	**Sentences**	**Density**
2014	4156	5373	257	16.17
2016	6253	7655	364	17.18
2018	4529	5870	251	18.04
2020	6251	7777	373	16.76
2022	9449	11864	462	20.45
2024	20114	25806	1006	19.99

7.3. Construction Density

We also computed the average number of construction matches per sentence (Density$_y$; Eq. 3) to estimate the “density” of constructional information encoded in each sentence in the WWF-LPR subcorpora. In Table 3, we can see that density remains relatively stable across the earlier reports (16.17-18.04 matches per sentence for 2014-2020), and then surges in 2022 (20.45 matches), remaining high in 2024 (19.99 matches).

Nominal density and descriptive packaging. In all years, the preference is strong for dense noun phrases, packaging premodification (for example, *premodifier_of_noun*) with postmodifying complements (for example, *valency_postmodifier*).

(premodification): “environmentally friendly food production”

(postmodification): “the underlying causes of unsustainable development”

Assertion-driven discourse and agency management. The prevalence of *declarative_statement* is coupled with a high incidence of elementary argument-structure patterns (*monotransitive*, [2] *intransitive*) that correspond to the structure of an exposition centered on declarative statements. Passive clauses ([138], *passive_construction*) are less frequent but of great importance, as they tend to focus on the result and the process rather than the actor, especially in cases where they are generic or institutionally diffuse.

(declarative statement): “Human wealth depends on nature’s health.”

(passive): “In this region, climate change is more frequently reported as a threat to populations in the LPI.”

Projection and conditional framing. Modal and hypothetical sentences, although less common, mark instances where the discourse shifts from description to projection. These sentences introduce a reasoning process, which is often conditional, projecting the environmental changes in terms of consequences to be taken into account. The relative infrequency of these sentences indicates that the reports emphasize assertive description over projection.

(hypothetical): “If we lived the lifestyle of a typical resident of the US, we would need 3.9 planets.”

7.4. Error Analysis and Discussion

Although overall extraction performance is consistent, the analysis of mismatches reveals some main sources of errors and limitations, offering useful directions for future work.

UD annotation inconsistencies. A first possible cause of discrepancy relates to the interaction of clause embedding and root assignment in UD. The extraction of *declarative_statement* directly targets finite verbs that are marked as syntactic roots in the sentence. The applicable `Grew` pattern is:

```
pattern {
  ROOT [];
  V [upos=/^(VERB|AUX)$/, VerbForm="
    Fin"];
  ROOT -[root]-> V;
  S [];
  V -[nsubj|nsubj:pass]-> S;
}
without { V [Mood="Imp"]; }
```

This way, only finite verbs functioning as the matrix root are extracted, excluding subordinate clauses by definition. However, take the following example:

“The result is that we miss the message.”

In the UD representation of this sentence, the verb *miss* is marked as the syntactic `root`, whereas in a canonical UD analysis the clause *that we miss the message* would normally be attached to the copular predicate *is* as a `ccomp`. Consequently, the `Grew` query correctly extracts this occurrence, since the structural requirement ROOT–root–V is met. The problem, therefore, is not in the definition or the logic of the query, but in the representation of the clausal hierarchy as a dependency structure.

Proper names and nominal classification. Another issue is related to proper names and how they combine with constructions like [101] *first_name_np*. The `Grew` query uses morphosyntactic features like `upos=PROPN` to constrain the search space. However, this feature does not make a distinction between personal names and other named entities. For example:

“The Living Planet Index (LPI) has declined by 52 per cent since 1970.”

In this case, the term *LPI* has the part-of-speech tag `PROPN`, making it relevant to the query, but it does not refer to a personal first name but to an institutional index. Although the extraction is formally correct at the morphosyntactic level, it is semantically not correct for the constructional definition. This could be ideally tackled by enriching queries with external knowledge bases such as WikiData[6].

Semantically constrained constructions. Some patterns cannot be operationalized purely on the basis of structural constraints (e.g., preposition + noun), but also need some semantic or lexical information. For example, the extraction of the pattern [39, 111] *with_partner* should not be

---

[6] https://www.wikidata.org

formalized solely based on the presence of a prepositional phrase and *with* as the head word, as many other unrelated patterns (instrumental, manner, risk, etc.) would also match. Instead, the pattern should be filtered to include words related to collaboration and partnership (e.g., *partner*, *stakeholder*, *organization*).

> **(wrong with partner extraction)**:
> "Species are threatened with the risk of extinction."

Also expressions like [37] *on_focus_area* cannot be analyzed in terms of the structural template *on* + *NP*, but rather require semantic filtering to select cases where *on* expresses topical focus rather than spatial/temporal relations. As already proposed in Pisciotta et al. 2025 in a pilot experiment on Italian, filtering results through WordNet synsets could represent an effective solution.

Structural limits of pattern matching (coordination). Lastly, among the patterns we examined there are constructions sensitive to structural complexity, e.g. additive coordination. The `Grew` patterns usually capture binary coordination configurations; that is, they involve two conjuncts in the coordinated structure. However, there is limited ability to capture more complex coordination chains, with three or more conjuncts. As a result, there is limited ability to capture multi-member coordination.

Lessons learned Constructions based on solid and well-defined dependency relations can be reliably identified in the UD-based operationalization, such as basic noun phrase modification, argument structure constructions like monotransitive and intransitive constructions, and declarative statements. All these exhibit high precision levels. More complex patterns, however, are difficult to handle and hard to model exhaustively in a single query, such as coordinations (especially with more than two conjuncts), long-distance dependencies, and clause embedding, which are more sensitive to root assignment and hierarchical representation in UD.

Two classes of constructions emerge: those with a transparent structure in UD that easily correspond to a simple graph configuration, and those without a direct correspondence to a graph. While for the former the link between CASA descriptions and dependency relations is robust and replicable, it is more challenging to extract the latter.

Furthermore, annotation inconsistencies and parser artifacts are relevant to constructions that rely on fine-grained morpho-syntactic information. Lastly, semantically restricted constructions, such as partnership or topical focus, cannot be fully expressed using only structural information.

8. Conclusion and Future Work

In this paper, we introduced a novel methodology based on the extraction of constructions and evaluate it on texts about environmental issues. Our main goal is to provide preliminary evidences to confirm the hypothesis that constructions can provide valuable insights into how environmental topics are framed. The methodology consists in semi-automatically convert qualitative construction descriptions, commonly found in constructicons such as CASA, into formal graph-based queries in `Grew`. The results of our queries appear to faithfully describe the structure of narratives in a selection of data from the TREEN corpus, the first and currently only available treebank for environment, indicating that our approach may be useful for the analysis of environmental discourse.

Our annotation procedure only involved precision: namely, how many hits among the ones matched are true cases of the examined construction. In future work we will expand the methodology to recall as well, trying to spot *false negatives*: cases of construction instances that the `Grew` pattern wasn't able to match. We also intend to apply our methodology on additional datasets of environmental and policy discourse to test its robustness on different datasets and genres. Another line of work will be related to extending our approach to other constructicons and checking its portability across languages, examining the potential of CASA-based constructions in UD-annotated corpora of other languages.

Limitations

Several limitations need to be pointed out.

Firstly, the precision figures are based on a relatively small manually annotated dataset (up to 30 instances per construction), which is more an approximation than a validation on the entire corpus.

Secondly, the approach is dependent on the quality of UD annotation. Inconsistencies in the annotation of dependencies or roots will impact the extraction results, especially for those constructions which are sensitive to the hierarchy or structure of the clauses.

Thirdly, the study is limited to the constructions included in the CASA Constructicon. CASA has developed a list of constructions in a structured and linguistically well-motivated way. However, the list is not exhaustive and covers only the whole spectrum of constructions which can possibly play a role in environmental discourse.

Lastly, the empirical study is based on a single genre of institutional discourse, WWF-LPR, limiting the immediate generalizability of the study to other types of discourse.

Acknowledgements

The work of Eliana Di Palma was funded by the project NEIKEA (CSP TRAPEZIO Call – Line 1: Paving the Way to Research Excellence and Talent Attraction).

We would like to thank Beatrice Bernasconi for her valuable support and guidance on constructional approaches, which was instrumental in the development of this work.

Bibliography

Eleni Antonopoulou and Kiki Nikiforidou. 2011. Construction grammar and conventional discourse: A construction-based approach to discoursal incongruity. *Journal of Pragmatics*, 43(10):2594–2609.

Sabine Bartsch, Changxu Duan, Sherry Tan, Elena Volkanovska, and Wolfgang Stille. 2023. The insightsnet climate change corpus (iccc). In *BTW 2023*, pages 887–900. Gesellschaft für Informatik e.V., Bonn.

Cristina Bosco, Muhammad Okky Ibrohim, Valerio Basile, and Indra Budi. 2023. How green is sentiment analysis? environmental topics in corpora at the University of Turin. In *Proceedings of the 9th Italian Conference on Computational Linguistics (CLiC-it 2023)*, Venice, Italy. CEUR.

Cristina Bosco, Adriana Silvina Pagano, and Elisa Chierchiello. 2025. Communicating urgency to prevent environmental damage: insights from a linguistic analysis of the WWF24 multilingual corpus. In *Proceedings of the 1st Workshop on Ecology, Environment, and Natural Language Processing (NLP4Ecology2025)*, pages 34–43, Tallinn, Estonia. University of Tartu Library.

Elisa Chierchiello, Patricia Chiril, and Adriana Pagano. 2025. Ontology-guided domain entity recognition in environmental texts: Evaluating syntax-driven and LLM approaches using BabelNet and GEMET. In *Proceedings of the Eleventh Italian Conference on Computational Linguistics (CLiC-it 2025)*, pages 233–244, Cagliari, Italy. CEUR Workshop Proceedings.

Marie-Catherine De Marneffe, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. 2021. Universal Dependencies. *Computational Linguistics*, pages 1–54.

Xu Du, Matthew Kowalski, Aparna S. Varde, Gerard de Melo, and Robert W. Taylor. 2020. Public opinion matters: mining social media text for environmental management. *SIGWEB Newsl.*, 2019(Autumn).

Charles J. Fillmore. 2008. Border conflicts: FrameNet meets construction grammar. In *Proceedings of the XIII EURALEX International Congress*, pages 49–68, Barcelona. Universitat Pompeu Fabra.

Charles J. Fillmore, Paul Kay, and Mary Catherine O'Connor. 1988. Regularity and idiomaticity in grammatical constructions: The case of let alone. *Language*, 64(3):501.

Charles J. Fillmore, Russell R. Lee-Goldman, and Russell Rhomieux. 2012. The FrameNet constructicon. In Ivan A. Sag and Hans C. Boas, editors, *Sign-Based Construction Grammar*, pages 283–322. CSLI, Stanford.

Kerstin Fischer and Morgan Aarestrup. 2021. Relationships between construction grammar(s) and genre: Evidence from an analysis of instagram posts. *Journal of Pragmatics*, 183:87–104.

Adele E. Goldberg. 2013. *Constructionist Approaches*. Oxford University Press.

Francesca Grasso, Stefano Locci, Giovanni Siragusa, and Luigi Di Caro. 2024a. Ecoverse: An annotated twitter dataset for eco-relevance classification, environmental impact analysis, and stance detection. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation, LREC/COLING 2024*, pages 5461–5472. ELRA and ICCL.

Francesca Grasso, Ronny Patz, and Manfred Stede. 2024b. NYTAC-CC: A climate change subcorpus based on new york times articles. In *Proceedings of the Tenth Italian Conference on Computational Linguistics (CLiC-it 2024)*, Pisa (Italy). CEUR.

Normunds Gruzitis, Dana Dannélls, Benjamin Lyngfelt, and Aarne Ranta. 2015. Formalising the Swedish constructicon in grammatical framework. In *Proceedings of the Grammar Engineering Across Frameworks (GEAF) 2015 Workshop*, pages 49–56, Beijing, China. Association for Computational Linguistics.

Bruno Guillaume. 2019. Graph Matching for Corpora Exploration. In *JLC 2019 - 10èmes Journées Internationales de la Linguistique de corpus*, Grenoble, France.

Thomas Herbst and Thomas Hoffmann. 2018. Construction grammar for students: A constructionist approach to syntactic analysis (casa). *Yearbook of the German Cognitive Linguistics Association*, 6(1):197–218.

Muhammad Okky Ibrohim, Cristina Bosco, and Valerio Basile. 2023. Sentiment analysis for the natural environment: A systematic review. *ACM Computing Surveys*, 56(4).

Laura A. Janda, Olga Lyashevskaya, Tore Nesset, Ekaterina Rakhilina, and Francis M. Tyers. 2018. Chapter 6. A constructicon for russian: filling in the gaps. In *Constructicography*, pages 165–182. John Benjamins.

Zhuo Jing-Schmidt. 2017. What are they good for? a constructionist account of counterfactuals in ordinary chinese. *Journal of Pragmatics*, 113:30–52.

Knud Lambrecht. 1994. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge University Press.

Adrieli Laviola, Ludmila Lage, Natália Duarte Marção, Tatiane Tavares, Vânia Almeida, Ely Edison Matos, and Tiago Timponi Torrent. 2017. The Brazilian Portuguese Constructicon: Modeling Constructional Inheritance, Frame Evocation and Constraints in FrameNet Brasil. In *Proceedings of the AAAI Spring Symposia*, number SS-17-02 in AAAI Technical Report. AAAI Press.

Russell Lee-Goldman and Miriam R.L. Petruck. 2018. Chapteramp;160;2. the framenet constructicon in action. In *Constructicography*, pages 19–40. John Benjamins.

Per Linell. 2009. *Grammatical constructions in dialogue*, page 97–110. John Benjamins Publishing Company.

Arthur Lorenzi, Vânia Gomes de Almeida, Ely Edison Matos, and Tiago Timponi Torrent. 2023. Modeling construction grammar's way into NLP: Insights from negative results in automatically identifying schematic clausal constructions in Brazilian Portuguese. In *Proceedings of the First International Workshop on Construction Grammars and NLP (CxGs+NLP, GURT/SyntaxFest 2023)*, pages 96–109, Washington, D.C. Association for Computational Linguistics.

Benjamin Lyngfelt, Lars Borin, Kyoko Ohara, and Tiago Timponi Torrent, editors. 2018. *Constructicography: Constructicon development across languages*, volume 22 of *Constructional Approaches to Language*. John Benjamins Publishing Company, Amsterdam.

Kiki Nikiforidou and Kerstin Fischer. 2015. On the interaction of constructions with register and genre. *Constructions and Frames*, 7(2):137–147.

Kyoko Hirose Ohara. 2013. Toward Constructicon Building for Japanese in Japanese FrameNet. *Revista Veredas*, 17(1):11.

Jan-Ola Östman. 2005. *Persuasion as implicit anchoring: The case of collocations*, page 183–212. John Benjamins Publishing Company.

Adriana Silvina Pagano, Patricia Chiril, Elisa Chierchiello, and Cristina Bosco. 2025. TreEn: A multilingual treebank project on environmental discourse. In *Proceedings of the Eighth Workshop on Universal Dependencies (UDW, SyntaxFest 2025)*, pages 80–96, Ljubljana, Slovenia. Association for Computational Linguistics.

Ludovica Pannitto, Beatrice Bernasconi, Lucia Busso, Flavio Pisciotta, Giulia Rambelli, and Francesca Masini. 2024. Annotating constructions with ud: the experience of the italian constructicon. *arXiv preprint arXiv:2411.07623*.

Florent Perek and Amanda L. Patten. 2019. Towards an english constructicon using patterns and frames. *International Journal of Corpus Linguistics*, 24(3):354–384.

Flavio Pisciotta, Ludovica Pannitto, Lucia Busso, Beatrice Bernasconi, and Francesca Masini. 2025. Constraining constructions with wordnet: pros and cons for the semantic annotation of fillers in the italian constructicon. In *Proceedings of the 13th Global Wordnet Conference*, pages 77–84.

Manfred Stede and Ronny Patz. 2021. The climate change debate and natural language processing. In *Proceedings of the 1st Workshop on NLP for Positive Impact@ACL-IJCNLP 2021*, pages 8–18, Online. ACL.

Tiago Timponi Torrent, Ely Edison da Silva Matos, Ludmila Lage, Adrieli Laviola, Tatiane Tavares, Vânia Gomes de Almeida, and Natália Sigiliano. 2018. *Towards continuity between the lexicon and the constructicon in FrameNet Brasil*, page 107–140. John Benjamins Publishing Company.

Nicolas Webersinke, Mathias Kraus, Julia Anna Bingler, and Markus Leippold. 2022. ClimateBert: A pretrained language model for climate-related text. *ArXiv*, abs/2110.12010.

Leonie Weissweiler, Nina Böbel, Kirian Guiller, Santiago Herrera, Wesley Scivetti, Arthur Lorenzi, Nurit Melnik, Archna Bhatia, Hinrich Schütze, Lori Levin, et al. 2024. Ucxn: Typologically informed annotation of constructions atop universal dependencies. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 16919–16932.

Alexander Ziem and Hans C Boas. 2017. Towards a constructicon for german. In *AAAI Spring Symposia*.

A. Construction List and Validation Results

Table 4: Top constructions extracted from the 2014 WWF Living Planet Report. Frequency is normalized per 10,000 tokens.

Construction	Matches	Freq/10k	Sent. Coverage (%)
premodifier_of_noun	666	1239.53	84.82
adjunct_base	529	984.55	77.82
adjective_base	450	837.52	77.82
indefinite_plural	393	731.43	67.32
valency_postmodifier	257	478.32	59.14
postmodifier_pp_base	250	465.29	58.37
mass_np	241	448.54	50.19
definite_singular	182	338.73	50.97
declarative_statement	168	312.67	49.42
monotransitive	124	230.78	40.08
third_person_singular_present	121	225.20	38.52
intransitive	109	202.87	35.80
indefinite_singular	92	171.23	27.24
subject_attribute_pp	90	167.50	25.29
asyndetic_coordination	89	165.64	20.62
general_present_tense	85	158.20	25.29
definite_plural_np	55	102.36	17.51
we_definite_reference	55	102.36	16.34
postmodifier_recl	54	100.50	18.68
relative_postmodifier	54	100.50	18.68

Table 5: Top constructions extracted from the 2016 WWF Living Planet Report. Frequency is normalized per 10,000 tokens.

Construction	Matches	Freq/10k	Sent. Coverage (%)
premodifier_of_noun	1061	1386.02	93.68
adjunct_base	734	958.85	77.20
adjective_base	668	872.63	82.14
indefinite_plural	644	841.28	78.57
mass_np	453	591.77	61.54
valency_postmodifier	407	531.68	65.11
postmodifier_pp_base	405	529.07	65.11
definite_singular	271	354.02	50.00
declarative_statement	206	269.11	44.23
third_person_singular_present	166	216.85	37.91
intransitive	157	205.09	35.44
monotransitive	151	197.26	34.07
asyndetic_coordination	126	164.60	20.60
subject_attribute_pp	126	164.60	26.10
indefinite_singular	121	158.07	27.47
general_present_tense	99	129.33	21.98
definite_plural_np	93	121.49	20.88
monotransitive_clausal_objects	60	78.38	15.11
postmodifier_recl	59	77.07	13.46
relative_postmodifier	59	77.07	13.46

Table 6: Top constructions extracted from the 2018 WWF Living Planet Report. Frequency is normalized per 10,000 tokens.

Construction	Matches	Freq/10k	Sent. Coverage (%)
premodifier_of_noun	690	1175.47	86.45
adjunct_base	505	860.31	75.30
indefinite_plural	473	805.79	74.10
adjective_base	459	781.94	77.69
valency_postmodifier	363	618.40	72.11
postmodifier_pp_base	361	614.99	71.71
mass_np	227	386.71	49.00
definite_singular	208	354.34	53.78
third_person_singular_present	161	274.28	51.39
declarative_statement	135	229.98	41.04
indefinite_singular	129	219.76	41.43
intransitive	120	204.43	39.44
monotransitive	101	172.06	34.66
asyndetic_coordination	98	166.95	20.32
subject_attribute_pp	71	120.95	23.90
general_present_tense	70	119.25	23.51
definite_plural_np	54	91.99	19.92
passive_construction	54	91.99	18.73
first_name_np	46	78.36	15.14
first_third_past_tense	46	78.36	15.94

Table 7: Top constructions extracted from the 2020 WWF Living Planet Report. Frequency is normalized per 10,000 tokens.

Construction	Matches	Freq/10k	Sent. Coverage (%)
premodifier_of_noun	1029	1323.13	88.47
indefinite_plural	705	906.52	71.58
adjunct_base	644	828.08	66.49
adjective_base	609	783.08	76.68
valency_postmodifier	477	613.35	66.22
postmodifier_pp_base	475	610.78	66.22
mass_np	429	551.63	65.68
definite_singular	213	273.88	39.95
third_person_singular_present	198	254.60	38.61
declarative_statement	190	244.31	39.68
asyndetic_coordination	158	203.16	20.91
intransitive	145	186.45	31.37
monotransitive	141	181.30	31.90
indefinite_singular	121	155.59	25.20
subject_attribute_pp	109	140.16	22.25
first_third_past_tense	89	114.44	19.03
general_present_tense	82	105.44	19.30
passive_construction	76	97.72	15.82
focus_np	74	95.15	19.84
first_name_np	70	90.01	13.94

Table 8: Top constructions extracted from the 2022 WWF Living Planet Report. Frequency is normalized per 10,000 tokens.

Construction	**Matches**	**Freq/10k**	**Sent. Coverage (%)**
premodifier_of_noun	1478	1245.79	90.26
adjunct_base	1100	927.17	81.82
indefinite_plural	1037	874.07	79.44
adjective_base	881	742.58	82.03
valency_postmodifier	716	603.51	72.29
postmodifier_pp_base	714	601.82	71.86
mass_np	547	461.06	59.74
definite_singular	358	301.75	52.81
declarative_statement	300	252.87	51.08
third_person_singular_present	279	235.17	47.62
intransitive	232	195.55	41.56
monotransitive	218	183.75	40.91
indefinite_singular	208	175.32	35.93
asyndetic_coordination	204	171.95	20.56
subject_attribute_pp	165	139.08	27.49
general_present_tense	164	138.23	30.09
first_name_np	132	111.26	19.48
last_name	132	111.26	19.48
definite_plural_np	117	98.62	22.51
passive_construction	109	91.87	20.78

Table 9: Top constructions extracted from the 2024 WWF Living Planet Report. Frequency is normalized per 10,000 tokens.

Construction	**Matches**	**Freq/10k**	**Sent. Coverage (%)**
premodifier_of_noun	3446	1335.35	91.65
indefinite_plural	2405	931.95	83.10
adjunct_base	2348	909.87	81.01
adjective_base	2009	778.50	80.52
mass_np	1633	632.80	69.28
valency_postmodifier	1257	487.10	66.10
postmodifier_pp_base	1250	484.38	65.90
definite_singular	711	275.52	47.02
declarative_statement	617	239.09	46.82
third_person_singular_present	543	210.42	40.26
monotransitive	508	196.85	42.15
asyndetic_coordination	500	193.75	28.33
intransitive	457	177.09	36.68
indefinite_singular	363	140.66	29.82
general_present_tense	337	130.59	27.93
subject_attribute_pp	313	121.29	23.06
definite_plural_np	246	95.33	20.87
monotransitive_clausal_objects	202	78.28	17.79
postmodifier_recl	197	76.34	17.20
relative_postmodifier	197	76.34	17.20

Table 10: Full validation results for all constructions, reported using the same precision metric as in Table 1.

Construction	C	N	Precision
first_name_np	0	30	0.00
last_name	0	30	0.00
on_focus_area	11	30	0.37
for_desired_object	12	30	0.40
subject_attribute_pp	14	30	0.47
monotransitive_clausal_objects	20	30	0.67
additive_coordination_base	21	30	0.70
postmodifier_acl_partger	21	30	0.70
with_partner	21	30	0.70
general_postmodifier_of_adjective	22	30	0.73
indefinite_plural	23	30	0.77
relative_postmodifier	25	30	0.83
intransitive	26	30	0.87
premodifier_of_adjective	26	30	0.87
focus_np	27	30	0.90
postmodifier_pp_base	27	30	0.90
adjunct_base	28	30	0.93
mass_np	28	30	0.93
valency_postmodifier	28	30	0.93
adjective_base	29	30	0.97
alternative_coordination	29	30	0.97
asyndetic_coordination	29	30	0.97
declarative_statement	29	30	0.97
monotransitive	29	30	0.97
numerical_plural	29	30	0.97
passive_construction	29	30	0.97
premodifier_of_noun	29	30	0.97

Construction	C	N	Precision
definite_plural_np	30	30	1.00
definite_singular	30	30	1.00
demonstrative_plural	30	30	1.00
demonstrative_singular	30	30	1.00
general_present_tense	30	30	1.00
general_quantifier_singular	30	30	1.00
genitive_np	30	30	1.00
indefinite_singular	30	30	1.00
it_definite_reference	30	30	1.00
most_least	30	30	1.00
perfective_construction	30	30	1.00
postmodifier_recl	30	30	1.00
progressive	30	30	1.00
superlative	30	30	1.00
third_person_singular_present	30	30	1.00
us_definite_reference	30	30	1.00
we_definite_reference	30	30	1.00
first_third_past_tense	29	29	1.00
they_definite_reference	29	29	1.00
ranking_np	26	28	0.93
general_quantifier_plural	28	28	1.00
to_recipient_goal	14	26	0.54
them_definite_reference	26	26	1.00
self_motion	17	25	0.68
general_past_plurals	24	24	1.00
general_quantifier_mass	24	24	1.00
complex_full_name	20	23	0.87
ditransitive_objtoinf	20	21	0.95
imperative_construction	21	21	1.00
self_change_resultative_base	14	19	0.74
full_name	2	18	0.11
attribute_construction_adj	10	18	0.56
x_is_adj_to	16	17	0.94
subject_attribute_adjp	2	15	0.13
caused_motion	12	14	0.86
adjectival_with_partner_base	4	10	0.40
ditransitive_obj_inf	7	10	0.70
realistic_condition	7	10	0.70
comparative_than	10	10	1.00
ditransitive_objnp	8	8	1.00
we_are_able_to_make	8	8	1.00
x_is_likely_to	8	8	1.00
itself_reflexive	1	7	0.14
change_of_state	0	6	0.00
caused_change_resultative	6	6	1.00
hypothetical	6	6	1.00
to_recipient_for_desired	1	4	0.25
about_reference_area	4	4	1.00
you_definite_reference	4	4	1.00
wh-cleft	0	2	0.00
ourselves_reflexive	2	2	1.00
ditransitive_wh-cl	1	1	1.00
exchange_x_for_y	1	1	1.00
first_person_singular_present	1	1	1.00
i-definite_np	1	1	1.00
it_cleft	1	1	1.00

Construction	C	N	Precision
resultative_attradj	1	1	1.00
wh-question_subject	1	1	1.00
with_emotion	1	1	1.00
yes_no_question	1	1	1.00