# MindSpaces

Art-driven adaptive outdoors and indoors design

H2020- 825079

# D5.5

# Text generation tools v1

| | |
|---|---|
| **Dissemination level:** | Public |
| **Contractual date of delivery:** | Month 19, 31 July 2020 |
| **Actual date of delivery:** | Month 19, 31 July 2020 |
| **Work package:** | WP5 MindSpaces adaptive environment development |
| **Task(s):** | T5.5: Text generation |
| **Type:** | Report |
| **Approval Status:** | Approved |
| **Version:** | 1.0 |
| **Number of pages:** | 62 |
| **Filename:** | D5.5_Text_generation_tools_v1_v1.0.docx |

**Abstract**

This deliverable elaborates on the basic text generation tools that are applied (i) as an explanatory system so that the users could have descriptive textual information during VR experience and (ii) as a service that addresses the multilingual language generation within artistic solutions in the scope of the project.

co-funded by the European Union

# History

| Version | Date | Reason | Revised by |
|---|---|---|---|
| V0.1 | 23.06.2020 | Creation of ToC | Alexander Shvets |
| V0.2 | 07.07.2020 | SoA, Basic techniques, NLG dataset for advanced techniques, Evaluations | Simon Mille |
| V0.3 | 23.07.2020 | Completed NLG parts and datasets, added references | Simon Mille |
| V0.4 | 24.07.2020 | Automated lexicon compilation | Beatriz Fisas |
| V0.5 | 26.07.2020 | Final version for internal review | Alexander Shvets |
| V0.6 | 29.07.2020 | Review | Luis Fraguada |
| V1.0 | 30.07.2020 | Final edit | Alexander Shvets |

# Author list


| Organization | Name | Contact Information |
|---|---|---|
| UPF | Alexander Shvets | alexander.shvets@upf.edu |
| UPF | Simon Mille | simon.mille@upf.edu |
| UPF | Beatriz Fisas | beatriz.fisas@upf.edu |

# Executive Summary

The deliverable 5.5 describes the progress on the task T5.5 “Text generation” of WP5. This task is responsible for the production of multilingual texts describing the changes in the environment during end-users’ VR experience and addresses the language generation within artistic solutions.

The text generation component is connected to the semantic reasoning framework (the result of T5.3 “Semantic representation and data integration”) that provides ontology-based structures generated in T5.4 “Semantic reasoning for emotion-based space adaptation” to be realised as natural language sentences. These structures represent the state of the user and the VR environment and the triggered rules for the change in the space to be applied.

The advances in three main subtasks are discussed in the course of the deliverable, namely: (i) the compilation and extension of the linguistic resources (generation lexica, graph transduction grammars) for the languages considered in MindSpaces (English, Spanish, French, Catalan, Greek); (ii) the improvement of multilingual discourse generators available in UPF with their further extension in accordance with the MindSpaces requirements; (iii) the development and realisation of a principled methodology for the mapping of ontological representations to conceptual structures appropriate for language generation.

The text generation component is being developed as expected, the provided functionality is in line with the project timeline. The evaluation shows promising results. The text generation pipeline was integrated into the project platform and was successfully applied within a preliminary implementation of an artistic solution.

# Abbreviations and Acronyms

| | |
|---|---|
| **ConS** | Conceptual Structures |
| **DIST** | Normalised Inverted Edit Distance |
| **DnS** | Description and Situation |
| **DSyntS** | Deep-Syntactic Structure |
| **dul** | DOLCE+DnS Ultralite |
| **EEG** | Elecroencephalography |
| **FORGe** | Fabra Open-source Rule-based Generator |
| **GP** | Government Pattern |
| **KB** | Knowledge Base |
| **LF** | Lexical Function |
| **LSS** | Lexical Semantic Structures |
| **LSTM** | Long Short-Term Memory |
| **MorphS** | Morphological Structure |
| **NLG** | Natural Language Generation |
| **NLP** | Natural Language Processing |
| **PoS** | Part of Speech |
| **PUC** | Pilot Use Case |
| **RDF** | Resource Description Format |
| **SemS** | Semantic Structure |
| **SR** | Surface Realisation Shared Task |
| **SSyntS** | Surface-Syntactic Structure |
| **SW** | Semantic Web |
| **UD** | Universal Dependency |
| **VR** | Virtual Reality |

# Table of Contents

# 1 INTRODUCTION

Text generation T5.5 being connected to other tasks in WP5 through ontology-based modelling T5.3 (Figure 1) plays an important role in attaining the project goals. The versatility of the project requires the production of multilingual texts both within the main workflow organised as a cooperation of different technologies comprising the MindSpaces platform and within various downstream applications that mainly aim to create functional artworks. This causes the necessity in supporting two types of approaches: a solid multistage text generation that allows for advanced control of outputs for long-term tasks and a statistical neural network-based realization. The latter might be readily tuned for the needs of particular spontaneous subproject at the cost of less control of the outcome content as it is highly dependent on the available data in the domain to be used for training the models.

The multistage approach is central in the project and is applied to describe the state of running experiments in MindSpaces, where subjects observe adaptive architectural and artistic solutions inside virtual reality (VR) and provide explicit and implicit feedback that helps designers and creatives to identify the most prominent design features for particular indoor and outdoor environments. In spite of the fact that these experiments are developed to not be very intrusive for the subjects, the number of operations performed by the Platfrom behind the scene is very high. They include the selection of predefined configurations to start with, the analysis and interpretation of behavioral patterns and EEG signals, detection of relevant insights in the results of textual analysis, and providing decisions made by the Reasoner to VR to apply a change in a space, etc. All this makes the experiment quite dynamic with some parts hidden from the experimenter's point of view. The automatically generated explanatory text highlighting the key events during end-users' VR experience is intended to ease tracking the experiment and drawing conclusions.

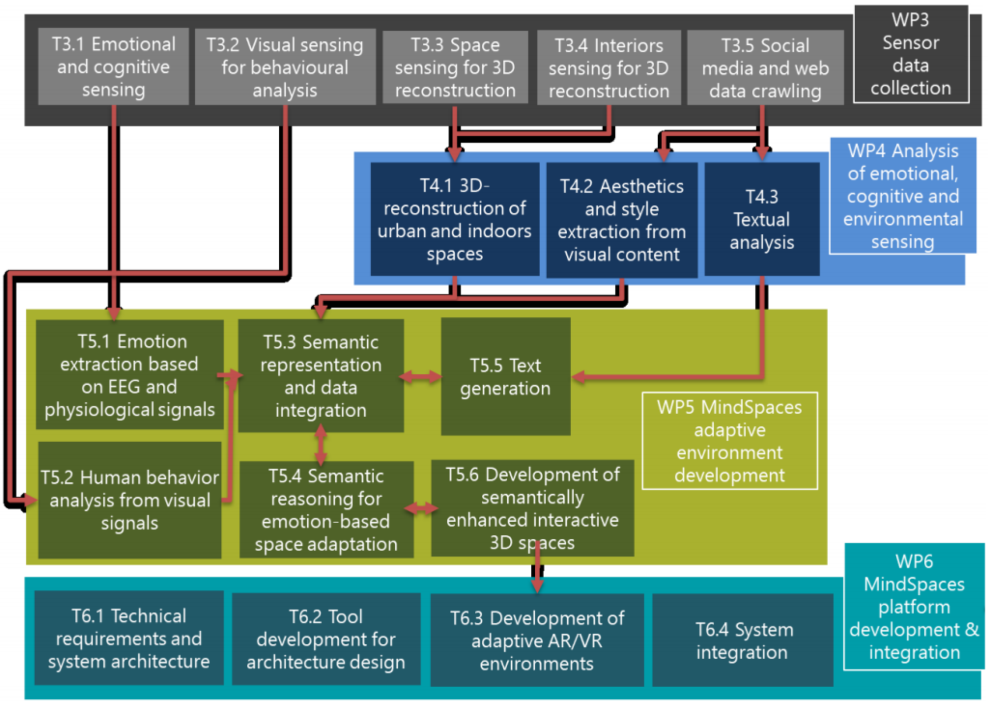



Figure 1: Position of text generation among other project tasks

A statistical approach is to be applied first for the personalised narrative generation, starting from the crawled data where users interact with the process at various stages (i.e., provide textual collections of their interest to be crawled, decide whether to add generated sentences to their story or to regenerate them, select some of the concepts automatically extracted from current narrative to further guide the crawling, etc.). Since external artists in the consortium have not finalized their proposals, additional usage is still expected. One of the envisioned applications is the creation of music with artificially voiced generated text.

The position of text generation component in the Platform corresponds to the position of the task T5.5 in Figure 1: it communicates with the Knowledge Base and takes as input either ontological representations or resulting structures (parse trees) of the Textual Analysis component and produces plain text as an outcome. Two technical requirements are addressed by the development of the component: (i) “Create a technique of the projection of ontology constructs to respective lexicalized semantic structures to support knowledge-driven content selection”, and (ii) “Create a linguistic generation service for the realization of acquired knowledge as natural language sentences”.

The remainder of the deliverable is structured as follows.

Section 2 provides an overview of state-of-the-art techniques in multilingual text generation and discusses advantages and drawbacks of certain approaches.

Section 3 proposes the method of projection of ontology constructs to lexicalized semantic structures, introduces graph-transduction framework, and presents a basic approach based on graph-transduction grammars with further discussion of the progress done within MindSpaces. Quantitative and qualitative evaluation with a standard challenging dataset is provided, and results are outlined.

Section 4 describes the work done towards advanced multilingual text generation, namely automated lexicon compilation, creation of the datasets for Surface Realisation Shared Task, and the first version of a statistical neural lineariser. We discuss the results of the shared task and introduce the best-participated model that is chosen as a basis for neural linearization in MindSpaces. The preliminary experiments within the implementation for artistic purposes are briefly presented.

Section 5 concludes the deliverable and outlines possible directions regarding future tasks.

# 2 STATE OF THE ART

In this section, we present the state of affairs in the area of lexical resources and state-of-the-art methods in multilingual text generation with an emphasis on approaches that allow handling language-independent ontological representations.

We started with the languages foreseen in the project and compiled available lexical resources for English, Greek, German, and Spanish. The existing work for remaining languages which have been recently added (French and Catalan), will be investigated during the course of the project.

## 2.1 Existing lexical resources

Good quality lexical resources are needed in order to obtain reliable semantic structures as input for text generation. We aim at compiling descriptions of lexical units that include their government patterns (GP) (or subcategorisation frames), that is, how many participants one unit usually has and how they combine with each other. There are many lexical resources which are useful for a variety of applications. We focus on the resources that can be used for both language analysis and language generation. Lexicons with more generic semantic information can be very useful, and those that include mappings to standard resources (BabelNet, PropBank, or VerbNet for instance) are preferred. The following tables compile the lexical resources relevant to our purposes.racteristics of Greek lexicons.

Table 1 describes the main characteristics of Greek lexicons.

Table 1: Greek lexicons

| GREEK LEXICONS | | | | |
|---|---|---|---|---|
| Name | Short description | format | size | license |
| EXIS[1]<br><br>GDT-LEXIS (Papageorgiou et al, 2006)<br>LEXIS-EmotionVerbs (Giouli and | A Greek Computational Lexicon of general language based on corpora, language with morphological, syntactic and semantic information.<br>GDT-LEXIS: a lexical resource with semantic information for verbal predicates.<br>LEXIS-Emotion Verbs: details the argument structure, distributional properties and possible | | Comprises ~60,000 entries with morphological information, of which a subset of 30,000 entries also have syntactic information and a further subset of 15,000 with semantic information.<br>In GDT-LEXIS: about | |

[1] http://www.ilsp.gr/en/infoprojects/meta?view=project&task=show&id=140

| | | | | |
|---|---|---|---|---|
| Fotopoulou, 2012) | transformations of greek emotion verbs. | | 800 verbs | |
| SKEL (Petasis et al, 2001) | Morphological lexicon that was used to develop a lemmatiser and a morphological analyser that were included in a controlled language checker for Greek. | | ~60.000 lemmas that correspond to ~710.000 different word forms. | |
| Conceptual Lexicon (Fotopoulou et al, 2014) | Encodes morphosyntactic and semantic properties of nominal and verbal multi-word expressions (MWEs). | | ~1000 entries | |
| EKFRASI (Tzortzi and Markantonatou, 2014) | Conceptually organised lexicon encoded with Protégé, Includes conceptual and lexical relations as well as their morphosyntactic properties. | | | |

Table 2 summarises the main characteristics of Spanish lexicons.

Table 2: Spanish lexicons

| SPANISH LEXICONS | | | | |
|---|---|---|---|---|
| Name | Short description | format | size | license |
| ANCORA_VERB_ES (Aparicio et al, 2008) | Semantic info, subcategorisation, Argumental patterns and thematic roles. Pbank id, Verbnet Id, Framenet id, Wordnet id. | XML | 2,820 verbs | Freely available |
| ANCORA_NOM_ES (Peris and Taulé, 2011) | Deverbal nouns: Denotative type, Wordnet synset, argumental pattern and thematic roles. Link to verb. | XML | 1,658 lemmas | Freely available |
| ANCORANET (Taulé et al, 2011) | Contains the AnCora-Verb lexical entries linked to different English knowledge sources: VerbNet, PropBank, FrameNet, WordNet 3.0 and OntoNotes. | XML | | Freely available |
| ADESSE (García-Miguel et al, 2010) | An online database for the empirical study of the interaction between verbs and constructions in Spanish: | | ~4,000 verbs | |

| | | | | |
|---|---|---|---|---|
| | Subcategorisation frames, diathesis alternations and syntactic semantic schemes. | | | |
| GLiCOm[2] | Computational lexicon of inflected wordforms in Spanish. The lexicon is distributed in two sublexicons: 1. word forms 2. verb-clitic combinations. | | 1,152,242 word forms, and 4,283,637 verb-clitic combinations | Freely available |

Table 3 describes the main characteristics of English lexicons, while Table 4 contains the German ones and Table 5 gives details about the multilingual resources.

Table 3: English lexicons

| ENGLISH LEXICONS | | | | |
|---|---|---|---|---|
| Name | Short description | format | size | license |
| PropBank / NomBank (Kingsbury and Palmer, 2002) / (Meyers et al, 2004) | Subcategorisation frames for verbs and nouns, correspondences between syntactic and semantic roles. | XML | 11,781 disambiguated lemmas | CC BY-SA 4.0 |
| VerbNet (Schuler 2005) | Classification of verbs into 270 semantic classes; Subcategorisation frames, diathesis alternations and syntactic semantic schemes. | XML | 2,380 disambiguated verbs | CC BY-SA 4.0 |
| Framenet[3] (Baker et al, 1998) | English resource based on frame semantics, which models “prototypical situations” with participants and their roles. | XML, HTML | 1,224 frames<br>13,640 lexical units<br>10,542 frame elements<br>1,876 frame-to-frame relations<br>20,229 annotated sentences | |
| WordNet[4] | A large lexical database of English. | 8 files in | 117 000 synsets | Freely |

2 https://www.upf.edu/documents/107805982/109136461/tec0128_glicom_tbadia.pdf/07632628-f275-425e-b59c-417433c6a327
3 https://framenet.icsi.berkeley.edu/fndrupal/
4 https://wordnet.princeton.edu/

| | | | | |
|---|---|---|---|---|
| (Fellbaum, 2005) | Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations. | ASCII format | | and publicly available for download |
| ConceptNet (Speer et al, 2017) | ConceptNet is a multilingual knowledge base, representing words and phrases that people use and the common-sense relationships between them. The knowledge in ConceptNet is collected from a variety of resources, including crowd-sourced resources (such as Wiktionary and Open Mind Common Sense), games with a purpose (such as Verbosity and nadya.jp), and expert-created resources (such as WordNet and JMDict). | JSON-LD a linked data format | 34 million edges (statements) | Freely available with Creative Commons Attributio n-Share-Alike license |

Table 4: German lexicons

| GERMAN LEXICONS | | | | |
|---|---|---|---|---|
| Name | Short description | format | size | license |
| IMSLex German Lexicon[5] | A lexical resource comprising morphological and syntactic information that links together previous resources from IMS-Stuttgart, and covers information on inflection, word formation and valence. Follows the LFG theoretical framework. | XML | 11,000 adjectives<br>1,000 adverbs<br>22,500 nouns<br>300 particles<br>10,000 proper nouns<br>6,000 verbs<br>167 derivation suffixes | Academic research license |
| HaGenLex (HAgen GErmaN LEXicon)[6] (Hartrumpf et al, | A domain independent computational lexicon with morphosyntactic and semantic information | Internal representation: standard typed feature | 12986 noun entries<br>6911 verb entries<br>3278 adjective entries | Contact the author: Rainer |

[5] https://pdfs.semanticscholar.org/1fee/63d8a6114720653c9e2327188491ccf77a92.pdf
[6] http://pi7.fernuni-hagen.de/research/hagenlex/hagenlex-en.html#HHO03

| | | | | |
|---|---|---|---|---|
| 2003) | (based on the MultiNet paradigm, which provides a hierarchy of 45 ontological sorts and a more than 100 semantic relations and functions). | structure formalism. Some expanded entries also have XML representations | 579 adverb entries | Osswald (rainer.osswald@fernuni-hagen.de) |
| GermaNet v.14.0[7] (Univ.Tübingen) (Hamp and Feldweg, 1997) | A lexical-semantic net that relates German nouns, verbs, and adjectives semantically by grouping lexical units that express the same concept into synsets and by defining semantic relations between these synsets. Related to Wordnet especially in the case of nouns. | Relational database and XML files | Synsets: 136263<br>Lexical units: 175000<br>Literals: 159359<br>conceptual relations: 150003<br>lexical relations: 12203 (synonymy excluded)<br>Wiktionary sense descriptions: 29549 | GermaNet is free for academic users but you have to sign a license |
| BilderNetle[8] (Roller and Schulte, 2013) | A Dataset of German Noun-to-ImageNet Mappings<br>ImageNet is a large-scale and widely used image database, built on top of WordNet, which maps words into groups of images, called synsets. Multiple synsets exist for each meaning of a word.<br>This BilderNetle dataset provides mappings from German noun types to images of the nouns via ImageNet. | | 2,022 word-synset mappings for 309 words | Freely available for education, research and other non-commercial purposes |

[7] http://www.sfs.uni-tuebingen.de/GermaNet/
[8] https://www.ims.uni-stuttgart.de/forschung/ressourcen/lexika/bildernetle.en.html

| | | | | |
|---|---|---|---|---|
| German Subcat Database extracted from MATE Dependency Parses[9] (Scheible et al, 2013) | Induced verb subcategorisation information from German MATE dependency parses, based on the SubCat-Extractor tool. The subcategorisation database is represented in a compact but linguistically detailed and flexible format, comprising various aspects of verb information, complement information and sentence information, within a one-line-per-clause style. | | | The SubCat-Extractor is freely available for education, research and other non-commercial purposes |

Table 5: Multilingual lexicons

| MULTILINGUAL LEXICON | | | | |
|---|---|---|---|---|
| Name | Short description | format | size | license |
| BabelNet (Navigli and Ponzetto, 2010) | Dictionary with fine-grained senses, definitions and mappings to VerbNet among others. | RDF / HTTP API | 284 languages ~6,000,000 concepts 10,000,000 named entities | CC BY-NC-ND 4.0 |
| UBY[10] (Gurevych et al, 2012) | A large-scale lexical-semantic resource for natural language processing (NLP) based on the ISO standard LMF. UBY combines a wide range of information from expert-constructed and collaboratively constructed resources for English and German. Currently, UBY integrates resources in English and German by linking them pairwise at the word sense level: English WordNet, Wiktionary, Wikipedia, FrameNet and the syntactically rich VerbNet, German Wikipedia, G_ Wiktionary, GermaNet, | UBY database | | Apart from GermaNet and IMSlex which are licensed under an academic research license,all resources in UBY are available |

9 https://www.ims.uni-stuttgart.de/forschung/ressourcen/lexika/subcat-database.en.html
10 https://dkpro.github.io/dkpro-uby/

| | | | | |
|---|---|---|---|---|
| | IMSLex-Subcat and OmegaWiki. | | | under open licenses, requiring either attribution or both, attribution and share alike |
| lemonUBY[11] | A Semantic Web version of UBY. | lemonUBY (an export of UBY data into lemon derived from UBY.) | | |

For the project, we will use primarily lexical resources that have an explicit mapping to standard English resources such as VerbNet.

## 2.2 Multilingual text generation

For targeting the development of a reusable Natural Language Generation (NLG) pipeline and its interface with the Knowledge Base (KB), we base our approach on the traditional view of NLG as a sequence of three subtasks: (i) content selection, which is responsible for determining the contents to be rendered as text, (ii) text planning, which takes care of packaging the contents into discursively organised units (i.e., sentences), and (iii) linguistic generation, which realises the contents as well-formed text (Rambow y Korelsky 1992). In MindSpaces, step (i) is carried out by the Reasoning module and/or the content selection module described in this deliverable, and steps (ii) and (iii) by the text generation module.

In general, each step can be performed using template-based, grammar-based or statistical systems, or a combination of these (Ballesteros, et al. 2015) (Gardent, et al. 2017a). Currently, a lot of research in the topic addresses the whole sequence as one step, and focuses on filling the slot values of pre-existing templates using neural network techniques (Nayak, et al. 2017). Few systems follow a theoretical framework, and most of them make extensive use of language models (i.e. use a large amount of reference texts) to statistically mimic correct language use (Gardent, et al. 2017b). The main problems with these approaches are their low portability to new languages and domains and the lack of control over the final output, but also the very limited amount of actual linguistic knowledge used during the generation process. A multilayer grammar-based generator does not require

[11] https://www.lemon-model.net/lexica/uby/

training material, allows for a greater control over the outputs (e.g., for mitigating possible errors or tuning the output to a desired style), and the linguistic knowledge used for one domain or language can be reused for other domains and languages. However, due to their complexity, such approaches have undergone few developments within the open-source community in recent years (Gatt and Krahmer 2018). The only grammar-based system used successfully in all recent NLG shared tasks is FORGe, the open-source generator being developed in the framework of MindSpaces, which addresses the last two NLG subtasks mentioned above, namely text planning and linguistic generation. FORGe, building on the lines of the Meaning-Text Theory (Mel'čuk 1988), is based on the notion of linguistic dependencies, that is, the semantic, syntactic, and morphologic relations between the components of the sentence. It was the best system at the WebNLG 2017 shared task (automatic verbalisation in English of several hundreds of pre-selected properties) according to all human evaluations, and was the most portable generator with the best results for all metrics on unseen data.[12] FORGe is a very promising system but currently handles only a small subset of abstract contents. The text planning layer in FORGe is embryonic, and its linguistic generation layers suffer from coverage issues due to the fact that this generator has been developed in the framework of EU projects that always target specific domains (Wanner, et al. 2010), (Bouayad-Agha, et al. 2012), (Wanner, et al. 2015)[13]. On complex general-domain inputs, for about 25% of the contents, the generator does not find an adequate syntactic structure and cannot generate complete sentences. Furthermore, its multilingual coverage is limited (Mille, et al. 2017). Thus, one of the main objectives of MindSpaces with respect to text generation is to improve the multilingual coverage and the quality of the UPF generator.

As far as input representations are concerned, an NLG pipeline needs to be fed with linguistic structures. These are quite different from the triples found on the KB, in which the properties are labelled with an open vocabulary and only two types of relations (Subject and Object) are used. The triples must be mapped onto linguistic concepts and relations, preferably according to standard lexico-semantic resources such as VerbNet (Schuler 2005), NomBank (Meyers, et al. 2004), or PropBank (Kingsbury and Palmer 2002) to ensure reusability. These resources can be used as interlingua thanks to the amount of multilingual resources connected to them. To the best of our knowledge, little research has been carried out so far on bringing together KB contents and standard linguistic resources in the context of NLG: on the one hand, standard Semantic Web (SW) approaches such as Lemon (Walter, Unger and Cimiano 2014) or word-embeddings-based lexicalization (Perez-Beltrachini and Gardent 2016) define their own lexicons to be associated with the properties, and on the other hand, linguistic resources such as VerbNet, NomBank and PropBank are not connected with reusable Knowledge Bases. Finally, even if the SW components were mapped to NomBank and PropBank entries, the syntactic information about the participants is not

[12] http://webnlg.loria.fr/pages/webnlg-human-evaluation-results.pdf

[13] See, e.g., the FP7 and H2020 projects PESCaDO, and KRISTINA.

expressed in these resources. This subcategorisation information can be derived from VerbNet, which is neither NLG nor dependency-oriented.

# 3 BASIC TECHNIQUES FOR MULTILINGUAL TEXT GENERATION

The multilingual text generation pipeline consists of two main parts: a knowledge-oriented module, in charge of mapping ontological representations onto linguistic representations (Section 3.1 ), and a linguistics-oriented module, in charge of transforming the deep linguistic representations into well-formed text (Section 3.2 ). For the implementation of the linguistic module, we are developing a graph-transduction environment tool (described in Section 3.2 ).

## 3.1 Projection of ontology constructs to lexicalised semantic structures

The mapping of the ontological representation to lexicalised semantic structures, referred to as *conceptual structures*, is the first step towards the projection of the input class and property assertions to language-oriented structures. The ontological-to-conceptual grounding is based on the Description and Situation (DnS) pattern of DOLCE+DnS Ultralite[14] (dul). Under the adopted DnS-based paradigm, each dul:Situation object corresponds to an n-ary linguistic predicate, with its participating entities' roles specified through the associated dul:Concept objects that are in turn defined (via dul:satisfied assertions) by corresponding dul:Description objects. When mapping to the conceptual structure, participating elements (classified as arguments) are mapped to linguistic arguments (i.e., labelled edges that link the predicate to the argument), while circumstantials (e.g., temporal attributes of an action, such as start time and duration) are treated as typed predicative nodes through which the linking between the predicate and the circumstantial entity is realised. Thereby, as opposed to the content structure, the conceptual structure encapsulates the first version of what will be found in the final sentence: only the elements which will be mentioned (explicitly or not) are kept. Some other elements are removed altogether (e.g., the situation and description elements), while others are captured in another form. For instance, the information related to the ontological type of retrieved data will only be realised as grammatical tense on the main verb of the sentences (in the case of habits, present), and not mentioned as such. Table 6 summarises the main transformation rules used for mapping the DnS-based representations to respective conceptual structures (ConS).

Table 6: DnS-based to ConS-based representation transformation mappings

| ***DnS-based representation*** | ***Conceptual structure representation*** |
| --- | --- |

[14] http://www.ontologydesignpatterns.org/ont/dul/DUL.owl

| | |
|---|---|
| Eventive relational contexts, i.e. dul:Situations (onto:Sleep, onto:Eat, onto:Walk, etc.) and participating entities | N-ary predicates and their arguments |
| Argumentative dul:Concept specialisations (context:Agent, context:Beneficiary, context:Theme, context:Destination, etc.) | Argument labels of referenced n-ary predicates (Argument1, Argument2, etc.) |
| Circumstantial dul:Concept specialisations (context:FrequencyAttribute, context:TemporalAttribute, context:TemporalPattern, etc.) | Typed predicative nodes (Frequency, StateTime/EndTime, TemporalOverlap/ TemporalOrder, etc.) |

Grounding the conceptual representation on the DnS model and retaining a high level of abstraction allows for a principle interface between the ontology and the semantic structures of the next layers while rendering it sufficiently generic, so as to allow for generation in any language. At this point in the project, the conceptual structures are in the form of simple predicate-argument templates associated with each property in the ontology. As an example, consider Figure 2 that illustrates a triple set as reformulated from the KB, which describes a situation with design parameters of the space, a detected state of mind of a subject, and a subsequent change in one of the parameters, and its respective "conceptual" representation, illustrated in Figure 3 as a set of populated predicate-argument templates.

*ceiling_height (room, high)*

*light_origin (room, below)*

*shape (room, quadratic)*

*colour (wall, yellow)*

*confidence (feel (subject, anger), low)*

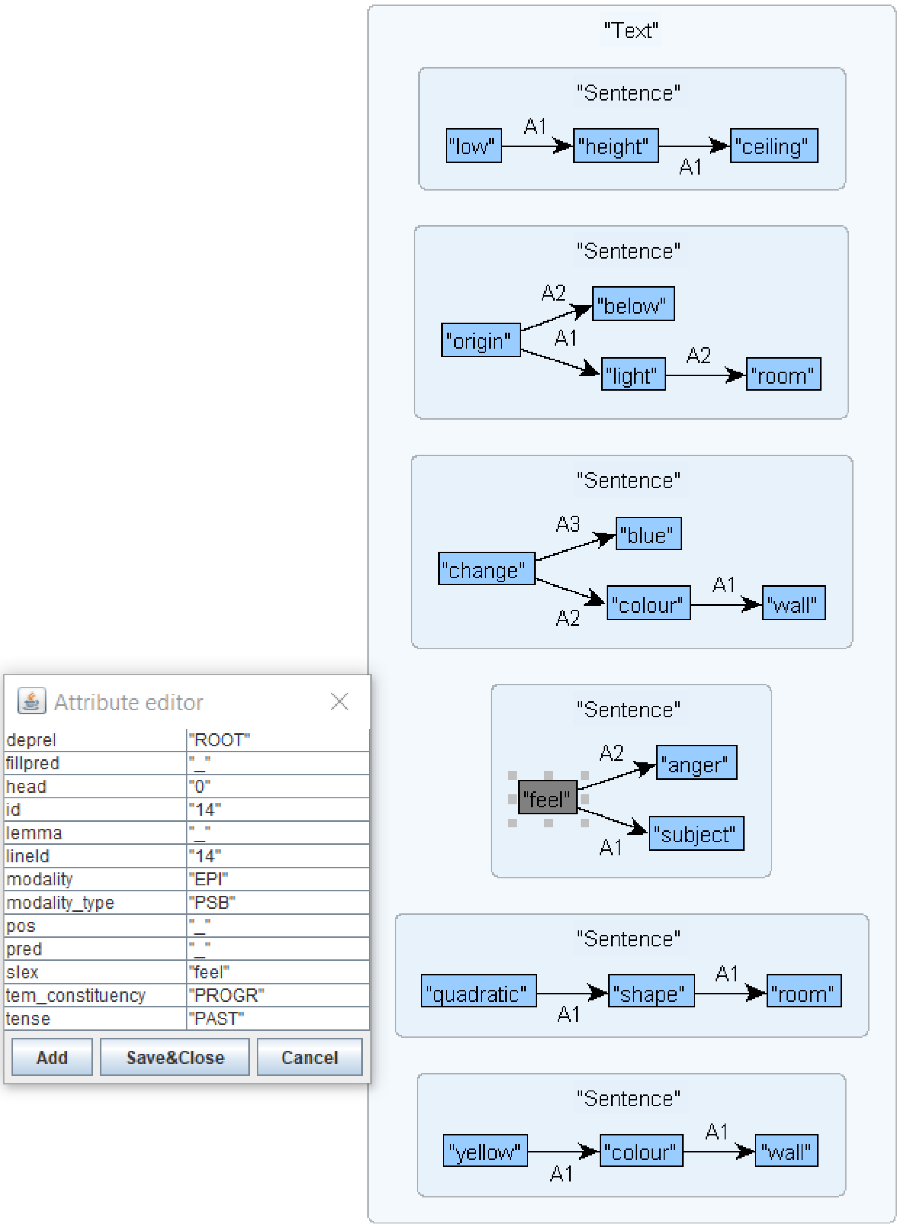
*change (wall_colour(room), blue)*

Figure 2: Six triples related to room parameters

Note that the conceptual structures are linguistic structures, since they contain only linguistic elements (meanings, or concepts). However, they are language-independent, in that the same input structure is used regardless of the target language. Multilinguality is dealt with during the next transition.

For the deeper stages of generation (i.e., for going from the ontology to the deep-syntactic structures used for surface generation, rule-based modules have been developed.

The DnS-based translation rules are manually crafted and realise the inverse transformation of the one applied during the analysis of the verbal content of user utterances. For the implementation of the ontology to conceptual structure translation, we foresee using the Jena API[15] for RDF for parsing the input DnS-based responses.




15 https://jena.apache.org

Figure 3: Six populated predicate argument structures corresponding to the triples in Figure 2

## 3.2 Graph-transduction framework

For the implementation of the graph-transduction rules that map the conceptual structures onto text, we have been re-implementing the graph transducer MATE (Bohnet and Wanner, 2010) in order for the transduction rules to be more expressive and compact, as well as for the tool to perform the transductions faster. We will refer to this new tool as "MATE-2" in this section.

MATE is a graph transducer programmed in Java. It contains different editors for graph construction, rule and lexical resource writing, a debugger, as well as a tool for regression tests. The rules (and their corresponding conditions) match a part of an input graph, and create a part of the output graph. The main problem with MATE is its speed, which was not adequate for a real-time system as potentially needed in MindSpaces. MATE-2 is currently in an advanced state but no publication describes it yet. Figure 4 shows a project open in MATE-2.

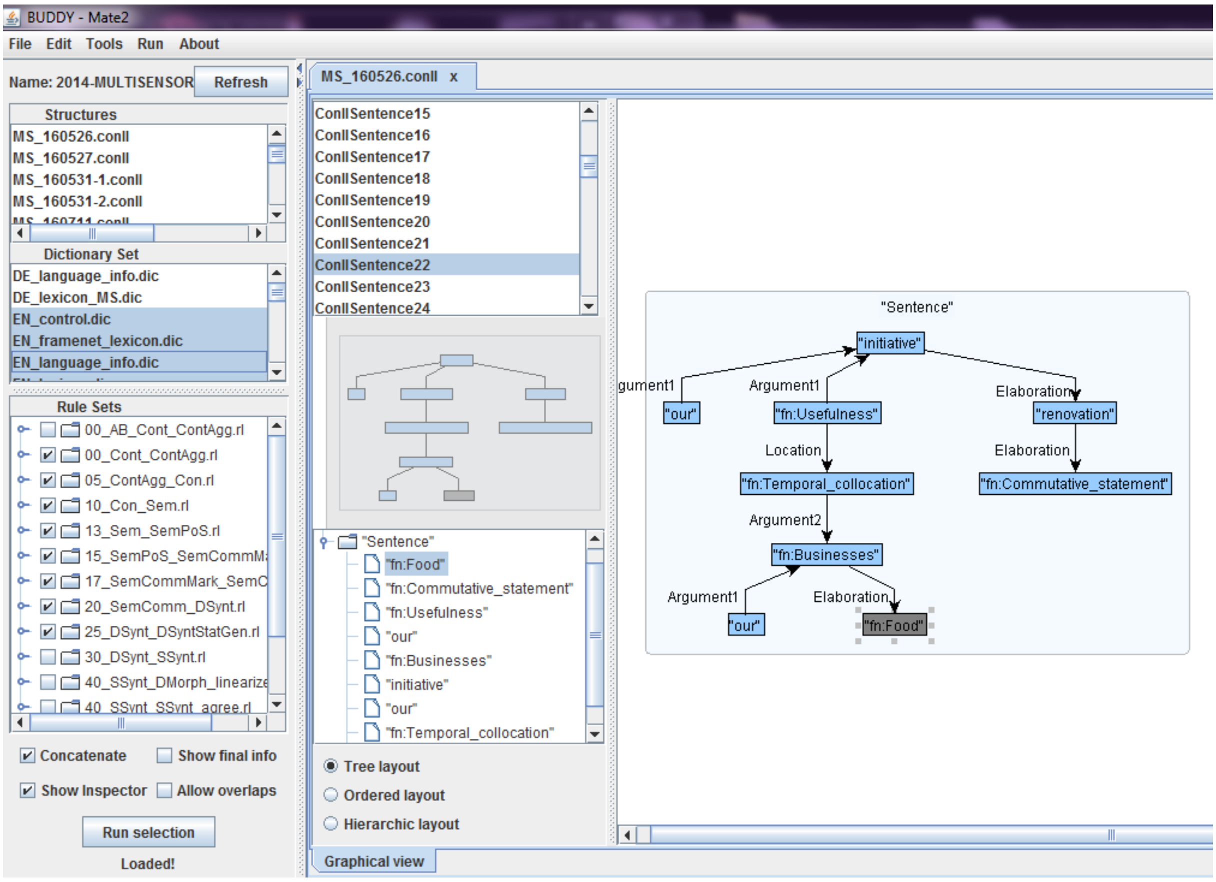



Figure 4: A screenshot of MATE-2 (Graph Editor view)

MATE-2 contains the following components:

- Project Browser

It is used to open and navigate through the different resources of a project. It is the leftmost interface control, which contains 3 list fields, some options, and the “Run Selection” button. Each list field corresponds to a resource used for the transduction:

  - Structures: a list of input structures on which the rules can be applied in order to create new structures;
  - Dictionary Set: a list of lexical resources used by the rules;
  - Rule Sets: a list of rule sets (= grammars). Each rule set performs one transduction.

- Resource Editor

  Each of the three resources can be opened in an editor tab by double clicking on it.

- Graph Editor

  The Graph Editor contains five fields (see Figure 4):

  - Graph List: the list of graphs contained in one file (top left);
  - Graph Global View: the global view of the selected graph (middle top left);
  - Graph Node List: the list of nodes of the selected graph (middle bottom left);
  - Graph layout options (bottom left);
  - Graph View: the complete graph view of the selected graph (right).

- Rule Editor

  The Rule editor contains two main fields and some parameters (see Figure 5):

  - Rule List: the rule tree (left).
  - Rule View: the complete rule view of the rule selected in the rule tree (right).

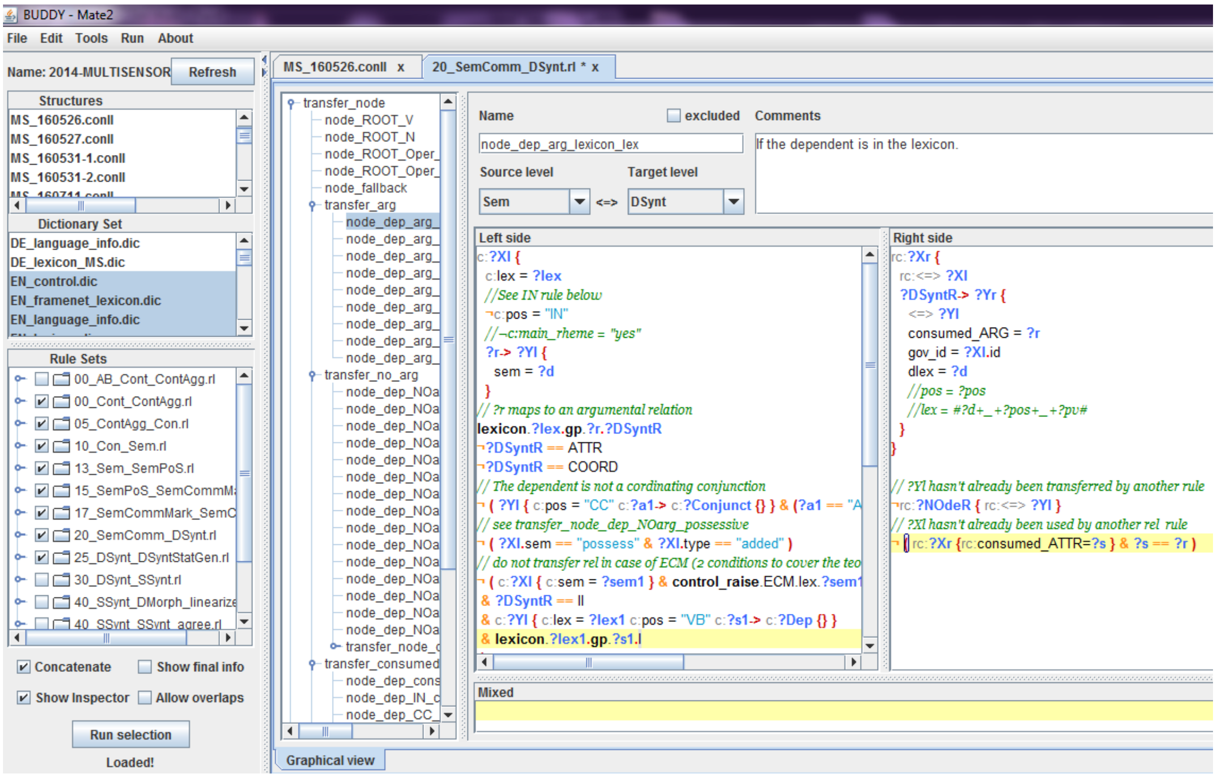


Figure 5: A screenshot of MATE-2 (Rule Editor view)

The Lexicon editor's advanced view is still at an early development stage at this point.

## 3.3 Graph-transduction grammars

This section focuses on the extension of UPF's multilingual discourse generators, developed for multilingual report generation in a series of European projects, to an incremental description generator that is coordinated with MindSpaces' Knowledge Base (T5.3, T5.4). The work has been carried out as foreseen during the first 18 months, thus, the present deliverable contains a description of the approach (Section 3.3.1 ) and the respective implementation (Section 3.3.2 ). This section also contains a part that summarises what has been advanced on during the first year and a half of MindSpaces, which includes a report on preliminary quantitative evaluations performed before the first Prototype (Section 3.3.3 ).

### 3.3.1 Approach

In our approach, the text generation consists of two sub-modules: sentence packaging (aka text planning) and linguistic generation. The latter is split into several modules that address the tasks of sentence structuring (choosing the words to be used and organising them syntactically), word ordering, and morphological agreement resolution. The advantage of splitting text generation into specific tasks is to allow for a precise and independent modelling of each level of language description (semantics, syntax, topology, morphology).

This is one of the central ideas of the Meaning-Text Theory (Mel'čuk 1988), which serves as a theoretical framework for the generator.

Text generation starts from the ontological assertions that comprise the selected contents of the Knowledge Base (T5.3, T5.4), and thus the ontological structures must be mapped to linguistic structures before the process can start. The generation is performed step by step, by successively mapping one level of representation onto the adjacent one:

- Ontology
- Conceptual Structure
- Semantic Structure
- Packaged semantic structure
- Deep-Syntactic Structure
- Surface-Syntactic Structure
- Morphologic Structure
- Sentence

In the following, we describe the role of each transition.

**From Conceptual Structure to Semantic Structure (SemS): choosing the meanings in each language**

The conceptual structure is mapped to a language-specific structure according to the available meanings (*semantemes*) in the concerned language (namely, English, Spanish, Greek, and Catalan). In order to illustrate the difference between a concept and a semanteme, consider the case of the concept of *measurement*. For example, the wind, as a physical event, can be measured, and this can be expressed in combination with the meaning *speed* in English. By contrast, in some languages, no meaning is available in order to realise *speed* in combination with *wind*. Mentioning *wind* with a rating is enough in order to understand that we are talking about wind speed. In English too, it is actually common not to mention *speed*, and the organisation of the concepts must allow for choosing one way or another of combining the meanings. In theory, a semanteme can be lexicalised by many different words, see for instance the semantic dictionary entry 'CAUSE': CAUSE { lex = cause_N | lex = cause_V | lex = contribute |lex = responsible | lex = due | lex = because | etc.}

However, in MindSpaces, a simplified and more practical view has been applied, considering the lexical units (i.e., words, as opposed to meaning units) such as 'cause_V' (*cause* as a verb) are the basic meaning units in the semantic structure. In practice, most of the time the semantic structure simply serves for the introduction of the lexical units in the target language, so the semantic structure for the generation in English in our running example would be the same as in Figure 3.

The semantic structure is unambiguous: each semanteme is the argument of a predicate and is numbered by the *valency* (or subcategorisation frame) of the predicate, through the relation linking the two of them. Each language has its own set of predicates, and each predicate has its own valency.

## Text planning / Sentence packaging: defining the boundaries of sentences

If several discursive units such as the one shown in Figure 3 (the group of nodes called "Sentence" is what we call here a discursive unit) are provided to the generator, each of them will be realised by default as an independent sentence. In order to group different units into complex sentences, we need to perform an "aggregation", or a "packaging" of the information, in two steps (Figure 6, Figure 7). First, we look for shared pairs of predicate and subject argument in the input: if the object arguments of two unlinked predicates have the same relation with their respective predicates, they will be coordinated. For instance, if the generator receives two separate units corresponding to *"the walls are blue"* and *"the ceiling is blue",* these two units will be rendered as one single sentence *"the walls and the ceiling are blue".*

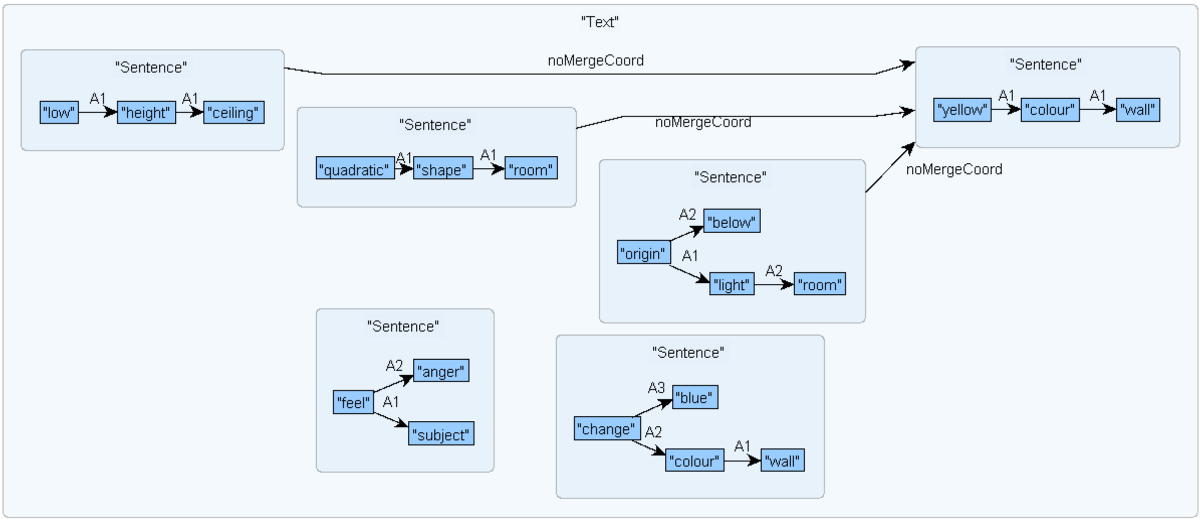



Figure 6: Sentence packaging through two-step aggregation of triples (step 1)

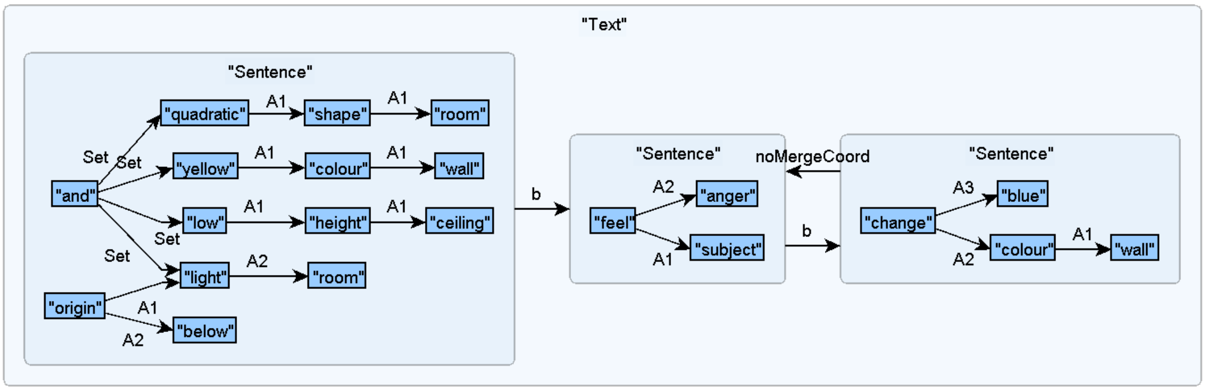



Figure 7: Sentence packaging through two-step aggregation of triples (step 2)

Second, we check if an argument of a predicate appears further down in the ordered list of discursive units. If so, the units are merged by fusing the common argument; during linguistic generation, this results in the introduction of postnominal modifiers such as relative and participial clauses or appositions. For instance, if there is information about the fact that the walls did not have an adequate colour, and that this colour was changed, *colour* is the common argument triggering the fusion: "*the **colour** of the wall, **which** was not adequate, was changed to blue*". In order to avoid the formation of heavy nominal groups, we allow at most one aggregation by argument. An example result of sentence packaging is shown in Figure 8. Referring expressions are introduced during the next steps.

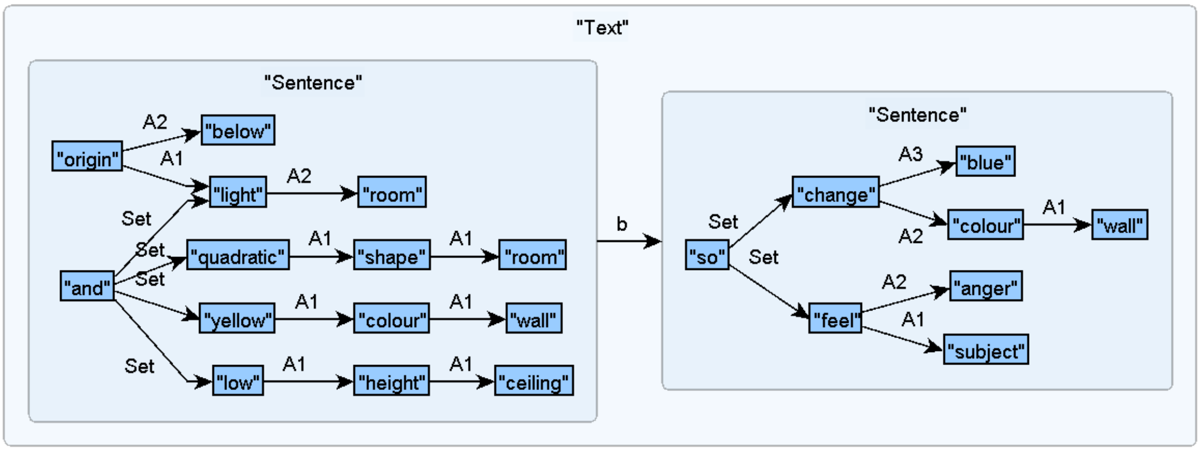



Figure 8: Semantic structures after sentence packaging

One more action is performed during text planning, which is the determination of the communicative structure. In order to realise a sentence, it is necessary to give it a communicative orientation: *what are we talking about? What do we say about it?* The former is marked as *theme* of the sentence, the latter as *rheme*, based on the semantic relations and the nodes in each connected graph. Anything that does not belong to any of these two spans is by default a *specifier*. Each lexical unit is included in a communicative span (theme, rheme, specifier), which can contain any number of lexical units.

For the structures in Figure 8, the nodes '*yellow*', and '*feel*' have been identified as being the main nodes of their respective sentences. All the nodes have been assigned a part of speech and linked to an entry in a lexical resource: for instance, *feel* is linked to the entry 'feel_VB_01' according to the PropBank nomenclature (Kingsbury and Palmer 2002).

**From Semantic Structure to Deep-Syntactic Structure (DSyntS): lexicalising and defining the sentence structure**

During the transition from Semantic Structure to Deep-Syntactic Structure, the semantic graph with the communicative structure is mapped onto a tree: the main node of the rheme will be the head/root of the sentence, that is, the main verb, while the rest of the rheme generally corresponds to the objects and adverbs, and the theme to the syntactic subject. From the root, the whole tree is built node by node.

A lexicon indicates what a syntactic predicate requires in order to form a correct sentence in a language (syntactic combinatorial). For instance, the verb '*begin*', as most verbs, requires a noun or a non-finite verb as its subject. The subject may also have arguments, also restricted by the syntactic combinatorial.

Only meaningful units (lexical units) are part of the DSyntS; in other words, there are no grammatical units that lack semantic content at this point (bound prepositions, auxiliaries, etc.). The DSyntS can also contain abstract lexemes (collocates), formalised as Lexical Functions (LFs). Those LFs are given a value (a concrete label) during the DSyntS-SSyntS mapping (see next subsection) based on the combination with other words. For instance, the abstract lexeme *Magn*, which means 'a high degree of', would be realised as '*heavy*' in combination with '*rain*', but as '*deep*' in combination with '*sleep*'. In our running example, '*be*' is introduced as a support verb for '*yellow*' to be realised as the main element in the sentence (adjectives generally cannot be the main element in a sentence).

Figure 9 shows that the main syntactic node of the sentences are the roots of the trees, and that all other elements are organised around each main node. Instead of a pure predicate-argument structure, the edge label reflects the syntactic structure of the sentence, in particular the opposition between arguments (*I*, *II*, *IV*) and modifiers (*ATTR*). Co-referring nodes are linked together with a blue-dotted line; these coreference links are used to introduce referring expressions (e.g., pronouns) in the next steps.

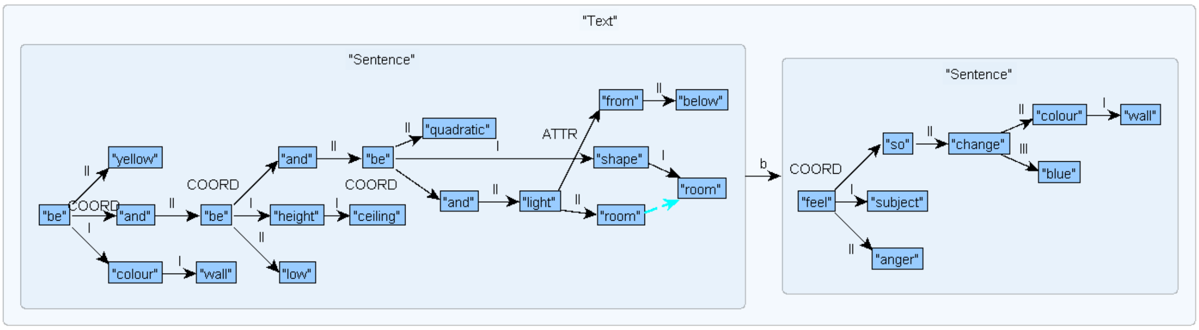


Figure 9: Deep-syntactic structures that correspond to the semantic structures in Figure 8

**From Deep-Syntactic Structure to Surface-Syntactic Structure (SSyntS): introducing all idiosyncratic information**

Once the structure of the sentence has been defined and all the meaningful words have been chosen, non-meaningful units need to be introduced. In the lexicon, an entry of a word indicates which preposition, case, finiteness, number, etc. has to be inserted on its dependent. For instance, the DSyntS configuration "colour-I->wall" means that the noun 'colour' has the noun 'wall' as its first ('I') argument, as in Figure 9. In this case, the entry of 'colour_NN_01' indicates that the dependent 'I' must be introduced by the preposition '*of*', which is a so-called governed (bound) preposition. Then other non-lexical nodes such as governed conjunctions, auxiliaries, determiners, expletive subjects, etc. are introduced. In our example, the modal auxiliary 'may' is introduced in order to verbalise the low confidence score seen in the ontological structure, and mapped to the modality features in Figure 3.

Lexical Functions must also be resolved during this transition: most words of the lexicon are the keywords of one or more LFs. The value(s) of the LFs is stored in the entry of a word:

'*heavy'* as the value of the LF *Magn* in the entry of '*rain'*, for instance. '*Pouring'* would be another value for the same LF of the same word.

Finally, the generic syntactic relations found in DSyntS are refined into more idiosyncratic relations that convey very accurate syntactic information, instead of semantic (i.e., argument numbers). For instance, the DSyntS relation 'I' can be mapped to *SBJ* (subject) if the verb is active, *OBJ* (object) if the verb is passive, *NMOD* if the head is a noun, etc. A *SBJ* has the syntactic property to trigger an agreement on the verb, to undergo demotion in some conditions, and to be realised before the verb in a neutral sentence. An *OBJ*, on the contrary, appears by default after the verb, can undergo promotion, and is cliticisable (Riemsdijk 2011) with an accusative pronoun. An *NMOD* cannot be promoted or demoted, does not trigger any agreement, and always has to be realised to the right of its governor. Figure 10 shows the surface-syntactic structure with functional words and language-specific syntactic relations.

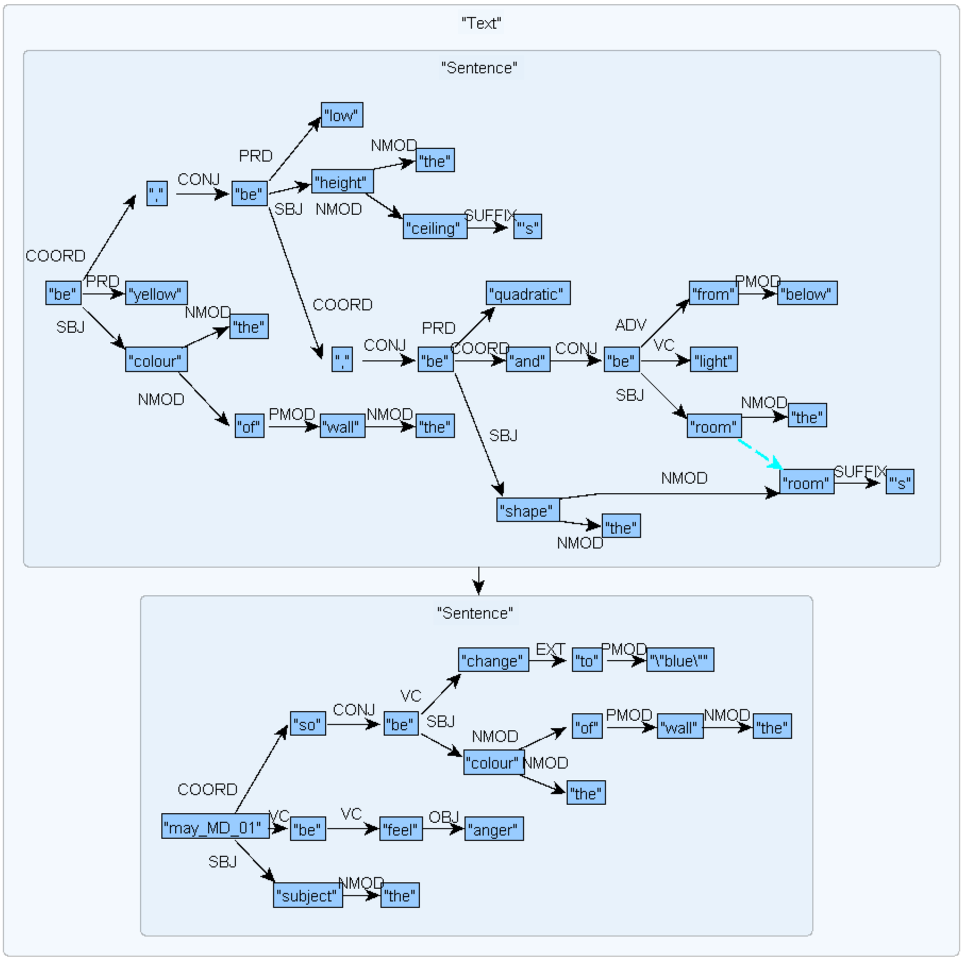



Figure 10: Surface-syntactic structures that correspond to the deep-syntactic structures in Figure 9

**From Surface-Syntactic Structure to Morphologic Structure (MorphS): resolving word agreements and word ordering**

Thanks to the idiosyncratic set of surface-syntactic relations, all agreements between the components of the sentence can be resolved. Every word of the sentence contains all the indications to make the production of the final form possible. This can be done either by creating a full-fledged dictionary containing entries under the form, e.g., '*be<VB><IND><PAST><3><SG> = was*', or by using some automata based on inflection schemas, such as Two Level Morphology, to automatically inflect forms. Capitalisation can be introduced when necessary.

Another advantage of using the idiosyncratic set of surface-syntactic relations is that the issue of order between the components of the sentence can be resolved effectively; for instance, in a given language, subject goes before its governing verb, a determiner before its governing noun, etc.

Figure 11 shows that at this level, the words carry all the necessary information for inflection, i.e. part-of-speech, mood, tense, person, and number (the small window on the top right of the figure, which shows information related to 'be'). The precedence relations are in red.

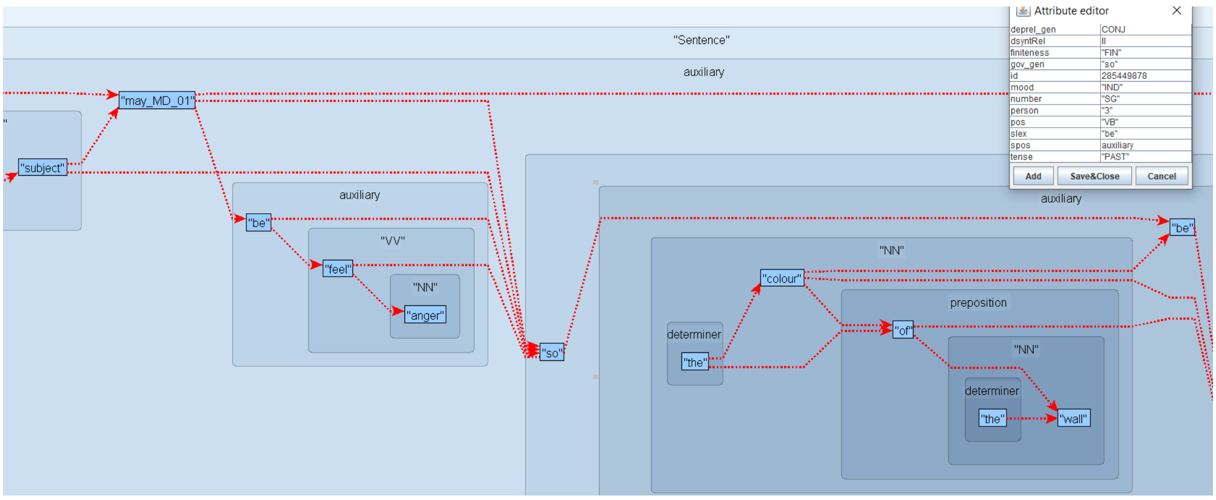



Figure 11: A (linearised) morphological structure that corresponds to the first surface-syntactic structure in Figure 10

**From Morphologic Structure to Sentence: finalising the sentence**

Once all the words are ordered, punctuation marks are introduced (periods and commas around descriptive modifiers), the final form of the words is retrieved, and the sentence is ready to be delivered to the next module. In the case of the running example shown throughout this section, the output would be the following:

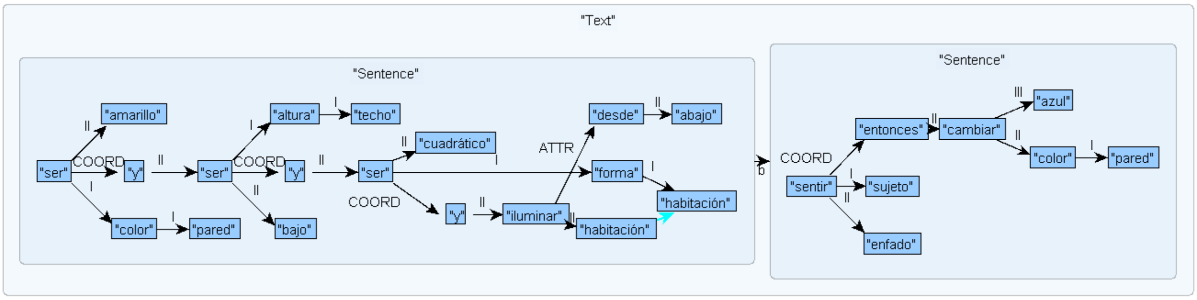
> *The colour of the walls was yellow, the ceiling 's height low, the room 's shape quadratic and the room was lit from below. The subject may be feeling anger so the colour of the wall is being changed to "blue".*

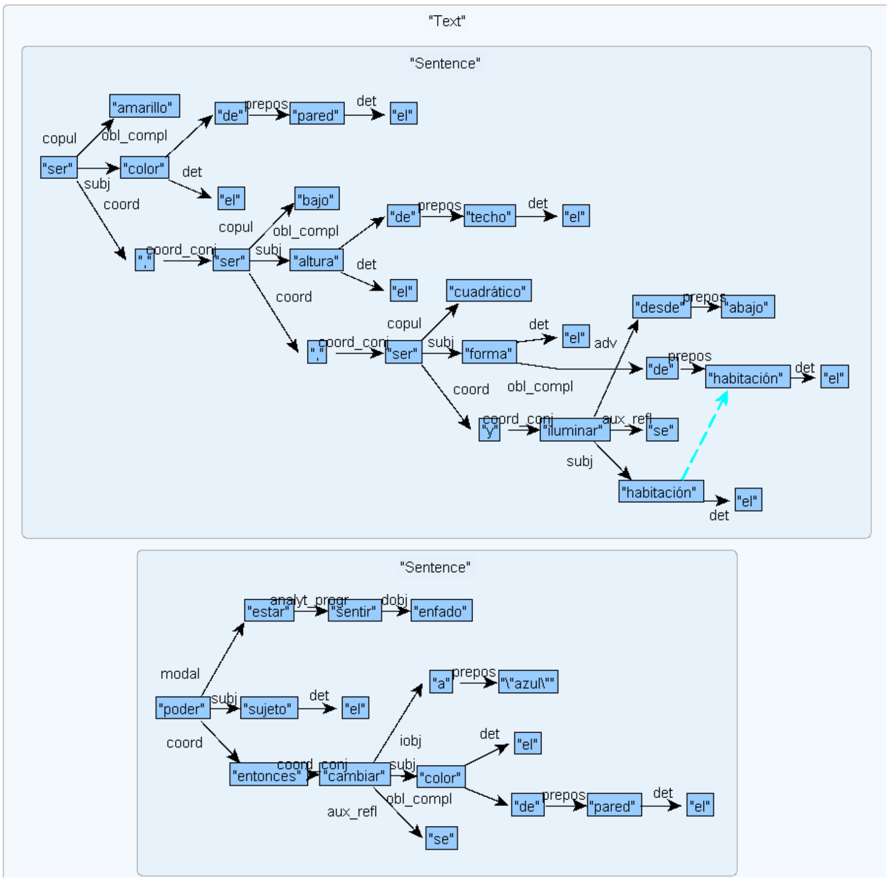
Other texts about the adaptation of the environment to the subject's detected position and emotional state are also currently covered:

*The subject was sitting. The subject was probably feeling surprise so the plant wall 's scale was changed to 0,8 and a lamp was removed.*

**Examples in Spanish and Catalan**

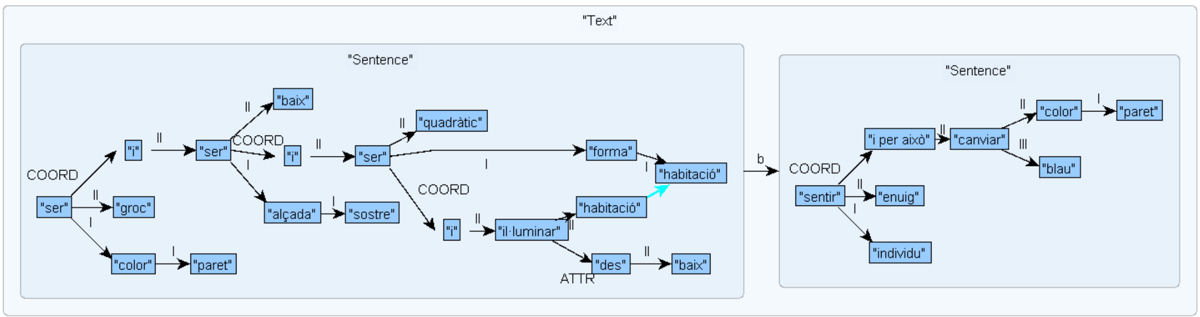
Note that in languages other than English, the structures start being different in Deep Syntax. Deep-Syntactic Structures (DSyntSs) are very similar across languages, as can be seen with the examples below, in which mostly the words change. However, when getting closer to the final sentence, the structures get increasingly different, as for instance, the Surface-Syntactic Structures (SSyntSs) below. Note in particular how the relation labels are different, and how the passive is realised in Spanish and Catalan (impersonal *se/es*). The morphological interactions are also quite richer in Spanish and Catalan, with concatenations (*de+el=del*, 'of+the=of.the' on Spanish; *la+alçada=l'alçada*, 'the+height = th'height', in Catalan, etc.). Since lexical resources for Catalan are not fully investigated at this stage, we use basic lexicon to cover the basic events in VR for the first Prototype. See Figure 12 and Figure 13 for examples.

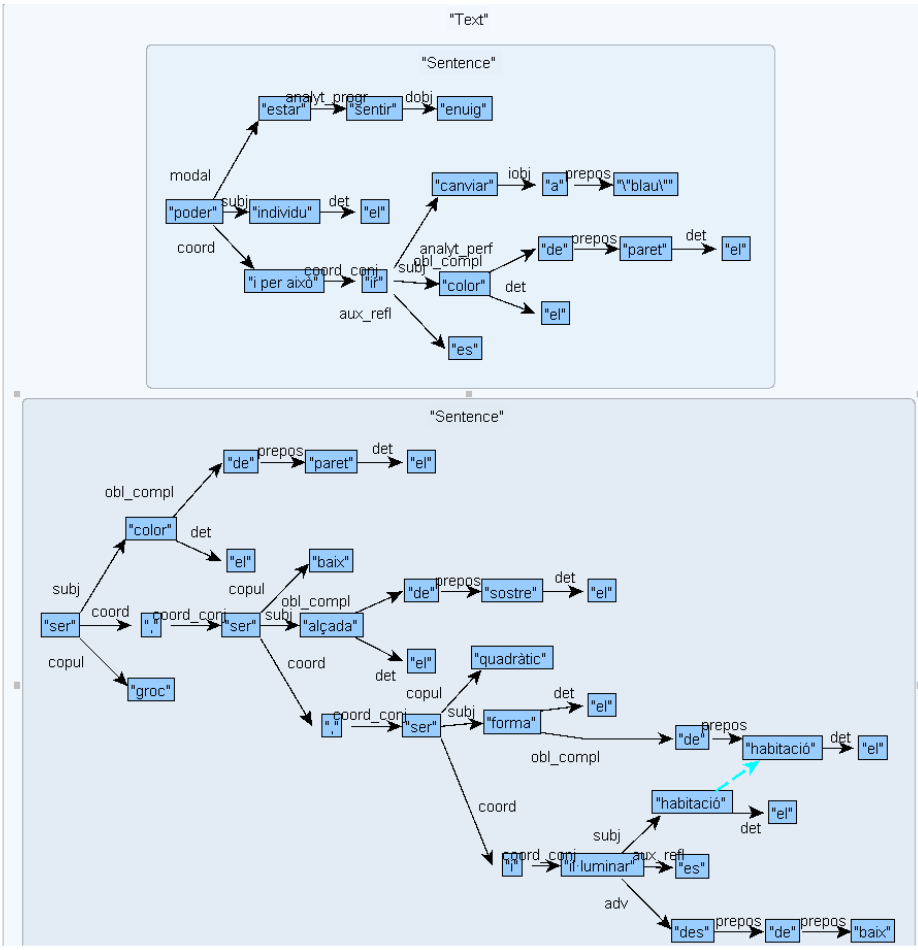





*El color de las paredes era amarillo, la altura del techo baja, la forma de la habitación cuadrática y se iluminaba la habitación desde abajo. El sujeto podía estar sintiendo enfado entonces el color de la pared se cambió a "azul".*

Figure 12: DSyntSs, SSyntSs and final text samples in Spanish






*El color de les parets era groc, l' alçada del sostre baixa, la forma de l'habitació quadràtica i s' il·luminava l' habitació des de baix. L' individu podia estar sentint enuig i per això el color de la paret es va canviar a "blau"*

.

Figure 13: DSyntSs, SSyntSs and final text samples in Catalan

### 3.3.2 Implementation with the FORGe generator

The basic text generation implementation consists of manually crafted graph-transduction grammars for each transition between two consecutive layers. In combination with the rules, dictionaries of two different types are required: one that describes the syntactic properties of these words (lexical dictionary), and one that contains the inflection patterns of each word (morphological dictionary). We manually crafted language-specific dictionaries that cover the texts foreseen for the first prototype.

In order to reach large-coverage, we have been experimenting on the extraction of subcategorisation patterns from lexico-semantic resources such as PropBank (Kingsbury and Palmer, 2002), NomBank (Meyers et al., 2004) and VerbNet (Schuler, 2005); see (Mille and Wanner, 2015). The sample entry in Figure 14 shows the syntactic properties of the verb 'give', which has three nominal arguments in government patterns with the third one being introduced by the preposition 'to' (*give something to someone*).

```
"give_VB_01":_verbExtArg_{
 vncls = "13.1"
  pbID = "give.01"
  pbsenseID = "01"
  lemma = "give"
  gp = {
    I = {
      pos = NN
      rel = SBJ
    }
    II = {
      pos = NN
      rel = OBJ
    }
    III = {
      pos = NN
      rel = IOBJ
      prep = "to"
    }
  }
}
```

Figure 14: Syntactic properties of the verb 'give' extracted from lexico-semantic resources

The method has proven to be useful in English, but not successful enough to be applied to other languages so far, especially given the lack of parallel resources in PropBank and NomBank for these languages. The generation module will cover all languages involved in interactions with the Knowledge Base in the project, that is, English, Catalan, Greek, French and Spanish, with different coverage, directly related with the size of the respective lexicons.

### 3.3.3 Advances in MindSpaces

In this section, we report on the advances made in the framework of MindSpaces during the first 18 months of the project. Since the work on text generation was initiated on Month 7 (July 2019), we use as a starting point the state of the generator as found in the final deliverable of the H2020 TENSOR project (project number 700024, D6.7, 30/06/2019).

#### Extensions to language-independent rules

In addition to adding functionalities to the generator, some detected issues after the evaluation at the end of TENSOR were fixed. Even though as shown in the TENSOR deliverable, FORGe received excellent evaluation marks at the WebNLG challenge (Gardent et al. 2017b), especially in the human assessments, according to which it was close to the quality of human-written text After an error analysis of FORGe's outputs, we found a series of general problems impairing the quality of the generated texts in terms of contents and grammaticality. In particular: (i) some properties were not verbalised due to the failure to produce relative clauses in some specific cases; (ii) the aggregations were at times excessive, erroneously merging verbs with different tenses (e.g., *X impacted Y, which was impacted by Z, instead of X and Z were impacted by Y*), failing to merge (e.g., *X was impacted by Y. Z was impacted by Y*), or leading to an ungrammatical outcome, with for instance the presence of several "*also*"; (iii) the construction of some relative clauses were faulty, as e.g., *X can a variation of which be Y,* instead of *X, which can be a variation of Y*; (iv) the referring expression module was applying excessively, resulting in ambiguous pronouns, and sometimes incorrectly pronominalising non-human entities with *he*, or failing to pronominalise locations, such as in "*A person is standing next to the desk. There is a plant next to the desk"*, the second sentence being more naturally rendered by "*There is a plant there"*; (v) some agreements were not solved (e.g., "*the main ingredient are*"); (vi) some determiners were erroneously introduced, and some others were not in the correct form ("*a"* instead of "an"). All these issues could potentially affect MindSpaces outputs would they be found in the inputs to the generator. Many occurrences of these issues were fixed in the grammars by modifying and adding rules, and some new features were added, as for instance new rules to cover more cases of embedded clauses generation. For developing the grammars, we used the collection of MindSpaces inputs, 6- and 7-triple inputs from the WebNLG training data, and the whole WebNLG development set. A qualitative evaluation of the new outputs is provided in the next Section.

**Extensions to the Spanish rules**

The most important rules added for Spanish are (i) rules introducing the surface-syntactic relations, based on which linear order and morphological agreements are resolved, (ii) rules for gender and number agreements in noun groups and auxiliary constructions, and (iii) word ordering rules. Note that the rules for Spanish also apply to other Romance languages with similar features (e.g., French, Italian, etc.).

For designing the rules, we followed the approach of AnCora-UPF (Mille et al., 2013), a Spanish dataset in which each dependency relation is associated with a set of syntactic properties. For instance, a subject is characterised by being linearised to the left of its governing verb (by default), by being removable, by triggering the number and person agreements on the verb, etc. During the linguistic generation stage, 27 out of the 47 relations proposed in AnCora-UPF - namely *adjunct, adv, agent, analyt_fut, analyt_pass, analyt_perf, analyt_progr, aux_phras, appos, attr, compar, coord, coord_conj, copul, det, dobj, iobj, modal, modif, obl_compl, obl_obj, prepos, punc, quant, relat, sub_conj,* and *subj* - are currently supported.

In order to generalise the ordering rules across languages, the dependencies were introduced in the lexicon with details about how they are linearised with respect to their governor (vertical ordering). Generic linearisation rules also apply. For instance, for the *copul* dependency (such as between *be* and *detected*), pronominal dependents are linearised BEFORE the finite verb, and the other dependents AFTER it. If several dependents end up at the same height with respect to their governor, they need to be ordered with each other. 21 rules were added to manage these horizontal orderings. They facilitate the ordering of, for instance, determiners before the adjectives, or small adverbial groups before the objects. Finally, 18 rules for resolving the agreements between verb and subject, adjective/determiner and noun, copulatives and subjects, etc. were implemented.

For instance, the structure “Una persona$_{\textit{-3-FEM-SING}}$ y un sofa$_{\textit{-3-MASC-SING}}$ *<-subj ser copul->* detectado”, will be linearised and inflected as follows: “Una persona y un sofa son detectados” (lit. “A person and a couch are$_{\textit{3-PL}}$ detected$_{\textit{MASC-PL}}$”).

**Crafting the Spanish dictionaries**

Several types of dictionaries are needed for generation: (i) a dictionary that maps the input meanings/concepts onto lexical units of a particular language (called concepticon), (ii) a dictionary that contains the combinatorial properties of each lexical unit (lexicon), and (iii) a dictionary with the full forms of the words (called morphologicon). Some other information, such as linearisation properties of dependencies, are also better stored in the lexicon in order to allow for more generic (hence less numerous) rules.

In our generation architecture, the input structured data (ontology substructures or DBpedia properties) are mapped to PredArg structures. For the WebNLG challenge, English was the only language to generate, so the labels of the nodes in the PredArg templates were in English. In order to take advantage of the templates developed for FORGe within WebNLG 2017, we also use these structures with English vocabulary as input to the generator. Thus, we manually crafted the concepticon (350 entries so far), in which the keys are the predicates from the templates, and the values are lexical units in Spanish; for instance, the predicate *locate* is mapped to the Spanish verb *estar_VB_04* (`be').

In the lexicon, lexical units such as *estar_VB_04* are described. This fourth entry for estar corresponds to a verb that has two arguments, the second being an adverb or a prepositional group. *estar_VB_01* is the simple copula, *estar_VB_02* is the existential be, which has only one argument, and *estar_VB_03* is the auxiliary. Each lexical unit contained in the concepticon is a key in the lexicon. A fine-grained lexicon has been crafted manually for the MindSpaces and WebNLG experiments, and we developed an automatic conversion of the large-scale AnCora-Verb (Aparicio et al, 2008) to obtain a large coverage resource. Finally, in order to store the surface forms of the inflected words, we crafted a very small morphological dictionary of about 600 entries to cover the needed forms in the experiments.

Table 7 and Table 8 summarise the state of the FORGe language generator at the beginning of the MindSpaces project (01/01/2019). As mentioned at the beginning of this section, we take as reference the state of the grammars as they were described in the final deliverable of the TENSOR project. Note that the method to count the rules and whether they are language-independent was updated since then[16], so the numbers in the first column (M0) are slightly different from the original ones, but they are fully comparable to the numbers in the column that corresponds to MindSpaces' month 18. TENSOR did not cover Greek, so we use as reference the state of the grammars as described in the intermediate deliverable of V4Design (H2020-779962, 30/04/2019, D5.2).

Table 7: Rule count after the first 18 months of MindSpaces

| | | **MindSPaces M0** | **MindSpaces M18** |
|---|---|---|---|
| Languages supported | | EN, ES, DE, EL | EN, ES, DE, EL, CA |
| Number of rules | all | 1,555 | 1,892 |
| % of language-inde pendent rules | all | Con-SMorph (1,555) : 70% | Con-SMorph (1,892) : 74% |
| | groups | 1 - Con-Sem (409) : 98%<br>2- Aggregation (216) : 100%<br>3 - Sem-DSynt (202) : 70%<br>4 - DSynt-SSynt (397) : 44%<br>5 - SSynt-DMorph (188) : 56%<br>6 - DMorph-SMorph (143) : 36% | 1 - Con-Sem (450) : 97%<br>2 - Aggregation (263) :100%<br>3 - Sem-DSynt (227) :78%<br>4 - DSynt-SSynt (617) : 55%<br>5 - SSynt-DMorph (218) : 53%<br>6 - DMorph-SMorph (117) : 50% |

In the above table, a quantitative assessment of the generator is reported, with a count of the rules and lexical entries and a description of the covered phenomena in the different languages. The rule sets have generally been made more language independent; only the SSynt-DMorph transition contains a higher proportion of language-specific rules. The reason is that the SSynt-DMorph transition is language-specific by nature, with the modelling of phenomena that are often highly idiosyncratic. Increasing the coverage of this grammar usually means adding language-specific rules, which makes their proportion increase. The number of rules during the DSynt-SSynt transition was substantially increased due to the addition of a new module that performs syntactic aggregations, which covers cases that the semantic aggregation grammars cannot see because of the fact that the final words are not in the structure at that point. For instance, the number of floors of a building would be represented by the property *floorCount*, and its area by the property *surfaceArea*. It is not obvious that these two properties can end up in the same sentence, but in English both properties can end up being verbalised with *has*, hence aggregation rules based on actual

[16] In particular, in the original count, the rules of two grammars that had the same function were counted, as opposed to only one of them now; also, in the original count, some rules were not detected as being language-specific, but are now.

words could be safely applied: *the building has X floors and a surface of Ym2*. The rules in the last transition have been slightly compacted.

Table 8 shows a more detailed analysis with respect to which phenomena are being covered by the generator in the different languages. At the beginning of the project, English was the most developed, and Spanish had a slightly higher coverage than French, with most basic features supported. There was nothing language-specific for Catalan. Little progress has been made on French so far, with an emphasis on improving the general quality of the generator and porting it to Spanish, Greek, and Catalan.

Table 8: Linguistic coverage in the different languages (Levels 0, 1, 2, 3, 4: Null, Reduced; Intermediate; Advanced; Very advanced). Green: advance made

| | **MindSpaces month ->** | **0** | **18** | **0** | **18** | **0** | **18** | **0** | **18** | **0** | **18** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **gr** | **Supported phenomena \| Language ->** | **CA** | **CA** | **EL** | **EL** | **EN** | **EN** | **ES** | **ES** | **FR** | **FR** |
| 2 | **Sentence planning** | 2 | 3 | 2 | 3 | 2 | 3 | 2 | 3 | 2 | 3 |
| 3 | **Lexicalisation** | 0 | 1 | 1 | 2 | 2 | 2 | 1 | 2 | 1 | 1 |
| 3 | **Sentence structures** | 0 | 2 | 1 | 2 | 2 | 3 | 1 | 2 | 1 | 1 |
| 3 | **Coordinations** | 0 | 1 | 0 | 1 | 2 | 2 | 1 | 2 | 1 | 1 |
| 4 | **Auxiliaries and modals** | 0 | 1 | 0 | 1 | 3 | 3 | 1 | 3 | 1 | 1 |
| 4 | **Relative clauses** | 0 | 1 | 0 | 1 | 2 | 3 | 1 | 2 | 1 | 1 |
| 4 | **Support verb constructions** | 0 | 1 | 0 | 1 | 1 | 2 | 1 | 2 | 0 | 0 |
| 4 | **Nominal compositionality** | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 4 | **Subcategorised information introduction** | 0 | 1 | 1 | 2 | 2 | 2 | 1 | 2 | 1 | 1 |
| 4 | **Structure well-formedness checks** | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| 4 | **Referring expressions** | 0 | 2 | 0 | 2 | 2 | 2 | 0 | 2 | 0 | 0 |
| 4 | **Syntactic aggregation** | 0 | 2 | 0 | 2 | 0 | 2 | 0 | 2 | 0 | 0 |
| 5 | **Linearisation** | 0 | 2 | 1 | 2 | 2 | 2 | 1 | 2 | 1 | 1 |
| 5 | **Number, person agreements: verbs** | 0 | 2 | 1 | 2 | 2 | 3 | 1 | 3 | 1 | 1 |
| 5 | **Number, gender agreements: adjectives** | 0 | 2 | 0 | 2 | 2 | 3 | 1 | 3 | 1 | 1 |
| 5 | **Case agreements: nouns, adjectives** | 0 | 2 | 1 | 2 | 2 | 2 | 0 | 2 | 0 | 0 |
| 6 | **Concatenations** | 0 | 2 | 0 | 0 | 0 | 0 | 1 | 2 | 0 | 0 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 6 | **Elisions** | 0 | 2 | 0 | 0 | 1 | 2 | 0 | 2 | 0 | 0 |
| | | | | | | | | | | | |
| | **MindSpaces UC1 support** | 0 | 1 | 0 | 1 | 0 | 2 | 0 | 2 | 0 | 1 |
| | **MindSpaces UC2 support** | 0 | 2 | 0 | 1 | 0 | 3 | 0 | 3 | 0 | 1 |
| | **MindSpaces UC3 support** | 0 | 2 | 0 | 1 | 0 | 3 | 0 | 3 | 0 | 1 |

### 3.3.4 Quantitative and qualitative evaluations of FORGe on a standard RDF dataset

The MindSpaces use cases require the generation of texts of one domain and with little variation in the inputs. In order to test the capacities of the MindSpaces generator on a wider scale, we evaluated it again on the WebNLG dataset (in the TENSOR deliverable we reported the best score obtained by FORGe at the WebNLG 2017 international challenge). The reason for using the WebNLG challenge dataset as reference basis is that it is the most recent and comprehensive dataset with respect to text generation from RDF data (a standard ontology model as the one used in the MindSpaces Knowledge Base) that has been specifically designed to promote data and text variety (Pérez et al, 2016). Moreover, it allows the direct comparison with the generators that participated in the challenge. In order to ensure future comparisons with machine learning-based systems in terms of their best obtained performance, only the seen categories subset of the original test set has been considered, i.e., only inputs with entities that belonged to DBpedia categories that were contained in the training data.

In this section, we detail how we built a new dataset for evaluating the outputs of the generator and describe the results of the automatic and human evaluations.

**Selection of triples for evaluation**

For evaluation purposes, we compiled a benchmark dataset of 200 inputs, i.e., sets of DBpedia triples, with sizes ranging from 1 to 7 triples, using as a reference pool the WebNLG challenge test set. The compilation methodology for our benchmark dataset implements a twofold goal: on one hand, we want to ensure that all properties appearing in the seen categories subset are included, on the other hand, and unlike the WebNLG human evaluation test set, we aim towards a more balanced number of inputs of different sizes. In practice, since the inputs of size 6 and 7 in the original seen categories subset of the WebNLG test set are 24 and 21 respectively, we chose to include all of them in the benchmark. 31 inputs for each of the remaining input sizes were subsequently added, by iterating over the reference test set and opting for the inclusion of inputs that: (i) contain different properties or properties combinations, rather than different property values, and (ii) contain, if none, the least possible number of properties that have been already selected in a previous iteration. In this way, the different input sizes are represented in a better proportion, avoiding possible biases that may be introduced when favoring some input sizes over other (indicatively in the WebNLG seen categories test set, the ratio of inputs of size 6 and 7 over those of smaller sizes ranges from 1 to 6 up to 1 to 9). Inevitably, the small number of inputs of size 6 and 7 initially available did not leave any space for selection. Hence, these inputs have a rather high degree of overlapping properties.

**Reference sentences**

The English reference texts are taken from the WebNLG dataset, for which there could be more than one reference per triple set. For Spanish, one single reference text was produced for each triple set, with natural and grammatical constructions containing all and only the entities and relations in the triples. The reference texts were written by a native Spanish speaker, having at hand the English references from the WebNLG challenge to serve as a potential model.

**Automatic evaluation**

The predicted outputs in English and Spanish were compared to the reference sentences in the corresponding language. Three metrics were used: (i) BLEU (Papine.ni et al., 2002), which matches exact words, (ii) METEOR (Banerjee and Lavie, 2005), which matches also synonyms, and (iii) TER (Snover et al., 2009), which reflects the amount of edits needed to transform the predicted output into the reference output. Table 9 shows the results of the automatic evaluation on the English and Spanish extensions proposed above using for each input its corresponding reference text(s). The first two rows show that in terms of automatic metrics, the extended FORGe and the 2017 FORGe have almost exactly the same scores on the English data (which are also very close to the WebNLG scores: 40.88, 0.40, 0.55). In other words, the quality improvements in English are not reflected by these metrics. To compare English and Spanish results, we calculated the scores using one sentence as reference (only one reference per text is available in Spanish). The English scores drop (third row) due to the way the scores are calculated by the individual metrics (BLEU matches n-grams in all candidate references, and METEOR and TER consider the best scoring reference). In the last row of the table, the scores of the Spanish generator look contradictory: the BLEU is 10 points below the English BLEU with the same number of reference (one), but METEOR is 8 points above, that is, the predicted outputs do not match the exact word forms, but they do match similar words. One reason for the low BLEU score could be the higher morphological variation in Spanish. However, the METEOR score is surprisingly high, actually even higher than the highest METEOR score at WebNLG, obtained by ADAPT and calculated with multiple references (0.44).

Table 9: English and Spanish scores according to BLEU, METEOR and TER, with one and all references on the 200-triples test set

| Reference set | BLEU | METEOR | TER |
|---|---|---|---|
| EN ($\text{All}_{FORGe-2017}$) | 39.87 | 0.40 | 0.58 |
| EN ($\text{All}_{FORGe-Ext}$) | 39.33 | 0.40 | 0.58 |
| EN ($1_{FORGe-Ext}$) | 29.18 | 0.38 | 0.65 |
| ES ($1_{FORGe-Ext}$) | 18.68 | 0.46 | 0.77 |

### Qualitative analysis of the results

In the 200 outputs of the 2017 generator, 275 errors were detected, compared to 166 in the current one in English (170 in Spanish), and 26.5% of the texts were error-free, as opposed to 43.5% now (45.5% in Spanish). In this section, we report on the examination of both English and Spanish outputs, in order to identify the main issues of the grammars in both languages. Outputs are available as supplementary material in the 2019 INLG paper (Mille et al., 2019b).

### English

The qualitative analysis of the generated English texts showed that the resulting texts are of a higher grammaticality and fluency than the 2017 ones. Below, we discuss the observed remaining errors and their respective causes.

*Determiners*

Although determiners are handled overall correctly, there are cases where a definite determiner should precede the mentioned NE. In some of these cases, for example in *acharya institute of technology was established in 2000*, the absence of the determiner can still be considered grammatically acceptable, while in others, for example in *arabian sea is located to the west of karnataka* and *st. louis is part of kingdom of france*, the determiner's absence is unequivocally erroneous. The missing determiner is traced back to the PredArg template that implements the involved DBpedia property and in particular to the assumptions underlying the semantic types of its respective arguments. For example, properties capturing information about administrative divisions (e.g., canton, state, city, country) and their respective *part-of relations*, as well as cardinal and intercardinal directions (e.g., west, southwest) range over entities denoting such subdivisions (i.e., names of cities, countries, regions, etc.) that in the general case do not admit a determiner. As a result, when an argument belongs to the exceptional cases, the generated text misses the determiner.

Definite determiners are missed with the property 'language', when referring to the language of a written work. The reason of this error lies in the discrepancy between the respective PredArg template that was defined based on the premise that the object value of this property is a language name (i.e., English, Italian), hence not admitting a determiner, and the form of the DBpedia language entities that in practice concatenate the language name with the word *language* (cf., *English language*). This type of error is the most frequent, being found about 65 times in the test set and representing about 40% of the total amount of errors (166).

This underlies the need for further normalisation of the DBpedia property values, so that during the PredArg templates instantiation, consistent linguistic features will be ensured for argument values of the same type.

*Tense*

Errors are observed with respect to the verb tense selection (6% of the errors). More specifically, in some cases the present tense is used instead of the past, as, e.g., in "*Alan Shepard, who graduated from NWC in 1957 with a M.A., is deceased. [...] He is a test pilot*."

This is a direct consequence of the fact that in the current implementation, tense selection does not take into account the temporal context as defined by the rest of the input triples.

*Aggregations*

Another type of error relates to the generation of unintuitive, yet still grammatical, constructs when aggregating the contents of more than one triple when certain properties are involved (11% of the errors). More specifically, when the property 'occupation' is selected to be expressed as a relative clause, it fails to append the occupation information to the referring entity as shown in "*Alan Bean, born in wheeler (Texas) on March 15, 1932, is from the United States (test pilot)*". A similar behaviour has been observed with the property 'category'. This is a result of the current implementation of aggregation that takes place in a single step and tries to avoid orphan clauses by attaching them to the closest reference head. Introducing iterative aggregation steps and incorporating semantic coherence information would mitigate such effects.

A related issue is, for instance, the way location information is verbalised in the presence of multiple subdivision references (15% of the errors), as for example, in "*the Acharya Institute of Technology is in Bangalore, Karnataka and India*", where the three involved location-denoting properties, namely '*city*', '*state*' and '*country*' have been aggregated in a semantics-agnostic manner. Navigating DBpedia and obtaining information about their interrelations would enable more fluent verbalisations. Fluency and meaning accuracy are also impacted when the input triples capture in practice *n*-ary relations. This is the case with the 'leader' and 'leaderTitle' properties, which in the absence of any semantic pre-processing before the instantiation of the PredArg templates result in verbalisations such as "*the leaders of Romania are the prime minister of Romania and Klaus Iohannis*", which does not communicate the fact that Klaus Iohannis is the prime minister.

*Subject/Object values*

Lastly, a number of disfluent verbalisations is the direct result of idiosyncrasies in the involved DBpedia properties and/or the respective subject and object values (4% of the errors). There are properties, that although meant to capture different types of information, are not used consistently, thus impacting the resulting verbalisations. The properties 'mainIngredient(s)' and 'ingredient(s)' are such an example, e.g., in an input about the dish "*Ayam Penyet*", which is described as having as main ingredient the "*fried chicken*" and as a further ingredient "*chicken*". Some minor errors such as unnatural word ordering (11%) or lexicalisations (8%) were also detected.

**Spanish**

The aforementioned errors listed for English are mostly independent of the language and thus also apply to Spanish, except from the first aggregation error, which does not appear due to a difference in the templates. The determiner error represents 30% of the total number of detected errors (51/170), the location aggregation 12%, the values and word choices 7%, the ordering 6%, the verbal tense 5%. However, despite its overall good quality, Spanish has some additional specific issues.

*English words*

There are some not-translated nouns (*52 minutes*) or phrases (*"está dedicado a Ottoman army soldiers killed in the battle of Baku"*), which, in addition of not being understandable, may produce subsequent morphological errors (21% of the errors).

*Morphology*

Morphological errors, mainly gender (invisible in English) and number disagreements, are found in the Spanish texts (5% of the errors). For example, in "*Dianne Feinstein es un senador de california*", (lit. 'Dianne Feinstein is $a_{MASC}$ $senator_{MASC}$ of California'), both *a* and *senator* should be feminine, but there is no information that D. Feinstein is a woman in the input.

*Complex relative clauses*

The main syntactic error is related to the genitive relatives with *cuyo* ('of which'), in particular when the antecedent is a location (5% of the errors). For example, in the sentence *"Alba Iulia, en el cual está el 1 Decembrie 1918 University"*, (lit. 'Alba Iulia, in the which is the 1 Decembrie 1918 University'), the proper pronoun should be *donde* 'where' instead of *en el cual*. Even when grammatically correct, sentences with these relative clauses tend to lack naturalness.

Other series of errors that produce sub-optimal Spanish constructions include occasional choice of a relative clause instead of a past participle modifier, and various other constructions that lack naturalness (10% of the errors).

# 4 TOWARDS AN ADVANCED MULTILINGUAL TEXT GENERATION

During the first half of the project, UPF has contributed to the methods of automated enhancement of the lexicon and the production of data that can be used to train statistical modules for the generation pipeline, in particular linearisers and morphology resolution tools. We used the freely available Universal Dependency datasets and converted them to be suitable for Natural Language Generation. With the obtained datasets, we organised an international shared task which took place in 2019 and was a success in terms of results and participation. Furthermore, an approach that showed the best result at the competition was adjusted to the needs of MindSpaces to generate personalised summaries from the crawled data within artistic solutions.

## 4.1 Automated lexicon compilation

In order to enrich the quantity (number of lexical entries) and the quality (the type of linguistic information) in the Spanish lexicon for the graph-transduction-based generator developed at UPF, we have converted the verbal and nominal AnCora lexicons into the UPF format.

The UPF ES_lexicon is a one-file dictionary that contains the syntactic description of lexical units in Spanish, especially, how they combine with other lexical units (i.e. required prepositions, functional words such as "*to*"). However, the Spanish lexicon has only little semantic information which has been added manually for lexical units in specific domains.

AnCora has currently two lexicons, which were developed together with the corpus: a verbal one, AnCora-Verb 2.0, and a nominal one, AnCora-Nom, where deverbal nominalizations are defined. AnCora-Verb includes for each verb its semantic class, its subcategorization, argumental structure and the roles of each of the arguments. AnCora-Nom includes for each noun, its denotative type, its WordNet synset, argumental structure and the roles assigned to the each of its arguments, as well as the information of the verb from which they derive. Both of them are a set of xml files, one for each of the lexical entries: 2828 xml files (one for each verb) and 1655 xml files (one for each noun). Finally, the xml file AnCora-Net-ES contains the links between verbal frames and PropBank rolesets, which have been manually validated, as well as the links to FrameNet, VerbNet, WordNet, and OntoNotes Groupings inherited from PropBank links using SemLink 1.1.

The automatic conversion went through a preliminary stage where the information in AnCora's lexicons was analysed and a proposal for matching it into the UPF format was set. The process went then through several rounds (detecting errors, exceptions and deciding to convert more information) until it was stable.

**Conversion of AnCora-Verb 2.0**

One of the main issues in the conversion of AnCora-Verb was matching the different classification of verbs in the UPF and AnCora lexicons. In fact, while UPF lexicons classify verbs only in two categories according to how arguments correspond to deep-syntax actants, AnCora-Verb classifies the verbs according to their lexical semantic structures (LSS) into 24 types. This was solved identifying the LSS in AnCora with an AO (external first argument) and tagging these as _verbExtArg_ , the label given to the verbs with an A0 in UPF lexicons. The _verbExtArg_ verbs match the semantic arguments with the deep-syntax actants as follows: A0 = I, A1 = II, A2 = III, A3 = IV, A4 = V, etc. The rest of the verbs were tagged as _verb_ and the correspondence of semantic arguments with deep syntax actants is: A1 = I, A2 = II, A3 = III, A4 = IV, etc.

The UPF lexicons have an attribute-value structure and all the new information transferred from the AnCora lexicons was assigned to new attributes with the anc- prefix (anc_sense, anc_type, anc_lss, anc_diathesis, etc.). Only the data that correspond to already existing attributes in the UPF lexicon were transferred to non-prefixed attributes (lemma, prep).

Although AnCora-Verb describes all the possible diathesis of a verb sense, in this first version of the converted lexicons we have only transferred the default diathesis linguistic information.

The government pattern structure in UPF lexicon will include the PropBank verbal senses, which have been split and transferred into two attributes *pbcls* and *pbid* from AnCora, as well as the syntactic function (anc_fucntion), and the thematic role (anc_theme) assigned to each argument in AnCora, together with the required prepositions. If more than one modifier (argM in AnCora) is possible, then each one of them is identified with a consecutive number.

Finally, the examples have also been transferred to UPF_AnCoraDict_verbs.

**Conversion of AnCora-Nom**

In the conversion of the AnCora-Noun-ES lexicon we have created 3208 entries in the UPF nominal lexicon, one for each lemma, sense, and diathesis.

In the same way as in the verbal lexicon, the linguistic information that has been transferred from AnCora to UPF format has been assigned to attributes prefixed with “anc” (anc_sense, anc_cousin, anc_denotation; anc_alternativelemma, anc_originalVerb, anc_diathesis, anc_plural, anc_lexicalized, anc_lextype, anc_theme).

Similarly to the verbal classification, UPF lexicons distinguish two types of nouns: if the noun has an arg0 then it is typified as _nounExtArg_, otherwise the noun is tagged as _noun_. This is the criteria that we have maintained in the conversion from AnCora to UPF-format.

If one sense has more than one possible diathesis, these will be converted into as many lexical entries as diathesis in UPF-lexicon. The identification of these entries will add a letter at the end.

The Wordnet id disambiguation is also a valuable piece of information that has been transferred, creating as many lines (wnet = “ ”) in UPF_AnCoraDict_nouns as possible according to Wordnet senses that are suggested in AnCora-Nom.

AnCora-Nom also offers a great number of examples for most lexical entries. We have also added these to the UPF-dict-nouns, but we have limited their number to a maximum of 10. AnCora offers valuable information regarding the syntactic function (cn), argument position (arg0, arg1, … , argM), and thematic role of each of the arguments of the noun exemplified. In this first version of UPF_AnCoraDict_nouns, the nominal arguments are identified with square brackets to which we have added the thematic role tag.

Finally, the information about the possible specifiers that the noun may possess has also been transferred into UPF_AnCoraDict_nouns as the value of the attribute “specifier” which has as many instances as many different types of specifiers are possible for the considered noun.

As a result of the conversion, two new lexicons have been generated in the UPF format: UPF_AnCoraDict_verbs.dic (3944 verbal entries) and UPF_AnCoraDict_nouns.dic (3208 nominal entries) with rich semantic information from the AnCora lexicons and can now be used together with UPF ES_lexicon in graph-transduction-based grammars and in other linguistic applications.

Example of a lexical entry in UPF_AnCoraDict_verbs.dic is shown in Figure 15.

```
"beber_VB_01":_verbExtArg_{
        example = "*0* no bebí en las fuentes comunistas"
        example = "Beber una botella de ginebra con Cayetana"
        example = "Otros también bebían"
        example = "*0* no podemos beber el agua"
        anc_sense = "1"
        lemma = "beber"
        anc_type = "verb"
        anc_lss = "A21.transitive-agentive-patient"
        anc_diathesis = "default"
        gp = {
                pb = {
                        "drink_VB_03"= {
                                pbcls = "drink"
                                pbId = "03"
                        }
                }
                 I = {
                        anc_theme = "agt"
                        anc_funct = "suj"
                }
                 II = {
                        anc_theme = "pat"
                        anc_funct = "cd"
                }
                 M1 = {
                        anc_theme = "loc"
                        anc_funct = "cc"
                        prep   = "en"
                }
                 M2 = {
                        anc_theme = "adv"
                        anc_funct = "cc"
                        prep   = "con"
                }
        }
}
```

Figure 15: Lexical entry '*beber*' in UPF_AnCoraDict_verbs.dic

Example of a lexical entry in UPF_AnCoraDict_nouns.dic is shown in Figure 16.

```
"aplazamiento_NN_02":_noun_{
	anc_sense = "2"
	anc_cousin = "no"
	lemma = "aplazamiento"
	anc_denotation = "event"
	anc_lexicalized = "no"
	anc_originalVerb = "aplazar_VB_01"
	wnet = "16:00690488"
	anc_diathesis = "default"
	anc_plural = "no"
	gp = {
		I = {
			anc_theme = "pat"
			prep = "de"
		}
		specifiers = "determiner(article)"
		specifiers = "void"
	}

	examples = {
		example = "El debate generado en Venezuela por el aplazamiento
[de las \" megaelecciones \" previstas para el pasado_domingo]-PAT ha revuelto el
escenario político nacional , con un sólo invitado de piedra : el presidente
Hugo_Chávez . "
		example = "El CNE fue acusado hace unos días por la generalidad de
las fuerzas políticas venezolanas de ser el verdadero y nico causante del
aplazamiento  [de las votaciones]-PAT debido_a la \" incompetencia \" de sus cinco
miembros , que complicaron las ya de_por_sí complejas elecciones múltiples con
continuas rectificaciones en las bases de datos del proceso automatizado . "
		example = "El aplazamiento  [de los comicios]-PAT fue recomendado
además por la misión de observadores electorales de la
Organización_de_Estados_Americanos ( OEA ) , para poder probar el programa
informático de conteo de votos que se usará en la nueva elección . "
	}
}
```

Figure 16: Lexical entry 'aplazamiento' in UPF_AnCoraDict_nouns.dic

## 4.2 Datasets for training statistical linearisers and deep generators

### 4.2.1 A new Universal Dependency-based dataset

Universal Dependencies is a generic framework for cross-lingual syntactico-semantic annotation that has been applied to over 80 languages so far, for a total of over 140 different treebanks. Most treebanks have been obtained through automatic conversions of other treebanks or obtained via automatic annotation. The resulting annotations are known to lack consistency and quality, but they have the advantage to provide a framework that reduces the differences across different languages. In MindSpaces, we further developed the first multilingual dataset for training statistical generators.

The annotated surface structures are syntactic trees with lemmas, part-of-speech tags, morphological and dependency information under the form of grammatical functions such as subject, object, adverbial, etc. We developed a converter for Universal Dependency (UD) structures to obtain parallel “deep” data and thus serve as input for deep generators as described in Section Surface generation. By using the structures at these two levels, we have two different outputs for Natural Language Generation:

- Shallow Track
  - Input: unordered UD dependency trees with lemmatised words that hold PoS tags and morphological information;
  - Task: determine word order and inflect words;
  - Languages: Arabic, Chinese, English (4 datasets), French (3), Hindi, Indonesian, Japanese, Korean (2), Portuguese (2), Russian (2), and Spanish (2).
- Deep track
  - Input: unordered predicate-argument tree with lemmatised content words that hold coarse-grained PoS tags and semantic information;
  - Task: introduce functional words, resolve morphological agreements, determine word order, and inflect words;
  - Languages: English (4 datasets), French (3), Spanish (2).

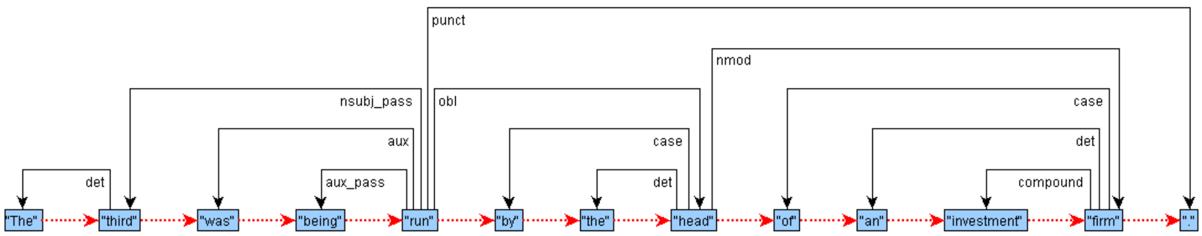
**General specifications of the data**

The shallow track structures are obtained by simply removing the order and surface forms information from the original structures.

The deep structures in this configuration consist of predicate-argument structures obtained through the application of graph-transduction grammars to the UD surface-syntactic structures. The deep and surface structures are aligned node to node. In the deep structures, we aim at removing all the information that is language-specific and oriented towards syntax:

- determiners and auxiliaries are replaced (when needed) by attribute/value pairs, as, e.g., Definiteness, Aspect, and Mood:
  - auxiliaries: was built-> build;
  - determiners: the building-> building;
- functional prepositions and conjunctions that can be inferred from other lexical units or from the syntactic structure are removed:
  - built by X-> built X;
- edge labels are generalised into predicate argument (semantics-oriented) labels in the PropBank/NomBank fashion:
  - subject(built, by X)-> FirstArgument(build, X).

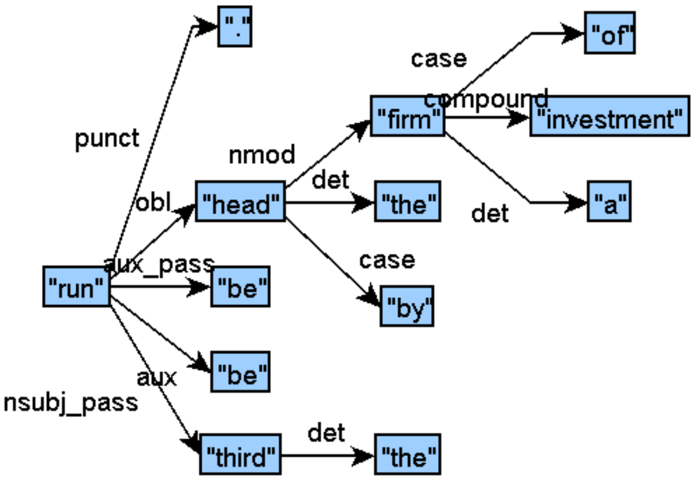
Figures Figure 17, Figure 18, and Figure 19 show original, surface, and deep structures respectively.




```
# sent_id = weblog-juancole.com_juancole_20051126063000_ENG_20051126_063000-0006
# text = The third was being run by the head of an investment firm.
```

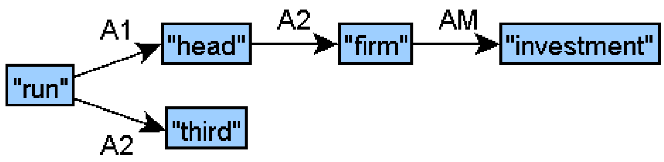
| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | The | the | DET | DT | Definite=Def\|PronType=Art | 2 | det |
| 2 | third | third | ADJ | JJ | Degree=Pos\|NumType=Ord | 5 | nsubj_pass |
| 3 | was | be | AUX | VBD | Mood=Ind\|Number=Sing\|Person=3\|Tense=Past\|VerbForm=Fin | 5 | aux |
| 4 | being | be | AUX | VBG | VerbForm=Ger | 5 | aux_pass |
| 5 | run | run | VERB | VBN | Tense=Past\|VerbForm=Part\|Voice=Pass | 0 | root |
| 6 | by | by | ADP | IN | _ | 8 | case |
| 7 | the | the | DET | DT | Definite=Def\|PronType=Art | 8 | det |
| 8 | head | head | NOUN | NN | Number=Sing | 5 | obl |
| 9 | of | of | ADP | IN | _ | 12 | case |
| 10 | an | a | DET | DT | Definite=Ind\|PronType=Art | 12 | det |
| 11 | investment | investment | NOUN | NN | Number=Sing | 12 | compound |
| 12 | firm | firm | NOUN | NN | Number=Sing | 8 | nmod |
| 13 | . | . | PUNCT | . | _ | 5 | punct |

Figure 17: Original UD structure in the CoNLL-U format (top: graphical representation)




| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | the | _ | DET | DT | Definite=Def\|PronType=Art | 2 | det |
| 2 | third | _ | ADJ | JJ | Degree=Pos\|NumType=Ord | 3 | nsubj_pass |
| 3 | run | _ | VERB | VBN | Tense=Past\|VerbForm=Part\|Voice=Pass | 0 | root |
| 4 | be | _ | AUX | VBD | Mood=Ind\|Number=Sing\|Person=3\|Tense=Past\|VerbForm=Fin | 3 | aux |
| 5 | be | _ | AUX | VBG | VerbForm=Ger | 3 | aux_pass |
| 6 | head | _ | NOUN | NN | Number=Sing | 3 | obl |
| 7 | . | _ | PUNCT | . | _ | 3 | punct |
| 8 | by | _ | ADP | IN | _ | 6 | case |
| 9 | the | _ | DET | DT | Definite=Def\|PronType=Art | 6 | det |
| 10 | firm | _ | NOUN | NN | Number=Sing | 6 | nmod |
| 11 | an | _ | DET | DT | Definite=Ind\|PronType=Art | 10 | det |
| 12 | investment | _ | NOUN | NN | Number=Sing | 10 | compound |
| 13 | of | _ | ADP | IN | _ | 10 | case |

Figure 18: Shallow track input in the CoNLL-U format (top: graphical representation)




| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | third | _ | ADJ | _ | Degree=Pos | 2 | A2 |
| 2 | run | _ | VERB | _ | Tense=Past\|Aspect=Progr | 0 | ROOT |
| 3 | head | _ | NOUN | _ | Number=Sing\|Definiteness=Def | 2 | A1 |
| 4 | firm | _ | NOUN | _ | Number=Sing\|Definiteness=Indef | 3 | A2 |
| 5 | investment | _ | NOUN | _ | Number=Sing | 4 | AM |

Figure 19: Deep track input in the CoNLL-U format (top: graphical representation)

### A converter for obtaining the Deep track structures

We developed graph-transduction grammar to implement the converter. The UD-based converter does not make any use of lexical resources; the predicate-argument relations are derived using syntactic cues only. The deep input is a compromise between (i) correctness and (ii) adequacy in a generation setup. The conversion of the UD structures into predicate-argument structures depends not only on the mapping process, but also on the availability of the information in the original annotation.

Table 10 shows the different labels that the UD-based graph-transduction grammars currently produce.

Table 10: Semantic labels in the output of the UD-based pipeline

| **Semantic label** | **Type** | **Description** | **Example** |
|---|---|---|---|
| A1/A1INV | Core | 1st argument of a predicate | build-> an architect |
| A2/A2INV | Core | 2nd argument of a predicate | build-> a building |
| A3/A3INV | Core | 3rd argument of a predicate | inaugurate-> on March 15 |
| A4, A5, A6 | Core | 4th to 6th arguments | *Very uncommon* |
| AM | Non-Core | None of governor or dependent are argument of the other | build-> next to the museum |
| LIST | Coordinative | List of elements | built-> and-> inaugurated |
| NAME | Lexical | Part of a name | Chrysler->Building |
| DEP | UKN | Undefined dependent | N/A |

The following phenomena should be highlighted:

- **Alignment between surface and deep nodes**

  On the deep nodes, we use one or more feature ids with attributed as suffix the line number of the corresponding surface nodes: on a deep node, id1=4|id2=15 means that this deep node is aligned with the surface nodes on the lines 4 and 15 of the corresponding surface structure. Only elements triggered by other elements (as opposed to be triggered by the structure of the sentence) are aligned with deep nodes. That is, a subcategorised preposition is aligned with a deep node, while a void copula or an expletive subject are not.

- **Core relations**

  Each defined core relation is unique for each predicate: there cannot be two arguments with the same slot for one predicate. If a predicate has an A2 dependent, it cannot have another A2 dependent, and it cannot be A2INV of another predicate.

- **Auxiliaries**

  Auxiliaries are mapped to the universal feature "Aspect".

- **Conjunctions/prepositions**

  The prepositions and conjunctions maintained in the deep representation can be found under a A2INV dependency. A dependency path Gov-AM-> Dep-A2INV-> Prep is equivalent to a predicate (the conjunction/preposition) with 2 arguments: Gov <-A1-Prep-A2-> Dep.

- **Modals**

  They are mapped to the universal feature "Mood".

- **Pronouns**

  - Relative: only subject and object relative pronouns directly linked to the main relative verb are removed from the deep structure.
  - Subject: a dummy pronoun node for subject is added if an originally finite verb has no first argument and no available argument to build a passive; for a pro-drop language such as Spanish, a dummy pronoun is added if the first argument is missing.

- **Punctuations**

  Only the final punctuations are encoded in the deep representations: the main node of a sentence indicates if the latter is declarative, interrogative, exclamative, suspensive, or if it is involved in a parataxis, with the feature "clause_type".

As described for FORGe in the previous Section, our conversion graph-transduction grammars are rules that apply to a subgraph of the input structure and produce a part of the output structure. During the application of the rules, both the input structure (covered by the left side of the rule) and the current state of the output structure at the moment of application of a rule (i.e., the right side of the rule) are available as context. The output structure in one transduction is built incrementally: the rules are all evaluated, the ones that match a part of the input graph are applied, and a first piece of the output graph is built Then the rules are evaluated again, this time with the right-side context as well, and another part of the output graph is built, and so on. The transduction is over when no rule is left that matches the combination of the left-side and the right-side. Consider, for illustration, a sample rule from the SSynt-DSynt mapping in Figure 20. This rule, in which we can see the left-side and the right-side fields, collapses the functional prepositions (?Xl, identified during the pre-processing stage with the BLOCK=YES attribute/value pair) with their dependent (?Yl). That is, a functional preposition such as by in built by Y is removed from the output structure and made to correspond with the right-side node Y (i.e., the dependent). The right-side context is indicated by the prefix rc: before a variable or a correspondence In practice, it means that the rule looks for the rc:-marked elements in the current state of the output structure, and builds the elements that are not rc:-marked, in this case the correspondence between the right-side Y and the left-side by, and the new feature *original_deprel*, which stores the left-side incoming dependency relation. A similar rule would apply to *firm* and *of*, *of* being the dependent in this configuration (see Figure 18). As a result of the application of this rule, only *firm* is left in Figure 19, which has a correspondence with both *firm* and *of* from Figure 18.

```
c:?Xl {                                      rc:?Yr {
  BLOCK = YES                                  rc:<=> ?Yl
  c:deprel = ?dep                              <=> ?Xl
  c:id = ?i1                                   original_deprel = ?dep
  c:?s-> c:?Yl {                             }
    c:id = ?i2
  }
}

(?s == PMOD | ?s == IM | ?s == SUB)
```

Figure 20: A sample graph-transduction rule; ? indicates a variable; ?Xl{} is a node, ?s-> is a relation, a=?b is an attribute/value pair

Table 11 sums up the current state of the graph-transduction grammars and rules for the mapping between surface-syntactic structures and UD-based semantic structures.

Table 11: Graph-transduction rules for UD-based deep parsing (* - includes rules that simply copy node features (~40 per grammar))

| Grammars | 2018 #rules* | 2019 #rules* | Description |
|---|---|---|---|
| Pre-processing | 76 | 112 | Identify nodes to be removed<br>Identify verbal finiteness and tense |

| | | | |
|---|---|---|---|
| SSynt-Sem | 120 | 168 | Remove idiosyncratic nodes<br>Establish correspondences with surface nodes<br>Predict predicate-argument dependency labels<br>Replace determiners, modality, and aspect markers by attribute-value feature structures<br>Identify duplicated core dependency labels below one predicate |
| Post-processing | 60 | 82 | Replace duplicated argument relations by best educated guess<br>Identify remaining duplicated core dependency labels (for posterior debugging) |

### 4.2.2 The second Surface Realisation Shared Task (SR'19)

In 2018 and 2019, UPF organised the first and second multilingual Surface Realisation Shared Task (SR'18 and SR'19). The UD data was processed as detailed in the previous sections, and a subset of the data shown above in this section was provided to the participants. Outputs were evaluated according to the three automatic metrics (BLEU, NIST, and Normalised Inverted Edit Distance – DIST) and two human evaluations (Meaning similarity and Readability). Human-produced outputs were also evaluated to serve (i) as reference score in Readability, for which evaluators were asked to rate from 0 to 100 with a slider the intrinsic quality of sentences, and (ii) as comparison for meaning similarity, for which evaluators were asked to rate from 0 to 100 (with a slider too) if the meanings of two sentences were the same or not, one being a system output and the other one being the human reference.

For SR'19 (Mille et al., 2019a), 33 international teams (from 17 countries) registered to the task. 14 of these teams submitted outputs, two of which withdrew their submissions at the last minute. New languages and features were introduced. Table 12 shows the results of the 12 teams in the Shallow Track according to the BLEU metric. 4 teams addressed all 29 datasets (11 languages), and 4 other teams addressed three and 9 languages. In Table 13, the scores for the Deep Track are presented. The increase in participation compared to SR'18 is clearly visible: 3 teams addressed the Deep Track, two of which for all datasets and languages. Finally, Table 14 shows the results of the human evaluations on English and Spanish (for which system outputs originating from gold-standard and silver-standard data were evaluated). One can notice that the gaps seen in the SR'18 evaluations between human texts and system outputs are closing: 0.78z in English and 0.65z in Spanish. A notable gap between human assessment (higher) and metric assessment (lower) of deep track systems can be observed, in particular for the best deep track systems. The biggest progress has been made in SR'19 for Deep track systems: not only we had multiple Deep Track systems to evaluate (compared to just one in 2018), but the best Deep Track system performed equally well or better than most Shallow Track systems for both Readability and Meaning similarity. Another notable development has been the introduction of silver-standard data. Even though the quality of the texts obtained when generating from automatically parsed data is lower than when using gold-standard data, the high scores according to human evaluations suggest that the shallow inputs could be used as pivot representations in text-to-text systems for paraphrasing, simplification or summarisation applications.

Table 12: SR'19 Shallow Track BLEU scores of the participating teams

| –T1-BLEU– | ADA | BME | CLa | CMU | Dep | Dip | IMS | LOR | RAL | OSU | Til |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ar_padt | | 26.4 | | | 23.01 | | **64.9** | 16.71 | | | 21.12 |
| en_ewt | 79.69 | 59.22 | 22.08 | 77.47 | 60.51 | 43.5 | **82.98** | 60.37 | 41.23 | 62.38 | 59.57 |
| en_gum | 81.39 | 57.57 | 15.32 | 82.39 | 66.06 | 44.24 | **83.84** | 60.7 | 46.68 | 49.91 | 59.39 |
| en_lines | 41.62 | 48.78 | 15.3 | 75.49 | 59.81 | 32.42 | **81** | 58.82 | 41.28 | 54.56 | 57.02 |
| en_partut | 51 | 61.37 | 10.07 | 78.98 | 62.68 | 35.11 | **87.25** | 53.64 | 48.43 | 7.37 | 64.87 |
| es_ancora | | 61.09 | | 76.47 | 59.29 | | **83.7** | 43.02 | | | 59.29 |
| es_gsd | | 53.74 | | 70.15 | 57.14 | | **82.98** | 53.16 | | | 54.48 |
| fr_gsd | | 43.8 | | 60.15 | 44.91 | 27.04 | **84** | 54.6 | | | 52.1 |
| fr_partut | | 49.17 | | 63.7 | 55.05 | 37.69 | **83.38** | 54.14 | | | 66.01 |
| fr_sequoia | | 46.72 | | 62.79 | 46.87 | 28.95 | **85.01** | 53.71 | | | 57.41 |
| hi_hdtb | | 63.63 | | | 64.07 | | **80.56** | 26.51 | | | 60.72 |
| id_gsd | | 54.22 | | | 63.71 | | **85.34** | 46.27 | | | 53.03 |
| ja_gsd | | 49.53 | | 63.59 | 50.19 | | **87.69** | 38.8 | | | 43.02 |
| ko_gsd | | 46.08 | | | 41.81 | | **74.19** | 37.85 | | | 2.14 |
| ko_kaist | | 47.23 | | | | | **73.93** | 39.75 | | | 1.39 |
| pt_bosque | | 39.53 | | | 39.82 | | **77.75** | 52.69 | | | 51.18 |
| pt_gsd | | 30.39 | | | 27.16 | | **75.93** | 33.45 | | | 40.48 |
| ru_gsd | | 54.58 | | | 32.04 | | **71.23** | 55.09 | | | 6.84 |
| ru_syntagrus | | 50.91 | | | | | **76.95** | 59.99 | | | 30.51 |
| zh_gsd | | 58.72 | | 68.54 | 59.64 | 32.87 | **83.85** | 48.21 | | | 53 |
| en_pud | 84.07 | 60.42 | 12.36 | 80.35 | | 45.61 | **86.61** | 61.43 | 46.84 | 67.91 | 63.29 |
| ja_pud | | 53.65 | | 66.52 | | | **86.64** | 41.72 | | | 44.37 |
| ru_pud | | 10.15 | | | | | **58.38** | 52.37 | | | 16.35 |
| en_ewt$_{HIT}$ | 77.21 | 58.07 | 21.21 | 76.6 | | 43.23 | **81.8** | 58.5 | 39.77 | 60.58 | 59.08 |
| en_pud$_{LAT}$ | 80.66 | 53.46 | 12.89 | 76.22 | | 44.06 | **82.6** | 55.4 | 41.5 | 66.18 | 57.92 |
| es_ancora$_{HIT}$ | | 61.26 | | 77.28 | | | **83.31** | 43.2 | | | 59.58 |
| hi_hdtb$_{HIT}$ | | 64.27 | | | | | **80.19** | 26.99 | | | 61.54 |
| ko_kaist$_{HIT}$ | | 46.72 | | | | | **74.27** | 41.83 | | | 1.73 |
| pt_bosque$_{STA}$ | | 40.42 | | | | | **78.97** | 53.64 | | | 52.79 |

Table 13: SR'19 Deep Track scores of the participating teams

| –T2– | BLEU | | | NIST | | | DIST | | |
|---|---|---|---|---|---|---|---|---|---|
| | IMS | RAL | Sur | IMS | RAL | Sur | IMS | RAL | Sur |
| en_ewt | **54.75** | 26.28 | 23.35 | **11.79** | 9.42 | 7.29 | **76.3** | 55.08 | 56.88 |
| en_gum | **52.45** | 26.17 | 17.97 | **11.04** | 9.14 | 5.88 | **73.07** | 51.64 | 49.45 |
| en_lines | **47.29** | 24.94 | 20.96 | **10.63** | 8.79 | 6.35 | **71.93** | 51.2 | 52.49 |
| en_partut | **45.89** | 23.82 | 17.19 | **9.03** | 7.67 | 4.66 | **67.45** | 48.88 | 47.2 |
| es_ancora | **53.13** | | 18.59 | **12.38** | | 5.66 | **68.58** | | 47.19 |
| es_gsd | **51.17** | | 18.69 | **10.82** | | 5.53 | **68.85** | | 48.06 |
| fr_gsd | **53.62** | | 15.83 | **10.79** | | 4.53 | **68.82** | | 47.93 |
| fr_partut | **46.95** | | 14.06 | **8.27** | | 3.61 | **68.99** | | 46.55 |
| fr_sequoia | **57.41** | | 18.52 | **11** | | 4.8 | **72.06** | | 50.94 |
| en_pud | **51.01** | 26.39 | 18.11 | **11.45** | 9.63 | 6.18 | **72.31** | 49.91 | 49.88 |
| en_ewt$_{HIT}$ | **53.54** | 24.54 | 22.42 | **11.55** | 9.19 | 6.9 | **74.99** | 52.54 | 54.86 |
| en_pud$_{LAT}$ | **47.6** | 24.18 | 17.3 | **11.08** | 9.21 | 6.16 | **71.65** | 50.14 | 50.17 |
| es_ancora$_{HIT}$ | **53.54** | | 21.1 | **12.36** | | 5.98 | **70.02** | | 48.57 |

Table 14: SR'19 human evaluations: Meaning Similarity (left) and Readability (right)

| Rank | Ave. | Ave. z | English n | N | System |
|---|---|---|---|---|---|
| 1 | 86.6 | 0.507 | 695 | 810 | ADAPT-T1 |
| | 85.6 | 0.503 | 672 | 768 | IMS-T1 |
| 3 | 82.5 | 0.407 | 702 | 812 | CMU-T1 |
| 4 | 80.6 | 0.324 | 718 | 826 | IMS-T2 |
| | 79.7 | 0.289 | 711 | 816 | TILBURG-T1 |
| | 79.3 | 0.276 | 753 | 859 | DEPDIST-T1 |
| | 78.4 | 0.255 | 720 | 836 | OSU-FB-T1 |
| | 77.0 | 0.222 | 702 | 816 | LORIA-T1 |
| | 73.5 | 0.164 | 695 | 796 | BME-UW-T1 |
| 10 | 72.9 | 0.110 | 680 | 795 | RALI-T1 |
| | 69.5 | −0.006 | 700 | 811 | DIPINFOUNITO-T1 |
| | 67.0 | −0.040 | 692 | 789 | SURFERS-T2 |
| | 68.3 | −0.052 | 707 | 808 | RALI-T2 |
| 14 | 60.9 | −0.216 | 752 | 885 | CLAC-T1 |
| 15 | 55.3 | −0.390 | 674 | 775 | LB-BASELINE-T1 |
| | 53.0 | −0.422 | 733 | 853 | LB-BASELINE-T2 |

| Rank | Ave. | Ave. z | English n | N | System |
|---|---|---|---|---|---|
| – | 71.1 | 0.585 | 824 | 1,281 | HUMAN |
| 1 | 67.9 | 0.507 | 477 | 564 | IMS-T1 |
| | 68.2 | 0.502 | 482 | 573 | ADAPT-T1 |
| 3 | 61.9 | 0.313 | 512 | 582 | IMS-T2 |
| | 62.5 | 0.285 | 500 | 575 | LORIA-T1 |
| | 62.4 | 0.260 | 506 | 589 | CMU-T1 |
| | 60.8 | 0.257 | 497 | 572 | SURFERS-T2 |
| | 60.5 | 0.211 | 516 | 591 | DEPDIST-T1 |
| | 59.2 | 0.160 | 516 | 594 | TILBURG-T1 |
| | 58.3 | 0.156 | 488 | 554 | BME-UW-T1 |
| | 57.4 | 0.121 | 507 | 583 | OSU-FB-T1 |
| | 57.5 | 0.096 | 497 | 569 | RALI-T1 |
| 12 | 50.3 | −0.117 | 494 | 549 | RALI-T2 |
| | 49.6 | −0.195 | 515 | 598 | DIPINFOUNITO-T1 |
| | 48.1 | −0.202 | 524 | 610 | CLAC-T1 |
| 15 | 37.8 | −0.594 | 492 | 569 | LB-Baseline-T2 |
| | 36.5 | −0.677 | 468 | 534 | LB-Baseline-T1 |

| Rank | Ave. | Ave. z | UD Spanish n | N | System |
|---|---|---|---|---|---|
| 1 | 81.1 | 0.378 | 620 | 716 | IMS |
| 2 | 75.8 | 0.168 | 655 | 753 | CMU |
| 3 | 72.2 | 0.006 | 614 | 708 | TILBURG |
| 4 | 70.6 | −0.080 | 617 | 704 | DEPDIST |
| | 69.1 | −0.111 | 623 | 705 | BME-UW |
| 6 | 63.2 | −0.302 | 625 | 706 | LORIA |

| Rank | Ave. | Ave. z | UD Spanish n | N | System |
|---|---|---|---|---|---|
| – | 89.0 | 0.582 | 389 | 438 | HUMAN |
| 1 | 86.5 | 0.517 | 511 | 584 | IMS |
| 2 | 78.9 | 0.236 | 523 | 601 | CMU |
| 3 | 72.1 | −0.009 | 513 | 596 | BME-UW |
| | 71.5 | −0.037 | 498 | 562 | TILBURG |
| 5 | 67.7 | −0.181 | 498 | 562 | DEPDIST |
| 6 | 60.6 | −0.458 | 506 | 577 | LORIA |

| Rank | Ave. | Ave. z | Pred. Spanish n | N | System |
|---|---|---|---|---|---|
| 1 | 82.7 | 0.394 | 686 | 799 | IMS |
| 2 | 78.4 | 0.272 | 683 | 804 | CMU |
| 3 | 70.3 | −0.042 | 688 | 803 | TILBURG |
| | 67.8 | −0.105 | 675 | 789 | BME-UW |
| 5 | 59.2 | −0.422 | 652 | 754 | LORIA |

| Rank | Ave. | Ave. z | Pred. Spanish n | N | System |
|---|---|---|---|---|---|
| – | 89.2 | 0.736 | 405 | 442 | HUMAN |
| 1 | 82.8 | 0.519 | 613 | 713 | IMS |
| 2 | 74.7 | 0.147 | 609 | 686 | CMU |
| 3 | 66.0 | −0.103 | 642 | 737 | TILBURG |
| | 64.7 | −0.169 | 640 | 734 | BME-UW |
| 5 | 53.8 | −0.531 | 594 | 670 | LORIA |

The participation level for such a task is rather high, and due to the success of these first two tasks, a third edition will take place in 2020.

## 4.3 First version of a neural lineariser

The IMS system (Yu, et al. 2019) obtained the best scores for all metrics on almost all datasets at SR'19: the only lower scores were the NIST score on RussianPUD, and the DIST score on EnglishGUM. IMS provided results for all 42 datasets and achieved high macro-average scores on both Shallow track (T1) and Deep track (T2) datasets, with 79.97 BLEU for T1, 51.41 BLEU for T2, 12.79 NIST for T1, 10.94 NIST for T2, 81.62 DIST for T1, and 71.16 DIST for T2. Due to the high quality of the generated outputs and the high adaptability of the system to new domains, it was taken as a core technique in MindSpaces within the artistic solution for PUC3 that consists in creating a personalised narrative based on arbitrary textual materials provided by a user.

IMS uses a pipeline approach consisting of linearisation, completion, inflection, and contraction. All internal models use the same bidirectional Tree- Long Short-Term Memory (Tree-LSTM) encoder architecture (Zhou, Liu, and Pan 2016). The linearisation model orders each subtree separately with beam search and then combines them into a full projective tree. The completion model generates absent function words in a sequential way given the linearised tree of content words; the inflection model predicts a sequence of edit operations to convert the lemma to word form character by character. The contraction model predicts begin-inside-outside tags to group the words to be contracted, and then generates the contracted word form of each group with a sequence-to-sequence model.

For the preliminary experiments, MindSpaces Web Crawler collected messages from Twitter channels of two favorite newspapers of one of the end-users participating in PUC3. The resulting collection contains 3236 and 3223 texts for each channel correspondingly. The messages mainly include titles and abstracts of news in French. They were processed by Textual Analysis component, and obtained parse trees were used for training IMS to generate coherent and grammatically correct texts in news domain.

Trained IMS was applied to unseen structures. For an example of the output, consider the following generated text:

> *Ce matin vers 9 h 30, un salarié qui travaille à La Défense, a alerté les forces de l'ordre qui interviennent désormais le doute pour lever. Lignes U, 1 et RER A ne marquent plus l'arrêt sur ce secteur.*

The generated texts were manually inspected and some suboptimal orders were detected in some cases. This is considered as a minor error to be revised at the course of the project. With this result, the neural lineariser is ready to be ported into the personalised narrative generation pipeline.

# 5 CONCLUSIONS

This document presents the progress attained in the first half of the project with regard to the task of multilingual text generation in WP5. The state of affairs at the beginning of the project is compared to the recent achievements that show significant improvements in all defined subtasks which were done in a timely manner in correspondence with the project timeline.

Five main improvements were made to the UPF FORGe multilingual discourse generator:

- the general coverage of the rules was improved: all grammars are now more complete, with 1,892 rules in total, as opposed to 1,555 at the beginning of the project;
- rules were made more language independent: now, 74% of the rules are language-independent, as opposed to 70% at the beginning of the project;
- updated rules caused the significant improvement of the quality of the English generator on a challenging dataset (about 40% of error reduction on a reference dataset);
- the coverage of the Spanish grammars has been significantly increased, and is now almost the same as for English;
- basic support has been added for Greek and Catalan, together with basic lexicons.

In addition, a new dataset for training statistical generation tools have been developed and validated through the organisation of the second Surface Realisation shared task which attracted 14 teams.

The MindSpaces use cases are currently covered in accordance with the progress in the definition of design parameters and their respective values that is more elaborated for the indoors environment (i.e., PUC2 and PUC3) while such parameters as "*origin of the light*", "*colour/height of the ceiling*", "*shape of the room*" are of limited applicability for PUC1. This reduces the variety in possible outcomes for outdoors environment that is reflected in Table 8.

In the scope of future work, we will aim in obtaining a similar coverage for Greek, Catalan, and French as for English and Spanish. The English and Spanish lexicon will be extended in case additional spatial design parameters are considered, especially with regard to PUC1. Statistical modules will be further trained using the input structures to be obtained by specific manipulations with parse trees in order to meet additional requirements based on the users' feedback.

Six publications in the framework of MindSpaces were produced:

- Mille, Simon, Stamatia Dasiopoulou, and Leo Wanner. (2019). A portable grammar-based NLG system for verbalization of structured data. In *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing (ACM, 2019)*, pp. 1054-1056.

- Shvets, A. Improving scientific article visibility by neural title simplification. In *Proceedings of the 8th International Workshop on Bibliometric-enhanced Information Retrieval (BIR 2019), the 41th European Conference on Information Retrieval (ECIR 2019)*, Cologne, Germany, pp. 140-147.
- Mille, S. (2019) Selected Challenges in Grammar-based Text Generation from the Semantic Web. In *G. S. Osipov et al. (Eds.): Artificial Intelligence: 5th RAAI Summer School, Dolgoprudny, Russia, July 4-7, 2019, Tutorial Lectures*, LNCS 11866 (Springer International Publishing), pp. 85-95.
- Mille, S., A. Belz, B. Bohnet, Y. Graham, L. Wanner. (2019). The Second Multilingual Surface Realisation Shared Task (SR'19): Overview and Evaluation Results. In *Proceedings of the 2st Workshop on Multilingual Surface Realisation (MSR), 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Hong Kong, China, pp. 1-17.
- Sobrevilla, M.A., S. Mille, T. Pardo. (2019). Back-Translation as Strategy to Tackle the Lack of Corpus in Natural Language Generation from Semantic Representations. In *Proceedings of the 2nd Workshop on Multilingual Surface Realisation (MSR), 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Hong Kong, China, pp. 94-103.
- Mille, S., S. Dasiopoulou, B. Fisas, L. Wanner. (2019). Teaching FORGe to Verbalize DBpedia Properties in Spanish. *In Proceedings of the 12th International Conference on Natural Language Generation*, Tokyo, Japan, pp. 473-483.

# 6 REFERENCES

Aparicio, J., M. Taulé, Martí M.A., "AnCora-Verb: A Lexical Resource for the Semantic Annotation of Corpora", Proceedings of 6th International Conference on Language Resources and Evaluation, 2008, Marrakesh, Morocco.

Baker, Collin F., Charles J. Fillmore, and John B. Lowe. "The berkeley framenet project", In Proceedings of the 17th international conference on Computational linguistics-Volume 1, pp. 86-90. Association for Computational Linguistics, 1998.

Ballesteros, M., B. Bohnet, S. Mille, and L. Wanner. "Data-driven sentence generation with non-isomorphic trees", *Proceedings of NAACL-HLT.* ACL, 2015. 387-397.

Banerjee, Satanjeev, and Alon Lavie. "METEOR: An automatic metric for MT evaluation with improved correlation with human judgments", In Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization, pp. 65-72. 2005.

Bohnet, B., and L. Wanner. "Open Source Graph Transducer Interpreter and Grammar Development Environment", Proceedings of LREC. Valletta, 2010.

Bouayad-Agha, N., G. Casamayor, S. Mille, and L. Wanner. "Perspective-oriented generation of football match summaries: Old tasks, new challenges", *TSLP* 9, no. 2 (2012): 1-31.

Fellbaum, Christiane. 2005. "WordNet and wordnets", In: Brown, Keith et al. (eds.), Encyclopedia of Language and Linguistics, Second Edition, Oxford: Elsevier, 665-670.

Fotopoulou A, Markantonatou S, Giouli V. "Encoding MWEs in a conceptual lexicon", Proceedings of the 10th Workshop on Multiword Expressions (MWE), 43-47, 2014.

Miguel, J.M.G., Vaamonde, G. and Domínguez, F.G. "ADESSE. A Database with Syntactic and Semantic Annotation of a Corpus of Spanish", Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC), Valletta (Malta), 2010.

Gardent, C., Shimorina, A., Narayan, S. and Perez-Beltrachini, L. "Creating training corpora for nlg micro-planning", *Proceedings of ACL.* ACL, 2017a. 179-188.

Gardent, C., Shimorina, A. Narayan, S. and Perez-Beltrachini L. "The WebNLG challenge: Generating text from RDF data", *Proceedings of the 10th International Conference on Natural Language Generation.* 2017b. 124-133.

Gatt, A., and E. Krahmer. "Survey of the State of the Art in Natural Language Generation: Core tasks, applications and evaluation", *Journal of Artificial Intelligence Research, 61*, 2018: 65-170.

Giouli,V.; Fotopoulou, A.; “Emotion verbs in Greek. From Lexicon-Grammar tables to multipurpose syntactic and semantic lexica”, Proceedings of the 15th EURALEX International Congress (EURALEX 2012), Aug 2012, Oslo, Norway.

Goutsos,D., “The Corpus of Greek texts: a reference corpus for modern Greek”, Corpora, Volume 5 Issue 1, page 29-44, May 2010.

Gurevych, I., Eckle-Kohler, J., Hartmann, S., Matuschek, M., Meyer, Ch.M., Wirth, Ch. “UBY – A Large-Scale Unified Lexical-Semantic Resource Based on LMF”, In: Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL), pp. 580–590, 2012. Avignon, France.

Hamp, B. and H. Feldweg. “GermaNet – a Lexical-Semantic Net for German”, in: Proceedings of the ACL workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications, pp. 9–15, 1997. Madrid, Spain

Hartrumpf, Sven, Hermann Helbig, and Rainer Osswald. “The Semantically Based Computer Lexicon HaGenLex - Structure and Technological Environment”, In: Traitement automatique des langues, 44(2), 2003, pp. 81-105.

Kingsbury, P. and M. Palmer, “From TreeBank to PropBank” in Proceedings of LREC, Las Palmas, 2002.

Mel’čuk I. “Dependency syntax”, State University of New York Press, Albany, NY. 1988

Mille, S., A.Burga, and L. Wanner (2013). “AnCora-UPF: A Multi-Level Annotation of Spanish”, In Proceedings of the 2nd International Conference on Dependency Linguistics (DepLing).

Mille, S., Wanner, L. (2015). “Towards large-coverage detailed lexical resources for data-to-text generation”, In Proceedings of the first workshop on data to text generation, Edinburgh, Scotland, UK.

Mille, S., R. Carlini, A. Burga, and L. Wanner. “FORGe at SemEval-2017 Task 9: Deep sentence generation based on a sequence of graph transducers.”, *Proceedings of the 11th International Workshop on Semantic Evaluation.* Vancouver, 2017. 920-923.

Mille, S., A. Belz, B. Bohnet, Y. Graham, L. Wanner. (2019a). “The Second Multilingual Surface Realisation Shared Task (SR'19): Overview and Evaluation Results”, In *Proceedings of the 2st Workshop on Multilingual Surface Realisation (MSR), 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp.1-17, Hong Kong, China.

Mille, S., S. Dasiopoulou, B. Fisas, L. Wanner. (2019b). “Teaching FORGe to Verbalize DBpedia Properties in Spanish”, *In Proceedings of the 12th International Conference on Natural Language Generation*, Tokyo, Japan, pp.473-483.

Meyers, A. R. R., C. Macleod, R. Szekely, V. Zelinska, B. Young and R. Grishman. "The NomBank project: An interim report", in Proceedings of HLT-NAACL 2004 workshop: Frontiers in corpus annotation, Boston, 2004.

Navigli R. and S. Ponzetto. "BabelNet: Building a very large multilingual semantic network", in Proceedings of the 48th annual meeting of the association for computational linguistics, Uppsala, 2010.

Nayak, N., D. Hakkani-Tur, M. Walker, and L. Heck. "To plan or not to plan? discourse planning in slot-value informed sequence to sequence models for language generation", *Proceedings of Interspeech.* 2017.

Papageorgiou, H., Desipri, E., Koutsombogera, M., Pouli, K. and Prokopidis, P. "Adding multi-layer semantics to the Greek Dependency Treebank", LREC 2006, Genoa, Italy.

Papineni, K., Roukos, S., Ward, T., and Zhu, W. (2002). "Bleu: a Method for Automatic Evaluation of Machine Translation", *Proceedings of the 2002 Conference of the Association of Computational Linguistics (ACL)*, p. 311-318.

Perez-Beltrachini, Laura, Rania Sayed, and Claire Gardent. "Building rdf content for data-to-text generation", In Proceedings of The 26th International Conference on Computational Linguistics (COLING 2016), Osaka, Japan. 2016.

Perez-Beltrachini, L., and C. Gardent. "Learning embeddings to lexicalise rdf properties", *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics.* 2016. 219-228.

Peris, A.; Taulé,M. "AnCora-Nom: A Spanish lexicon of deverbal nominalizations", Procesamiento del Lenguaje Natural, Vol. 46:11-18, 2011.

Petasis, G., Karkaletsis, V., Farmakiotou, D., Androutsopoulos, I. and Spyropoulo, C.D. "A Greek morphological lexicon and its exploitation by natural language processing applications", Panhellenic Conference on Informatics, 2001, pp. 401-419.

Rambow, O., and T. Korelsky. "Applied text generation", *Proceedings of the third conference on Applied natural language processing.* Association for Computational Linguistics, 1992. 40-47.

Riemsdijk, H. ed., 2011. "Clitics in the Languages of Europe", Vol. 20, No. 5. Walter de Gruyter.

Roller, Stephen, Sabine Schulte im Walde (2013). "A Multimodal LDA Model integrating Textual, Cognitive and Visual Modalities", In: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). Seattle, WA.

Scheible, Silke, Sabine Schulte im Walde, Marion Weller, Max Kisselew. "A Compact but Linguistically Detailed Database for German Verb Subcategorisation relying on Dependency Parses from a Web Corpus", Tool, Guidelines and Ressouce. In: Proceedings of the 8th Web as Corpus Workshop. Lancaster, UK, July 2013.

Schuler, K. "VerbNet: A broad-coverage, comprehensive verb lexicon", Univeristy of Pennsylvania, 2005.

Snover, Matthew G., Nitin Madnani, Bonnie Dorr, and Richard Schwartz. "TER-Plus: paraphrase, semantic, and alignment enhancements to Translation Edit Rate", Machine Translation 23, no. 2-3 (2009): 117-127.

Speer, Robyn, Joshua Chin, and Catherine Havasi. 2017. "ConceptNet 5.5: An Open Multilingual Graph of General Knowledge", In proceedings of AAAI 31.

Taulé, M., Borrega, O. and Martí, M.A. "AnCora-Net: Integración multilingüe de recursos lingüísticos semánticos", Procesamiento del Lenguaje Natural, Vol. 47:153-160, 2011.

Tzortzi, K., Markantonatou, S., "Development of a Conceptual Lexicon with ontological techniques", Proceedings of Terminology and Ontology: theories and applications TOth 2014, Chambéry, France.

Walter, S., C. Unger, and P. Cimiano. "M-ATOLL: A Framework for the Lexicalization of Ontologies in Multiple Languages", *Semantic Web Conference.* Riva, 2014.

Wanner, L., B. Bohnet, N. Bouayad-Agha, F. Lareau, and D. Nicklaß. "MARQUIS: Generation of user-tailored multilingual air quality bulletins", *Applied Artificial Intelligence* 24, no. 10 (2010): 914-952.

Wanner, L., Rospocher, M., Vrochidis, S., Johansson, L., Bouayad-Agha, N., Casamayor, G., Karppinen, A., Kompatsiaris, I., Mille, S., Moumtzidou, A. and Serafini, L. "Ontology-centered environmental information delivery for personalized decision support", *Expert Systems with Applications* 42, no. 12 (2015): 5032-5046.

Yu, X., Falenska, A., Haid, M., Vu, N. T., and Kuhn, J. (2019). "IMSurReal: IMS at the Surface Realization Shared Task 2019", In Proceedings of the 2nd Workshop on Multilingual Surface Realisation (MSR 2019), pp. 50-58.

Zhou, Y., Liu, C., and Pan, Y. (2016). "Modelling Sentence Pairs with Tree-structured Attentive Encoder. In Proceedings of COLING 2016" the 26th International Conference on Computational Linguistics: Technical Papers, pp. 2912-2922.