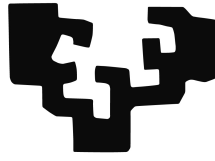


eman ta zabal zazu



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

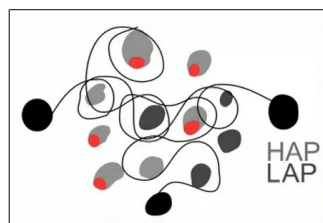
University of the Basque Country

Department of Computer Systems and Languages

A Study Towards Spanish Abstract Meaning Representation

Author: Noelia Migueles-Abraira

Advisors: Rodrigo Agerri and Arantza Diaz de Ilarraza



A thesis submitted in fulfillment of the requirements for the degree of
Master of Science (MSc) in Language Analysis and Processing

Donostia-San Sebastián, June 2017

DECLARATION

I hereby confirm that the thesis presented here is my own work, except for any Internet sources, published or unpublished works which are clearly indicated and acknowledged as such.

Donostia-San Sebastián, June 2017

Noelia Migueles-Abaira

ABSTRACT

Taking into account the increasing attention that researchers of Natural Language Understanding (NLU) and Natural Language Generation (NLG) are paying to Computational Semantics, we analyze the feasibility of annotating Spanish Abstract Meaning Representations. The Abstract Meaning Representation (AMR) project aims to create a large-scale sembank of simple structures that represent unified, complete semantic information contained in English sentences. Although AMR is not destined to be an interlingua, one of its key features is the ability to focus on events rather than on word forms. They do this, for instance, by abstracting away from morpho-syntactic idiosyncrasies. In this thesis, we investigate the requirements to – and we come up with a proposal to – annotate Spanish AMRs, based on the premise that many of these idiosyncrasies mark differences between languages. To our knowledge, this is the first work towards the development of Abstract Meaning Representation for Spanish.

Keywords: AMR, sembank, Spanish, annotation, semantics, NLP

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deep gratitude to my supervisors – Dr. Agerri and Dr. Diaz de Ilarraza – for their valuable and constructive suggestions during the planning and development of this research project. Their professionalism, attention to detail, and willingness to give their time so generously were greatly appreciated.

I would also like to offer my sincerest thanks to all my professors at UPV/EHU and to the IXA Research Group, especially to Gorka Labaka and Uxoia Iñurrieta for their technical and/or theoretical support during the annotation phase.

And, last but not least, I owe my special thanks to Martha Palmer and Julia Bonn, who provided help and advice during the entire development of this thesis.

CONTENTS

1 Introduction	9
1.1 Motivation and Goal.....	9
1.2 Thesis Outline.....	10
2 Background	11
2.1 Computational Semantics.....	11
2.2 Abstract Meaning Representation.....	13
2.2.1 How does AMR Treat Different Linguistic Phenomena?	15
2.2.2 The AMR Editor and Smatch.....	19
3 Related Work	20
3.1 The Treebank Semantics System.....	20
3.2 The Groningen Meaning Bank.....	21
3.3 Universal Conceptual Cognitive Annotation.....	23
3.4 Other Approaches and Related Research.....	25
4 Methodology	27
4.1 Data Preparation.....	28
4.2 Annotation Procedure.....	29
4.3 Data Analysis.....	31
5 Evaluation	34
5.1 Phenomena and Extended Guidelines.....	34
5.1.1 NP Ellipses.....	37
5.1.2 Third Person Possessive Pronouns.....	38
5.1.3 Third Person Clitic Pronouns.....	39
5.1.4 <i>Se</i> Usage.....	41
5.1.5 Gender.....	44
5.1.6 Verbal Periphrases and Locutions.....	47
5.1.7 Double Negatives.....	49
5.2 Quantitative Results.....	50
5.2.1 Results.....	51
5.2.2 Conclusions.....	53

6 Discussion	57
6.1 Research Questions	57
6.2 Comparison with Other Works	58
6.3 Limitations and Open Issues	60
7 Conclusions and Future Work	62
7.1 Summary and Contributions	62
7.2 Future Work	63
7.3 Closing Remarks	63
References	65
List of Abbreviations	6
List of Tables and Figures	8
Appendix 1: Annotations	70
Appendix 2: Conversions	94

List of Abbreviations

AMR	Abstract Meaning Representation
BLT	Basic Linguistic Theory
BOW	Bits of Wisdom
CL	Computational Linguistics
DAG	Directed Acyclic Graph
DAS	Degree of AMR Similarity
DMRS	Dependency Minimal Recursion Semantics
DRS	Discourse Representation Structure
DRT	Discourse Representation Theory
DSS	Degree of Syntax Similarity
ED	Event Detection
FOL	First-Order Logic
GMB	Groningen Meaning Bank
HRG	Hyperedge Replacement Grammar
IAA	Inter-Annotator Agreement
LDC	Linguistic Data Consortium
MT	Machine Translation
NE	Named Entity
NED	Named Entity Disambiguation
NLG	Natural Language Generation
NLP	Natural Language Processing
NLU	Natural Language Understanding
NP	Noun Phrase
PB	PropBank
POS	Part of Speech
PSS	Partial Structural Similarity
PTB	Penn Treebank
SCT	Scope Control Theory
SMT	Statistical Machine Translation

SNLU	Statistical Natural Language Understanding
SRL	Semantic Role Labelling
SSS	Substantial Structural Similarity
TSD	Total Structural Difference
UCCA	Universal Conceptual Cognitive Annotation
WSD	Word Sense Disambiguation

List of Tables and Figures

Tables

- 1 Text statistics
- 2 *Se* usage with examples according to Lozano (2005)
- 3 Rules for gender annotation of Spanish entities
- 4 Degrees of AMR Similarity (expressed in percentages)
- 5 Degrees of Syntax Similarity (expressed in percentages)

Figures

- 1 Levels of linguistic analysis according to Thomas and Cook (2005)
- 2 Pipeline view of an NLP system according to Bates (1995)
- 3 Equivalent formats for representing the meaning of “The boy wants to go” according to Banarescu et al. (2013)
- 4 An AMR, with its different key elements
- 5 Meaning representation output of the Treebank Semantics System for the sentence “Mary has been meaning to go for a week”
- 6 Example of the game “Names” of Wordrobe
- 7 DRS for sentence “No man loves Mary”
- 8 Example of an UCCA annotation for the sentence “Big dogs love bones”
- 9 Methodology stages
- 10 Step-by-step procedure on how to annotate verbal locutions
- 11 Correlation between DAS and DSS (expressed in percentages)

1 INTRODUCTION

1.1 Motivation and Goal

“Do you wish me a good morning,
or mean that it is a good morning whether I want it or not;
or that you feel good this morning;
or that it is a morning to be good on?”
– J.R.R. Tolkien

The massive use of data for supervised and unsupervised learning is essential in the fields of Computational Linguistics (CL) and Natural Language Processing (NLP). Undoubtedly, the construction of the first large-scale syntactic treebank (Marcus et al., 1993) in the early 1990s marked a milestone in such fields. Since then, the research and technical development of parsed corpora has become one of the primary driving forces behind the development of new NLP tools as well as a means to perform linguistic research.

Many NLP challenges involve enabling computers to derive meaning from natural language input and others involve generating natural language. But, unfortunately, syntactic parsers are ill-suited for producing meaning representations. After all, human communication is subjective and ambiguous by nature, and it is of no surprise that the analysis that these parsers provide is not enough to capture meaning.

The lack of a unified sembank of natural language sentences paired with their sentential, logical meanings is what led to the appearance of Abstract Meaning Representation (AMR) – a semantic representation language introduced by Banarescu et al. (2013). This approach promotes the representation of the logical meaning of sentences as single rooted, directed graphs – or AMRs – with labeled nodes (concepts) and edges between them (relations). These incorporate semantic roles, among other linguistic phenomena. In a propositional-style logic, Abstract Meaning Representation (AMR) is able to capture who is doing what to whom in a sentence. But there is a problem with this approach: it is exclusively designed to annotate English sentences.

To overcome this fundamental limitation, we decided to jump on the bandwagon and explore the possibility of annotating Spanish AMRs. Being one of the most spoken languages in the world, Spanish should not miss the semantic revolution. Even if it is not considered to be an interlingua, the minds behind AMR are considering the use of Natural Language Understanding (NLU) and Natural Language Generation (NLG) in a semantics based Machine Translation (MT) system. Recent studies for adapting it to other languages (Uresova et al., 2014; Li et al. 2016) have been done. But at the moment, there is only a small, publicly available corpus in a language other than English¹.

Thus, the central goal of this thesis is to investigate how to create Spanish AMRs in order to build a sizable Spanish sembank. We address this objective by answering the questions below:

- Is it possible to follow the current guidelines to annotate Spanish AMRs? And if not, how can the guidelines be refined in order to annotate Spanish AMRs? Also, what resources do we need to carry out such task?
- How similar are English and Spanish AMRs?
- What can be learned from the gathered information for future annotation efforts?

We believe this study could not only help lay the groundwork for building a large semantic bank for Spanish but also could be used as a reference for other languages.

1.2 Thesis Outline

The structure of this thesis is organized as follows. After providing the motivation for this project and our research questions in Chapter 1, Chapter 2 introduces the theoretical background of this thesis. In Chapter 3, we provide an overview of related work. Specifically, we review research on other major approaches to broad-coverage semantic representation. Chapter 4 is devoted to describing the methodology and the dataset used in this study. The qualitative and the quantitative results obtained are reported in Chapter 5. In Chapter 6, we discuss our findings in the light of our research questions as well as limitations of our approach. Finally, Chapter 7 presents conclusions and directions for future research.

¹ The Chinese Abstract Meaning Representation (CAMR) Bank is available at: <http://www.cs.brandeis.edu/~clp/camr/camr.html>

2 BACKGROUND

2.1 Computational Semantics

Any language can be analyzed through various levels of linguistic analysis. Inconsistency in the use of terminology makes it difficult to determine the exact number of these. One way to divide linguistics according to the level of language being studied considers that phonetics, phonology, morphology, syntax, semantics, and pragmatics constitute the levels of language analysis, as we can see in Figure 1.

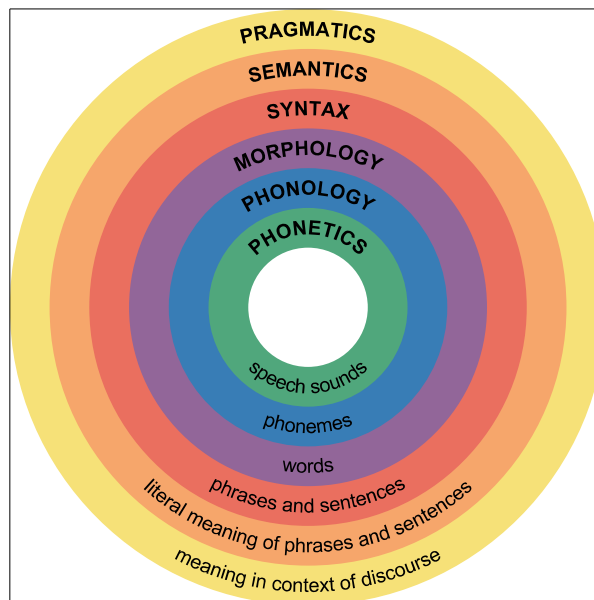


Figure 1. Levels of linguistic analysis according to Thomas and Cook (2005)

Both CL and NLP are concerned with the way computers and human (natural) languages interact in all dimensions, from phonetics to pragmatics. Figure 2 shows a pipeline view of the components of a generic NLP system, where each of the boxes corresponds to a particular type of processing that comprise a natural language analysis. Most of these systems include some kind of data preprocessor (including at least sentence segmentation, tokenization, and part-of-speech (POS) tagging). This initial preprocessing is an important step to ensure output quality. If we compare Figure 1 and Figure 2, we can see how an NLP system usually

goes from the smallest units of language to the largest ones, just as the aforementioned linguistics levels (phonetics being the smallest and pragmatics being the largest).

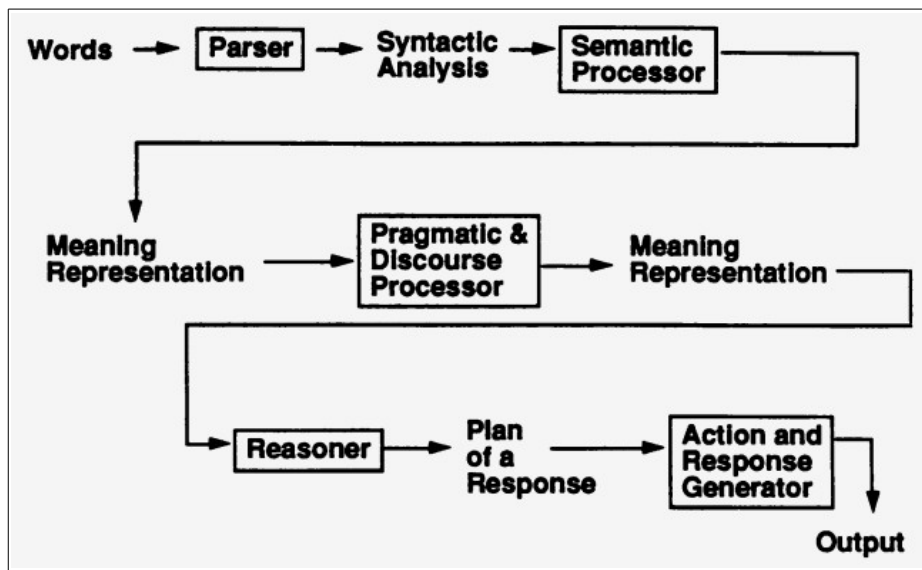


Figure 2. Pipeline view of an NLP system according to Bates (1995)

Blackburn and Bos (2013) define computational semantics as a discipline that combines insights from formal semantics, computational linguistics, and automated reasoning whose goal is to construct semantic representations for expressions of human language in an automatic way.

Because the meaning of a sentence depends so closely on its syntactic structure, there is no doubt that syntactic parsers play an important role in representing such meaning. Even if the analysis that these type of parsers return is far from being a complete one, phrase structure still helps to identify the semantic relationship that a predicate has with its given arguments in the description of a situation – also known as semantic roles (Crystal, 2008).

For instance, consider the following examples from Matthews (as cited in Chomsky, 1996):

- (1) The window broke
- (2) A hammer broke the window
- (3) The workman broke the window with a hammer
- (4) The window broke with a hammer

Syntactically speaking, the window is represented as the verb's subject in the first and fourth sentences and as its direct object in the second and the third sentences. All these sentences indicate that there is a broken thing: the window. However, a syntactic analysis is not able to depict this.

Who did what to whom, how, when, where, why, and with what consequences? It is amazing how a single sentence can tell so much. The ties between events and its participants cannot simply be ignored. Proper automatic detection of these is crucial for NLP systems to use information encoded in text effectively. To address this issue, a large variety of resources have been created to allow automatic semantic processing of text throughout the past few years. However, combining their information is not an easy task to do since each of these has its own idiosyncrasies.

On the one hand, we could assume that the problem of automatic syntactic analysis is rather solved based on the fact that there are many statistical parsers trained on manually annotated syntactic treebanks. The accuracy of state-of-the-art syntactic parsers – like Charniak and Johnson (2005) – is around 90%. These parsers are constantly improving in accuracy through the use of different techniques. A treebank is, by definition, a database of sentences which are annotated, often in the form of a tree. The term is believed to have been coined in the 1980s by Geoffrey Leech (Sampson, 2002). One of the most well-known English language treebanks is the Penn Treebank (PTB) (Marcus et al., 1993).

On the other hand, this is not the case for automatic semantic analysis. Some claim that this is because semantic annotation is balkanized (Banarescu et al., 2013). In other words, it is divided into separate annotations. The lack of a unified model makes it hard to integrate the various kinds of annotation data. There has been a shift in the direction of research to overcome this challenge. To be precise, a shift towards graph-based parsing, aiming a more direct semantic analysis of whole natural language sentences.

2.2 Abstract Meaning Representation

AMR stands for Abstract Meaning Representation. The concept of AMR was firstly introduced back in 1998 by Irene Langkilde and Kevin Knight (1998). However, it has

evolved since that time. For the purpose of this work, we follow a modern definition of AMR. This considers AMR as a semantic representation language that appears based on the assumption that we lack a simple readable semantic bank – or sembank – of natural language sentences “paired with their whole-sentence, logical meanings” (Banarescu et al., 2013). Thus, they recently started annotating the logical meaning of sentences, for which a particular sentence is encoded as a Directed Acyclic Graph (DAG). The ultimate goal is to encourage advances in different NLP tasks, including Statistical Natural Language Understanding (SNLU) or Statistical Machine Translation (SMT), amongst others. The idea is to do this by enabling rapid human annotation of corpora with broad coverage. Some phenomena that AMR deals with include discourse connectives, semantic roles, intrasentential coreference, named entities (and wikification), questions, negation, and modality, among others.

The fundamentals of the annotation scheme are the following:

- AMR features a three-way format: a traditional logic format, an AMR format, and a graph format. These three are equivalent. An example can be seen in Figure 3:

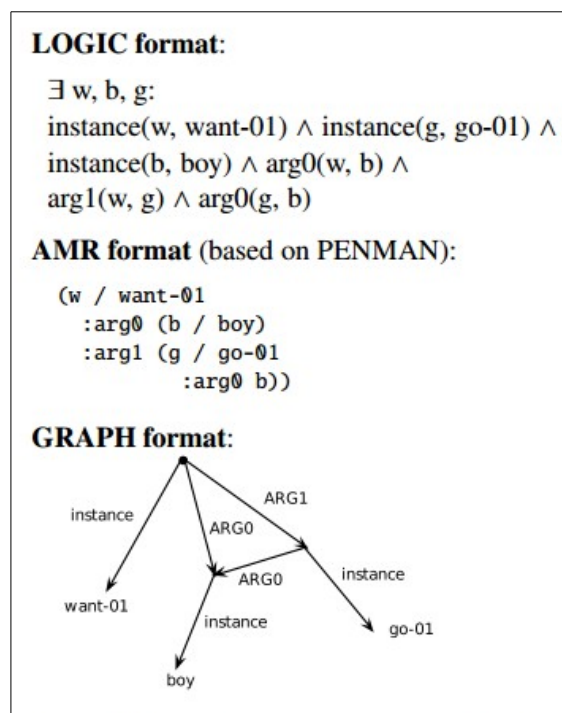


Figure 3. Equivalent formats for representing the meaning of “The boy wants to go” according to Banarescu et al. (2013)

The AMR format is based on PENMAN notation (Matthiessen et al., 1991) so that it is easy for human reading and writing. At the same time, a conventional graph notation is used in an effort to make it easier for programs to traverse. The latter is based on neo-Davidsonian event representations (Davidson, 1969).

- AMR strives to capture many aspects of meaning in a single simple data structure. To do that, it abstracts away from morpho-syntactic idiosyncrasies. After all, it focus on logic rather than in syntactic representation.
- AMR uses PropBank (PB) framesets (Palmer et al., 2005). Thus, each frame presents annotators with a list of senses. Seemingly, each sense has its own definitions. And each definition has its own numbered core arguments.
- AMR makes use of approximately a hundred semantic relations. Some examples of role categories are core “:ARGX” roles (frame arguments), non-core roles (general semantic relations), roles for quantities, roles used in date-entity, roles of the form “:opX,” roles of the form “:prep-X,” multi-sentence roles, and a conjunction role. Simple roles often correspond to a reified concept or concepts.
- AMR does not dictate a mandatory way of applying rules. Instead, it promotes personal interpretation of researchers about how strings are related to meanings.
- AMR is not an interlingua. In fact, it is biased towards English.

2.2.1 How does AMR Treat Different Linguistic Phenomena?

Every AMR has a single root node at the top of the graph, which is considered to be the focus. Each node in the graph has a variable and represents a semantic concept. We could think of variables as instances of concepts. A slash is used to indicate this. Variables are re-used if something is referenced multiple times. This is known as reentrancy. It is worth noting that it does not matter where the concept label goes. The relations between these concepts are denoted by graph edges. Concepts include PB framesets and English words, whereas semantic relations include different types of roles, which are marked by a colon prefix. Some relations

– known as constants – get no variable, just a value. An interesting feature of AMR is the fact that roles can be inverted. Inverse roles are useful for maintaining a single rooted structure. It is also possible to convert a role into a concept. This is called reification and it can be used to make a relation the focus of an AMR fragment. Yet not all relations have reifications. Figure 4 shows an example of this description.

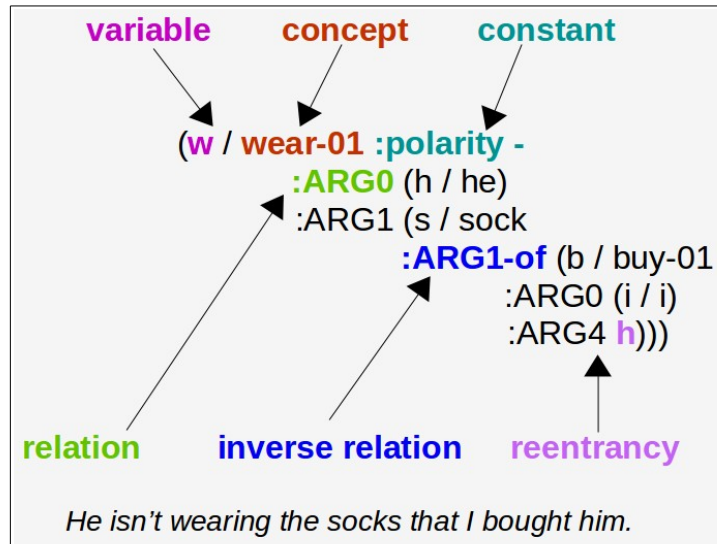


Figure 4: An AMR, with its different key elements.

As mentioned above, AMR abstracts in numerous ways. Not only it does not account for tense or number but also does not care about word category or word order. The former decision was made, taking into account that the English verb tense system does not generalize well cross-linguistically. The latter has to do with the fact that AMR leaf-labels are thought to be concepts rather than words. Thus, AMR assigns the same conceptual structure to different word categories.

Because all mentions of a term go to the same variable, AMR does drop articles too. Most prepositions are dropped, except for time and location prepositions. Again, the idea behind these abstractions is to keep a simple representation and to assign the same AMR to sentences that have the same basic meaning. This abstraction is good in the sense that it speeds up the annotation process.

AMR covers different phenomena apart from within-sentence coreference, inverse relations, focus, and reified concepts. Some of these phenomena are listed below. For more

information, the reader is referred to the AMR guidelines², where each phenomenon is described in more detail.

Named Entities

In AMR, the type of a given Named Entity (NE) is identified from the list of categories provided by AMR. As a last resource, the type “thing” is used. Then, the NE is followed by a role “:name” and a concept of the same name. This concept, in turn, gets “:opX” relations to the strings of their name as used in the sentence, whether the NE has only a single word or two. NEs are often referred to in different ways. Moreover, hyphenated compounds or possessive suffixes are not splitted.

Wikification is used to avoid ambiguity and to unify different surface forms of a NE to a canonical form: a “:wiki” role. If a NE is not covered by Wikipedia, then it is marked as “:wiki -.”

Copulas

AMR tries to eliminate purely grammatical words such as copulas. It does this in different ways. For example, “noun is noun” constructions are usually represented with the role “:domain.” Same role is used in predicate adjectives, as long as there is no adjective frame available.

Even though the existential “be” is kept and annotated with the concept “be-02,” the existential “there is” and the existential “there are” are dropped in AMRs and, as a result, have simple AMR representations.

Modality

Modality is represented by means of concepts. Each modal verb corresponds to at least one concept and, likewise, the same concept can correspond to more than one modal verb. E.g. the concept “possible-01” could represent verbs like “can,” “could,” “may,” or “might.”

2 Available at: <https://www.isi.edu/~ulf/amr/help/amr-guidelines.pdf>

Negation

Negation is represented logically in AMR. Thus, the role “:polarity -” and its reified concept “have-polarity-91” denote negation. Some verbs allow negative raising, which consists in transferring the negative element from a verb to another. For instance, representing the sentence “I don’t think I will go” as “I think I won’t go.”

Wh-questions, Other Interrogatives, Imperatives, and Exclamatives

The special concept “amr-unknown” is used to capture wh-questions. On the other hand, yes-no questions and yes-no embedded clauses are treated separately with the AMR relation “:mode interrogative,” whereas questions as requests are treated with the relation “:mode imperative.” The relation “:mode imperative” is also used for imperatives as well as for exclamative imperatives. Finally, AMR uses the role “:mode expressive” to mark exclamational words that express emotions, which do not particularly refer to a clear event, object, or property.

Implicit Arguments and Relations

Both roles and concepts may be implicit, depending on the circumstances. If a role is clearly implied by the sentence, then it must be annotated. Generally, implicit concepts are not introduced. But occasionally, we have to “hallucinate” the relationships that are grammatically underspecified. E.g. when building AMRs for proper names or “-er” nouns.

How to Treat Some Nouns, Adjectives, and Adverbs

Certain nouns and adjectives invoke predicates. Thus, AMR maps many nouns and adjectives to their verbal forms. This process is called verbalization. AMR treats light-verb constructions in a similar way, it strips them away and it verbalizes the given noun. Of course, AMR does not break down the meaning of a noun or an adjective when it substantially differs from the verbal form.

On a slightly different note, adverbs ending in “-ly” normally get stemmed to adjective form. Again, these are kept if the meaning of the word is significantly distinct.

2.2.2 The AMR Editor and Smatch

A web-based editor has been developed to construct AMRs via text commands and graphical buttons. The AMR Editor provides a dictionary that is full of examples with explanations and a search engine to see how certain phrases were handled in the past, among other features. These resources are of great help to annotators. Not to mention that they are important to ensure annotation consistency and Inter-Annotator Agreement (IAA).

Cai and Knight (2013) introduce the Smatch metric to assess both IAA and automatic parsing accuracy. This metric computes the degree of overlap between two AMRs. In order to do that, Smatch executes a brief search to compare two AMR's triples with respect to their previously calculated precision, recall, and F1 score. Obtaining the variable mapping which yields the highest F1 score is essential to match up variables from two input AMRs.

3 RELATED WORK

In CL and NLP, trees have been the traditional means to represent natural language. However, with a gradual shift of focus from syntactic analysis towards semantic interpretation due to its potential applications in fields like Text Summarization, Question Answering or Information Extraction, other parsing methods have come on the scene to offer a more appropriate representation.

To set an example, graph-based methods for semantic analysis – and NLP in general – have become more and more popular during the last years. Graphs seem a convenient option when it comes to capture semantic structures since an entity can play multiple roles in a sentence. Apart from graphs, there are also other non-tree structures which primary aim is to go beyond shallow representations and, in turn, to cover deeper analyses.

In this chapter, we compare AMR to some of the major approaches to broad-coverage semantic representation in NLP: the Treebank Semantics System (section 3.1), the Groningen Meaning Bank (section 3.2), and Universal Conceptual Cognitive Annotation (section 3.3). At the same time, we briefly comment on other approaches and related research in section 3.4.

3.1 The Treebank Semantics System

Butler and Yoshimoto (2012) propose a sembank generation technique that emanates from the information contained in conventional syntactic treebanks. This system is called Treebank Semantics and it is built primarily on Butler’s Scope Control Theory (SCT) (Butler, 2010). Its overall idea is to mimic natural language by means of supporting the extension, manipulation, and reduction of scope interactions. In other words, by means of controlling the scope.

This system converts constituent syntactic trees into expressions of a small formal language based on SCT. Such language is the input to the system, which can be subsequently processed

to output meaning representations that are based on Davidsonian predicate logic (Figure 5). Once again, the main goal is leading to the creation of a broad-coverage semantic corpus.

```
∃t1e2e3(  
  week(t1) ∧  
  has_been_meaning(e3, mary, go(e2, mary)) ∧ is_contained_in(for(e3), t1))
```

Figure 5: Meaning representation output of the Treebank Semantics System
for the sentence “Mary has been meaning to go for a week”³

This data conversion into meaning representations is done automatically. Therefore, we could think of it as a syntax-to-semantics automatic translator. The Penn Treebank (PTB) is an example of treebanks from which information has been obtained to carry out this project. Currently, there is a partially parallel corpus of English and Japanese meaning representations automatically generated by the system and then human checked. It is available online⁴.

Future work will deepen the conversion of annotations of different schemes to corresponding meaning representations content.

3.2 The Groningen Meaning Bank

Another approach that aims to create a large scale semantic bank is the one introduced by Basile et al. (2012): the Groningen Meaning Bank (GMB) project. Comparing AMR and GMB is interesting because they are similar, yet subtly distinct. In spite of sharing a common objective, which is to integrate various phenomena into a single formalism, they differ in the way they handle language.

To start with, GMB’s units of annotation are texts rather than isolated sentences. To get an approximation of the target annotations, GMB depends on a robust pipeline which contains a few tools like C&C tools and Boxer (Curran et al., 2007; Bos, 2008). Unlike AMR, GMB follows a more conventional approach that integrates part of speech tags and syntactic structure, among other levels of annotation.

³ Retrieved from: <http://www.compling.jp/ts/overview.html>

⁴ The Treebank Semantics Corpus is available at: <https://github.com/ajb129/tscorpus>

One of the major features of GMB is that its annotation is semi-automatic. Initial annotations in GMB are automatically generated by the toolchain mentioned above. At the same time, their output can be corrected or refined. Those changes and corrections are defined as Bits of Wisdom (BOWs) and are made by human annotators – who may be experts or not – as well as by external tools, like a Word Sense Disambiguation (WSD) system, at any stage of the processing. A BOW is simply a traceable database entry that improves the quality of annotation by giving advice on a particular decision regarding linguistic interpretation. There is currently no formal guide used to identify errors. To provide such “wisdom,” expert annotators edit the annotation in a web-based interface named GMB Explorer⁵. On the other hand, non-experts play an online game called Wordrobe (Venhuizen et al., 2013) with a purpose for crowdsourcing. Thus, they perform such corrections while playing and betting on the correctness of their answers. Figure 6 shows an example of a question that a player can encounter. If there is a contradiction between BOWs, preference is given to more recent BOWs and to expert annotators by a judge component.

Figure 6: Example of the game *Names of Wordrobe*⁶

The representation format used by GMB is based on a formal semantic theory known as Discourse Representation Theory (DRT) (Kamp and Reyle, 1993) and it comes in the form of Discourse Representation Structures (DRSs). Not only can these be extended to add particular phenomena like neo-Davidsonian events (Davidson, 1969) but they also can be translated into First-Order Logic (FOL), which is very convenient since it makes the use of existing inference engines possible. Figure 7 shows an example of a DRSs.

5 GMB Explorer: <http://gmb.let.rug.nl/explorer/explore.php>

6 Retrieved from: <http://wordrobe.housing.rug.nl/Wordrobe/Member/WordrobePlayPage.aspx>

Because GMBs are constrained by syntactic structure, their content is mostly compositional. In that sense, the non-compositional style proposed by AMR is more straightforward, since annotators encode what they think the meaning of a sentence is directly. However, the use of different tools combined with human help allows GMB to integrate quite a variety of phenomena, including some aspects that AMR does not cover explicitly like presupposition or scope. Similarly, AMRs are able to display certain phenomena that DRSs cannot at the moment. For instance, verb phrase ellipsis is not a problem for AMR. It is noteworthy that a suggestion has been made to represent projection and scope phenomena in AMR by extending its syntax and translation to FOL (Bos, 2016).

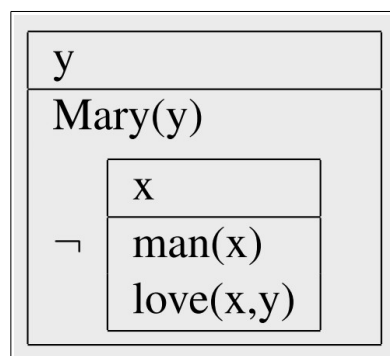


Figure 7: DRS for the sentence “No man loves Mary” (Venhuizen, 2014)

One of the project’s primary concerns is to reduce the need for manual correction and, in turn, to build a high quality gold-standard resource in the future. Their hypothesis is that the annotation accuracy improves by increasing the amount of BOWs that are applied. Therefore, an increment of BOWs will lead representations to come closer to a gold standard. According to Weck (2015), for future work, they are interested in finding more sophisticated ways of resolving the disagreement among BOWs as well.

3.3 Universal Conceptual Cognitive Annotation

Abend and Rappoport (2013a) present Universal Conceptual Cognitive Annotation (UCCA), a multi-layered framework for semantic representation where each layer specifies the encoded relations. Following that idea, it incorporates a syntactic hidden layer to learn the mapping between form and meaning. Yet, it diverges from the two aforementioned approaches in the sense that it is less coupled with a particular syntactic theory. After all, learning algorithms allow UCCA to automatically deduce syntactic categories and structures from semantic ones.

The theory behind UCCA is Basic Linguistic Theory (BLT), a term coined by Dixon (2005) that takes semantic similarity as its primary criterion to perform intra- and cross-linguistic categorization of constructions.

Just as AMR, UCCA uses Directed Acyclic Graphs (DAGs) which are intended to abstract away from specific syntactic constructions in order to represent semantic relations. As its name suggests, they refer to “conceptual” notions as opposed to “syntactic” ones. Units, terminals, and categories comprise a DAG. Units are the graph’s nodes. These are either terminals or several elements jointly viewed as a single entity according to some semantic or cognitive consideration. Hierarchy is formed by using units as arguments or relations in other units. Terminals are the graph’s leaves. These are atomic meaning-bearing units. The definition of terminals include arbitrary morphemes, words, multi-word chunks. Edges bear a category, indicating a unit’s role. Thus, categories annotated over outbound edges represent the internal structure of a unit whereas the ones annotated over inbound edges represent the roles a unit plays in the relations it participates.

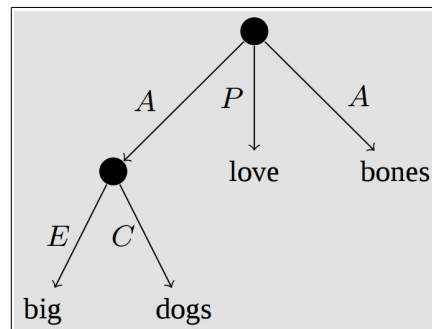


Figure 8: Example of an UCCA annotation for the sentence
 “Big dogs love bones” (Abend and Rappoport, 2013b)

As we have previously mentioned, UCCA is built as a multi-layered structure. Some of the most relevant relations, like argument structures and their interrelationship, are represented in its foundational layer. This is inspired mostly by theories like Cognitive Grammar (Langacker, 2008) that consider an utterance’s meaning to be related to the mental representations that it evokes rather than to its reference in the world. It is also inspired by work in linguistic typology along with neuroscience. This layer is designed to cover an entire text, which is seen as a collection of scenes. A scene describes a temporally persistent state or a temporally evolving event. Every scene contain a main relation. It may contain at least one participant or none. Scenes can contain secondary relations too. Apart from the scene

elements, the complete set of categories in UCCA’s foundational layer incorporates elements of non-scene units, inter-scene relations, and other units. The first is composed of sub-units of non-scene units. The second includes more complex cases that involve units that participate in more than one relation. And, last but not least, there are other units which do not introduce a new relation or entity into the scene. For more details about these categories, the reader is referred to the UCCA guidelines⁷.

A web-based application is available for the annotation process. Because of this reduced set of categories, UCCA makes annotation accessible for non-expert annotators. This is, undoubtedly, a great advantage over other annotation schemes that involve far more elaborate representations that require the work of experts, especially for large projects. On the other hand, it limits the coverage of semantic phenomena – even if it takes texts rather than sentences as the units of annotation. It is worth mentioning that there is no standard evaluation to measure IAA per se. In fact, it is hard to compute due to the way the UCCA interface is built – by allowing more than one acceptable, non-contradictory analysis.

Future work will focus on UCCA’s cross-linguistic portability and how to further reduce manual annotation through the use of a range of techniques. Moreover, they will explore the use of UCCA to apply it to several semantic tasks. E.g. MT.

3.4 Other Approaches and Related Research

The classical approach to derive a formal semantic representation from a natural language expression was to do this derivation from syntactic analyses using hand-written rules. During the last decades, the focus has been shifted towards the use of computational resources. Indeed, the task is now to automatically map strings to semantic representations. To do this, a system needs annotated corpora to learn from. As we have shown, graph-based methods are just one of the many modern approaches to infer semantic meaning.

Besides the theory of AMR, there are other popular graph-based formalisms for semantic representations that have potential applications in NLP or NLU. Some examples include Dependency Minimal Recursion Semantics (DMRS) (Copestake, 2009), Hyperedge

⁷ Available at: http://www.cs.huji.ac.il/~oabend/ucca/guidelines_corpora.pdf

Replacement Grammar (HRG) (Chiang et al., 2013), or S-graph Grammars (Koller, 2015), inter alia. Some of them, like the one proposed by Oepen et al. (2015), have released corpora.

In the words of Bos (2016), AMR is the “new kid on the semantics block.” Therefore, it is not surprising that everyone wants to be friends – and play – with the new kid. A wide range of parsers to generate AMR formalism from plain text have arisen with the publication of public corpora⁸ as well as with the appearance of SemEval tasks⁹ (Flanigan et al., 2014, Artzi et al., 2015; Wang et al., 2016; Damonte et al., 2017). In addition, others like Song et al. (2016) or Flanigan et al. (2016) have come up with ways to generate English sentences from AMRs.

Proposals have been made for improving AMR. For instance, to reduce the impact of noise (Goodman et al., 2016) or to deal with projection phenomena and quantifier scope (Bos, 2016). Furthermore, AMR has been used for different purposes: to link entity mentions (Pan et al., 2015) or to improve Event Detection (ED) (Li et al., 2015), just to set a couple of examples.

Even though AMR is highly biased towards English, it abstracts away from morpho-syntactic idiosyncrasies which account for many of the cross-lingual differences. Taking this fact into account, it is expected that many relations and given structures will be similar or even identical across languages. As a result, cross-linguistic research has been conducted to see if it is viable to adapt AMR to other languages in order to use it for different purposes, like MT (Xue et al., 2014; Vanderwende et al. 2015; Saphra and Lopez, 2015). As of now, there is a Chinese AMR corpus that has been developed (Li et al., 2016).

As a “kid,” AMR has a bright future ahead of it.

8 Many of which can be found at: <http://amr.isi.edu/download.html>

9 SemEval 2016: <http://alt.qcri.org/semeval2016/task8/> and SemEval2017: <http://alt.qcri.org/semeval2017/task9/>

4 METHODOLOGY

As previously explained, the central part of our work is to investigate how to create Spanish AMRs in order to build a sizable Spanish sembank. Even if AMR abstracts away from certain language idiosyncrasies that usually present a problem in cross-lingual NLP tasks, AMR is not meant to be an interlingua. We know that AMR is closer to English than to any other language and, considering the fact that English differs from Spanish in many ways, we deduce that the correlation between AMR annotation standards – as they are – and Spanish is not as high as it could be. To set an example, Spanish morphosyntax allows the use of a variety of pronouns (e.g. enclitic or proclitic pronouns) that English do not have as well as it allows a less restrictive, more frequent use of nominal ellipsis in comparison to English. These differences imply the need for new standards to capture Spanish language effectively. Thus, in this chapter, we proceed to describe the methodological procedures used in this linguistic study to see if the semantic representation language proposed by Banarescu et al. (2013) can be refined to provide a broad coverage of the Spanish language and if so, how to face this challenge.

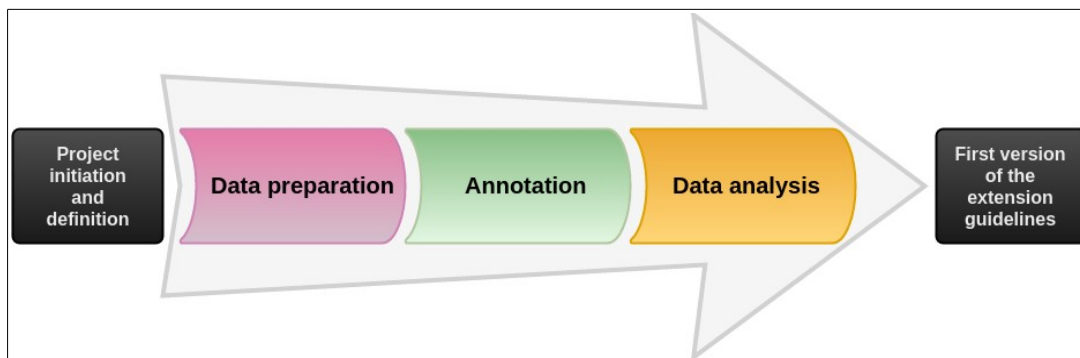


Figure 9: Methodology stages

Our methodology can be summarized by the application of the following three stages (illustrated in Figure 9). The different stages of this methodology are described in the subsequent sections.

4.1 Data Preparation

Our data consists of a set of 50 parallel bilingual (English and Spanish) AMR-annotated sentences (Table 1). Originally, we wanted to work with the AMR corpora released by the Linguistic Data Consortium (LDC) because it offers texts from different domains. Unfortunately, it is not in the public domain. For now, only two AMR corpora are publicly available¹⁰ for researchers: Bio AMR Corpus and *The Little Prince* Corpus. The first one includes texts – mostly articles and papers – from the biomedical domain, whereas the second one, as its name suggests, contains the full text of Antoine de Saint-Exupéry’s famous novel *The Little Prince*. All 50 sentences paired with their whole-sentence, logical meanings were selected from the latter since the language found in this corpus was closer to an everyday language than the technical language of biomedical literature. The translation of these was done and revised manually. We chose to do this over using an already translated version of public domain for the sole reason that the English and Spanish versions show contrasting styles of writing (e.g. a single sentence in the Spanish version can be composed of two or three sentences of the English one).

The selection of these sentences was not a random process. First, we translated about 250 sentences from *The Little Prince*. Then, we made a comparison between the resulting pairs of English-Spanish sentences. Out of those, we narrowed down the number to 50 – resulting in the selection of sentences that include a wide range of linguistic phenomena, including nominal ellipses, clitic pronouns, gender, verbal periphrases and locutions, double negatives, nominalization and verbalization, affixes, and some key words that have a special treatment in AMR. To choose these, we also paid attention to the level of structural cross-lingual variation that the sentence pairs exhibit, just so we could study both language pairs whose structures align well and those whose structures do not align well. Declarative, interrogative, exclamatory, and imperative sentences are covered.

For the purpose of this project, we consider that this number of sentences – in spite of being small – is good enough to detect a reasonable amount of Spanish-specific constructions that the current English-only version of AMR is not able to represent and to compare English and Spanish AMRs according to their structural similarity.

¹⁰ Available at: <https://amr.isi.edu/download.html>

Language	Spanish	English
Number of sentences	50	50
Number of words	422	483
Min. # of words per sentence	2	2
Max. # of words per sentence	26	27
Avg. sentence length	8	9
Most common sentence length	6	7
Number of nouns	74	71
Number of pronouns	51	70
Number of verbs	100	117
Number of adjectives	30	48
Number of adverbs	31	23

Table 1: Text statistics of our data

Because there was only one person performing the annotation and, hence, no way of providing a gold standard, we decided that the best option for us was to translate English sentences that already have their corresponding AMR representations annotated into Spanish. We did this in order to avoid biasing the analysis by relying on a single person to create all the content (both English and Spanish AMRs). This was also the most convenient option in terms of time consumption and productivity.

4.2 Annotation Procedure

This section summarizes the annotation procedure that was followed. The procedure involves three linear phases: the preparation phase, the problem identification phase, and the annotation phase.

The main purpose of the preparation phase is to train human annotators on how to annotate English AMRs, based on the premise that this would serve as a foundation for developing the Spanish version of AMR. This phase is further divided into a training task and a practice task. For the first task, we adopted a self-training method as far as the tool usage and the annotation scheme were concerned. The learning material that we used – including video tutorials and guidelines – is all freely available online¹¹. Once that an annotator gets familiarized with the content of the learning material, they are ready to put into practice the

11 Tutorial: <https://github.com/nschneid/amr-tutorial>, guidelines: <http://amr.isi.edu/language.html>, and editor overview: <http://amr.isi.edu/editor.html>

acquired knowledge. Then, in the second task, learning is reinforced through hands-on experience by annotating English AMRs.

The main focus of the problem identification phase is to identify linguistic phenomena in Spanish that cannot be represented when applying the current AMR guidelines. Of course, the goal of this phase is to find a way to solve these “problems” that arise in the attempt to annotate Spanish AMRs. During the preliminary stage of this phase, we manually created a sample of Spanish AMRs according to the annotation scheme introduced in Chapter 2. Next, we identified any missing aspects of meaning that AMR fails to represent and that cannot be ignored. After carrying out this task, we designed an extension of the annotation scheme to potentially meet those needs. This refinement in AMR annotation standards stays true to the syntax of the AMR.

During the annotation phase, we followed this extension. The final annotation of Spanish AMRs can be found in Appendix 1. In addition, details of the proposed changes are discussed in Chapter 5.

The annotation was performed manually by a human annotator, without any tool support. We took advantage, however, of the AMR Editor since it is a great resource for assisting annotators. Among other features, this web-based editor contains a rather interconnected documentation with full examples as well as a search function that allows users to see how given phrases were handled in the past. We did not use this tool as an editor simply because it does not work for Spanish. To facilitate the annotation process, the adaptation of the current editor or the development of a new annotation tool would have been ideal, but we agreed that this is outside of the scope of this thesis.

As we mentioned in Chapter 2, AMR makes extensive use of PropBank framesets to represent some semantic concepts. The AnCora corpus (Taulé et al., 2008) is the only available resource that uses Propbank-style roles for Spanish. At first, our plan was to use this as a reference in order to manually map the core arguments to Spanish numbered verb senses. But then, during the problem identification phase, we realized that many of the mappings that we needed to obtain from AncoraNet were inaccurate. Not only some words lacked senses but also some of the mappings were assigned over and over again for different – or even the same – senses of a given word. Not to mention that AncoraNet is connected to an outdated

version of the PropBank’s inventory. For instance, the entry for the word *querer* has two senses: the first one is connected to the English senses “mean.01” (current “mean.02”) and “try.01,” whereas the second one is connected to the English sense “love.01” twice. Yet, it is not connected to senses like “want.01” or “wish.01.” Because of these issues, we made the choice to manually map core arguments to Spanish unnumbered verb senses from scratch. We did this in the following way: we translated a given word from Spanish into English, then look for the appropriate sense of such word in the current PropBank inventory and, finally, we used its corresponding roles for the Spanish sense. To avoid confusion for the reader, we decided not to number the Spanish verb senses that we used to annotate our data.

4.3 Data Analysis

Considering that there is no widely-used metric to evaluate the structural overlap of AMRs of an English sentence and its translation, and due to the nature of this project, we took a visual comparison approach that analyzes our data quantitatively to evaluate bilingual structural similarity between pairs of AMR graphs. A similar approach was used by Uresova et al. (2014).

For this comparative analysis, we focused on the number of bilingual AMR pairs that show what we call a Substantial Structural Similarity (SSS) or a Partial Structural Similarity (PSS). The first category corresponds to AMR pairs that are considered structurally equal in the sense that all concepts and relations are aligned, whereas the second one refers to pairs of AMRs which are partially similar in structure. Furthermore, we distinguish the latter pairs according to whether or not they have the same top node because there is only one concept of focus for each AMR and it happens to be represented at their top-level root. Thus, these can be subcategorized into PSS 1 and PSS 2, respectively. Whenever a pair of English-Spanish AMRs shows major differences in structure, we mark them. More precisely, we were interested in pinpointing differences regarding framesets, non-core roles, and what we call non-role concepts – that is, concepts that do not bear roles (e.g. the concept “lemon”). We believe a close look at these differences will pave the way for a unified annotation scheme. If the bilingual AMR pairs are completely different, then we categorize them as pairs that display a Total Structural Difference (TSD).

Additionally, we took into account the Degree of Syntax Similarity (DSS) of the bilingual sentence pairs to study any possible correlation between the syntactic similarity of pairs of English-Spanish sentences and their corresponding pairs of AMRs. Thus, we differentiate between three DSSs: degree 1 or DSS 1 (extremely or very similar), degree 2 or DSS 2 (somewhat or slightly similar), and degree 3 or DSS 3 (not at all similar). Once again, we followed a visual comparison approach to make these distinctions. To do this, we focus on elements that appear on a sentence, rather than the order in which they appear.

To better understand this classification, consider the following pairs of bilingual AMRs (Examples 5, 6, 7, and 8).

(5) It is a pair of bilingual AMRs with SSS and a DSS 2:

Pero el principito no respondió

(c / contrastar

:ARG2 (r / responder :polaridad -

:ARG0 (p / príncipe

:mod (p2 / pequeño)))

But the little prince made no reply

(c / contrast-01

:ARG2 (r / reply :polarity -

:ARG0 (p / prince

:mod (l / little)))

It is easy to see that both AMRs are completely equivalent in the fifth example. However, their syntactic structure is only somewhat similar. The English version includes the verb “to make a reply” – which is simplified to the verb “to reply” in the AMR after removing the light verb “to make” – but in Spanish, one would not find the combination of *hacer una respuesta*. Thus, we consider this pair of sentences to be of degree 2 due to such difference regarding idiomaticity.

(6) It is a pair of bilingual AMRs with PSS 1 and a DSS 3:

¡Qué gracioso!

(g / gracioso

:grado (t / tan))

That is funny!

(f / funny

:domain (t / that))

In the sixth example, the pair of AMRs do not share the same structure but we still find some similarity between them. Also, their concept of focus is the same: “funny” or *gracioso*. On the other hand, by looking at the syntax, we can see that these two sentences are quite different regarding that sense. In fact, the literal translation of the Spanish sentence into English would be something like “How funny!”

(7) It is a pair of bilingual AMRs with PSS 2 and a DSS 1:

Olvidar a un amigo es triste	To forget a friend is sad
(t / triste	(s / sad-02
:campo (o / olvidar	:ARG0 (f / forget-01
:ARG1 (p / persona	:ARG1 (p / person
:ARG0-de (t2 / tener-rol-rel-91	:ARG0-of (h / have-rel-rol-91
:ARG2 (a / amigo))))	:ARG2 (f2 / friend))))

Then, in the seventh example, we also find a partial similarity between this pair of bilingual AMRs. Not only that but also they do not have the same top node: the Spanish AMR uses the concept *triste* followed by the role *:campo* (“:domain”), whereas the English AMR uses the frameset “sad-02” with its correspondent argument. Of course, part of this dissimilarity has to do with the fact that we did not consider the use of adjectival relations other than the role “possible-01” to annotate the modal verb “can.” The sentences, however, are structurally similar. This is because all elements are present in both sentences.

And, last but not least, the eighth pair of AMRs is an example of a bilingual pair that shows a TSD. Interestingly, this is a case where, syntactically speaking, the sentences are equivalent.

(8) It is a pair of bilingual AMRs with TSD and a DSS 1:

Caminarás cuando quieras descansar y el día durará tanto como quieras	When you want to rest, you will walk and the day will last as long as you like
(d / durar	(w / walk-01
:ARG1 (d / día)	:ARG0 (y / you)
:ARG2 (t / tanto	:time (w2 / want-01
:grado (i / igual)	:ARG0 y
:comparado-con (q / querer	:ARG1 (r / rest-01
:ARG0 (t2 / tú)))	:ARG1 y))
:condición (c / caminar	:ARG0-of (c / cause-01
:ARG0 t2	:ARG1 (l / last-01
:tiempo (q / querer	:ARG1 (d / day)
:ARG0 t2	:ARG2 (t / temporal-quantity
:ARG1 (d / descansar))))	:degree (e / equal)
	:compared-to (l2 / like-02
	:ARG0 y))))

After carrying out this last step in the process, we found very interesting results. It is important to point out that these should be taken as the first step towards the development of AMR guidelines for Spanish. Such results are introduced in the next chapter.

5 EVALUATION

In this chapter, we present two types of evaluation: qualitative and quantitative. As described in the methodology, an interplay of theory and practice served as a basis to draw theoretical conclusions. On the one hand, we expanded the AMR annotation guidelines that were heavily biased towards English. The linguistic phenomena that we covered is described in the first section. On the other hand, we applied this extension to annotate Spanish AMRs and we analyzed the resulting 50 pairs of bilingual AMRs that we obtained. In the second section, we report on our quantitative analysis.

5.1 Phenomena & Extended Guidelines

In this section, we introduce Spanish linguistic phenomena that cannot be represented in AMR under the current guidelines listed in the following 7 subsections, together with a proposal to extend the annotation scheme for each of them. But, before we start, it is important to point out the reason why we chose to adapt the current guidelines by converting some roles, reifications, modals, and special words into their Spanish counterpart.

Indeed, AMR is not meant to function as an international auxiliary language. As a consequence, we find that most resources like the guidelines or the editor are available exclusively in English. Therefore, we find all concepts and relations in English as well.

As we have mentioned in Chapter 3, we have manually mapped core arguments to Spanish unnumbered verb senses from scratch. Hence, concepts are mostly annotated in Spanish. For many words like “cat” (*gato*) or “like” (*gustar*), this rule works well.

(9) (c / cat) → (g / gato)

(10) (l / like-01) → (g / gustar)

However, this is not always the case. Here the problem comes when we attempt to annotate certain roles and relations that either receive a special treatment and/or lead to confusion.

For instance, many roles like “:prep-X” present a problem. Not only because the annotator may not know to what such role is referring to but also because the equivalences between Spanish and English prepositions might not be straightforward, one-to-one correspondences. To set an example, the English equivalents for the Spanish preposition *en* could be either “in,” “inside,” “into,” “within,” or “by.” In such case, an annotator would have to figure out which preposition should be used – and how, because not all prepositions are annotated with the role “:prep-X.” It is needless to say that making this type of decisions is beyond the annotator’s competence.

Other roles, at the same time, correspond to at least a reified concept – that is, a role that has being converted into a concept (for instance, “be-located-at-91” is the reified concept of the role “:location”). Because the purpose of these concepts is to connect to more than two concepts, or to be at the top of the AMR, they have their own core “:ARGX” roles. However, since they are special concepts, they have no direct connection with any Spanish sense.

(11)

El sobre está en el cajón.

(b / be-located-at-91

:ARG1 (s / sobre)

:ARG2 (c / cajón))

The envelope is in the drawer.

(b / be-located-at-91

:ARG1(e / envelope)

:ARG2 (d / drawer))

Similarly, AMR represents some special words (Examples 12, 13, and 14) as well as syntactic modals (Examples 15, 16, and 17) using specific concepts or relations. Consider the following examples:

(12) please → :polite +

(13) according to → (s / say-01)

(14) but → (c / contrast-01)

(15) would → :mode imperative (question as request)

(16) shall → (o / obligate-01)

(17) should → (r / recommend-01)

The effect of annotating the aforementioned roles and relations is illustrated by means of an example. As shown in the following pair of AMRs, the concept “contrast-01” refers to the word “but” and the role “:polarity -” is used to represent negation. Then, the word “what” is annotated with the concept “amr-unknown,” “without” with its corresponding “:prep-X” role,

and the conjunction “and” with a concept of the same name followed by a couple of operators “:opX.” Moreover, the role “:poss” is used to indicate possession. The reader is referred to the guidelines for detailed information about these roles and concepts.

(18)

<p>Pero no sé qué hacer sin ti y sin tu música</p> <p>(c / contrast-01</p> <p style="padding-left: 2em;">:ARG2 (s / saber :polarity -</p> <p style="padding-left: 4em;">:ARG1 (y / yo)</p> <p style="padding-left: 2em;">:ARG2 (h / hacer</p> <p style="padding-left: 4em;">:ARG0 y</p> <p style="padding-left: 4em;">:ARG1 (a / amr-unknown)</p> <p style="padding-left: 2em;">:prep-without (a / and</p> <p style="padding-left: 4em;">:op1 (t / tú)</p> <p style="padding-left: 4em;">:op2 (m / música</p> <p style="padding-left: 6em;">:poss t))))</p>	<p>But I don't know what to do without you and your music</p> <p>(c / contrast-01</p> <p style="padding-left: 2em;">:ARG2 (k / know-01 :polarity -</p> <p style="padding-left: 4em;">:ARG1 (i / I)</p> <p style="padding-left: 2em;">:ARG2 (d / do-02</p> <p style="padding-left: 4em;">:ARG0 i</p> <p style="padding-left: 4em;">:ARG1 (a / amr-unknown)</p> <p style="padding-left: 2em;">:prep-without (a / and</p> <p style="padding-left: 4em;">:op1 (y / you)</p> <p style="padding-left: 4em;">:op2 (m / music</p> <p style="padding-left: 6em;">:poss y))))</p>
--	--

The AMR on the left clearly shows the inconsistency that surrounds the annotation process. It urges to ask ourselves: “How come that there are concepts written in both Spanish and English in the same AMR?” To start with, the mixture between English-Spanish concepts and relations is rather confusing. Not to mention that having everything in Spanish would seem much more intuitive for Spanish annotators. To solve this issue, we propose the conversion of roles, reifications, modals, and special words. The idea is basically to define their equivalent roles and concepts in Spanish. Hence, from now onwards, we will apply such conversions. An example of the resulting annotation would be as follows:

(19)

Pero no sé qué hacer sin ti y sin tu música

(c / contrastar

:ARG2 (s / saber :polaridad -

:ARG1 (y / yo)

:ARG2 (h / hacer

:ARG0 y

:ARG1 (a / amr-desconocido)

:prep-sin (y2 / yo

:op1 (t / tú)

:op2 (m / música

:posee t))))

Since the resulting roles and concepts are equivalent to the English ones, these conversions are relatively easy to make, computationally speaking. A list of conversions is available in Appendix 2.

5.1.1 NP Ellipses

Because of the nature of Spanish grammar, nominal ellipsis in Spanish is the order of the day. In most cases, we will find either a conjugated verb and/or a clitic pronoun that indicates person. So, technically, there is no need to add a Noun Phrase (NP) unless we need to clarify the subject.

- (20) *Yo tengo prisa. Tengo cita con el dentista.*
I am in a hurry. (I) have an appointment with the dentist.
- (21) *Carla tiene prisa. Tiene cita con el dentista.*
Carla is in a hurry. (She) has an appointment with the dentist.
- (22) *Ellos tienen prisa. Tienen cita con el dentista.*
They are in a hurry. (They) have an appointment with the dentist.

In the first example above, *tengo* indicates the first person singular of the verb *tener* (“to have”). If we consider the second sentence: *Tengo cita con el dentista*, we could perfectly annotate it using the available AMR guidelines.

- (23)
- | | |
|-----------------------------------|---|
| Tengo cita con el dentista | I have an appointment with the dentist |
| (t / tener | (h / have-03 |
| :ARG0 (y / yo) | :ARG0 (i / I) |
| :ARG1 (c / cita | :ARG1 (a / appointment-02 |
| :prep-con (d / dentista))) | :ARG0 i |
| | :ARG1 (d / dentist))) |

The problem comes when we try to annotate the second or the third examples. *Tiene* indicates the third person singular of the verb *tener* whereas *tienen* indicates the third person plural of the same verb. In the second example, in English, we know that we are talking about a female entity. In Spanish, however, we do not know whether is a he or a she. Same happens in the third example since they could be translated as *ellos* (masculine or generic masculine) or *ellas* (feminine). In any of those examples, it is not clear who has the appointment.

- (24)
- | | |
|-----------------------------------|------------------------------------|
| Tiene cita con el dentista | Tienen cita con el dentista |
| (t / tener | (t / tener |
| :ARG0 X | :ARG0 X |
| :ARG1 (c / cita | :ARG1 (c / cita |
| :prep-con (d / dentista))) | :prep-con (d / dentista))) |

Whenever there is a nominal ellipsis, we propose to use a concept *ente* (“being”) that is mapped to a non-core role *:sinnombre* (“:nameless”) and followed by a concept of the same name. This decision is based on the idea that not including an entity¹² that performs an action in the annotation – when the sentence evidently states that there is one – would lead to an inaccurate semantic meaning representation. Below is an example of the proposed annotation of a third person nominal ellipsis.

(25)

Tiene(n) cita con el dentista ✗

(t / tener
:ARG1 (c / cita
:prep-con (d / dentista)))

Tiene(n) cita con el dentista ✓

(t / tener
:ARG0 (e / ente
:sinnombre (s / sinnombre))
:ARG1 (c / cita
:prep-con (d / dentista)))

5.1.2 Third Person Possessive Pronouns

When it comes to annotate ellipses in possessive NPs, another issue arises. This has to do with the third person singular and plural possessive pronouns *su* and *sus*. The problem, once again, is not knowing much about the possessor. For instance, *su* could be translated as “his,” “her,” “its,” “they,” or even “your” (formal).

(26)

Su casa es grande

(g / grande
:campo (c / casa
:posee X))

Your house is big

(b / big
:domain (h / house
:poss (y / you)))

His house is big

(b / big
:domain (h / house
:poss (h2 / he)))

Her house is big

(b / big
:domain (h / house
:poss (s / she)))

Its house is big

(b / big
:domain (h / house
:poss (i / it)))

Their house is big

(b / big
:domain (h / house
:poss (t / they)))

¹² When we use the word “entity” or the concept *ente* (“being”), we refer to real or imaginary, concrete or abstract, animate or inanimate beings.

Thus, there is a need to include an external role as the current AMR guidelines leave so much room for ambiguity – and misunderstanding – regarding these types of constructions.

When there is ellipsis in possessive NPs, we also annotate an entity with the concept *ente*. This, however, is tagged with the non-core role *:sinespecificar* (“:unspecified”). The latter is followed by the possessive pronoun in singular form.

(27)

Su casa es grande
 (g / grande
 :campo (c / casa
 :posee (e / ente
 :sinespecificar (s / su))))

If *su* refers to an omitted entity that has been previously mentioned – and, therefore, annotated – in the sentence, then the possessor must be such entity rather than *su*. For instance, consider the following AMR, meaning “She put on her wedding dress.” Because we have already annotated the concept (*e / ente :sinespecificar (s / se)*), we annotate the possessor with the same variable *e*. Otherwise, we would be indicating that *se* and *su* refer to two different entities – which is not the case.

(28)

Se vistió con su vestido de novia ✗

(v / vestir
 :ARG0 (e / ente
 :sinespecificar (s / se))
 :ARG1 e
 :prep-con (v / vestido
 :posee (e2 / ente
 :sinespecificar (s / su))
 :mod (n / nupcial)))

Se vistió con su vestido de novia ✓

(v / vestir
 :ARG0 (e / ente
 :sinespecificar (s / se))
 :ARG1 e
 :prep-con (v / vestido
 :posee e
 :mod (n / nupcial)))

5.1.3 Third Person Clitic Pronouns

A similar complication occurs when attempting to annotate third person clitic pronouns. Consider the following examples in which *lo*, as a clitic pronoun, fails to provide much information of the entity that it refers to because it is semantically underspecified. As far as we know, it could be equally associated to an object than to a pet or even a male human.

- (29) *Quiero mi dinero de vuelta. Mándamelo ahora.*
I want my money back. Send it to me now.
- (30) *Quiero a mi hijo de vuelta. Mándamelo ahora.*
I want my son back. Send him to me now.

Even if we know that *lo* is an entity that is mentioned in the sentence, we do not really know who or what it is as long as there is no context to take into account – which is the issue that we encounter when we are annotating AMRs. In the example below, for instance, we could break down the word *mándamelo* into three components: *manda* + *me* + *lo*. *Manda* being the second person singular of the verb *mandar* (“to send”), *me* being an enclitic pronoun that means “to me” and *lo* being another enclitic pronoun that refers to an entity in third person singular. Namely, *manda* marks who the doer of the action of sending is (you), *me* refers to the receiver of what is being sent (me), and *lo* indicates the entity that is being sent (it/him).

- (31)
- | | |
|---|---|
| <p>Mándamelo ahora
(m / mandar
:modo imperativo
:ARG0 (t / tú)
:ARG1 X
:ARG2 (y / yo)
:tiempo (a / ahora))</p> | <p>Send it to me now
(s / send-01
:mode imperative
:ARG0 (y / you)
:ARG1 (i / it)
:ARG2 (i2 / I)
:time (n / now))</p> |
| | <p>Send him to me now
(s / send-01
:mode imperative
:ARG0 (y / you)
:ARG1 (h / he)
:ARG2 (i2 / I)
:time (n / now))</p> |

Once again, following the idea that ignoring a present entity would lead to an inaccurate semantic meaning representation, we generally annotate enclitic and proclitic pronouns in the same way that we annotate third-person possessive NPs. To set an example:

- (32)
- | | |
|---|--|
| <p>Mándamelo ahora
(m / mandar
:modo imperativo
:ARG0 (t / tú)
:ARG1 (e / ente
:sinespecificar (l / lo))
:ARG2 (y / yo)
:tiempo (a / ahora))</p> | <p>Mándaselo ahora
(m / mandar
:modo imperativo
:ARG0 (t / tú)
:ARG1 (e / ente
:sinespecificar (l / lo))
:ARG2 (e / ente
:sinespecificar (s / se))
:tiempo (a / ahora))</p> |
|---|--|

The AMR on the right shows that normally it is fairly easy to annotate clitic pronouns, including *se*. Unfortunately, this is not always the case since *se* itself can be used for different purposes. In the following section, we discuss how to treat different types of *se*.

5.1.4 *Se* Usage

With eleven uses (Table 2), *se* is quite possibly the most versatile pronoun in the Spanish language.

Type of <i>se</i>	Description	Examples
<i>Se</i> as an indirect object	Substitutes <i>le</i> or <i>les</i> when they appear together with direct object pronouns	<i>Se lo regala</i> ("(S)he/they gave it to (someone)")
Reflexive <i>se</i>	The subject and the object have the same referent	<i>Se ducha</i> ("(S)he showers")
Reciprocal <i>se</i>	Expresses mutual action or relationship	<i>Se miran</i> ("They look at each other")
<i>Se</i> as a component of a pronominal verb	Acts as a morpheme and has no syntactic function	<i>Se quejan</i> ("They complain")
Intransitive <i>se</i> with a voluntary subject	Makes the verb intransitive; the subject performs the action	<i>Se traslada</i> ("(S)he relocates")
Intransitive <i>se</i> with an involuntary subject	Makes the verb intransitive; the subject experiences or receives the action	<i>Se alegraron</i> ("They rejoiced")
Emphatic <i>se</i>	Intensifies the meaning of a verb	<i>Se bebió una botella de vino</i> ("(S)he drank a bottle of wine")
Accidental <i>se</i>	Corresponds to non-intentional actions	<i>Se les derrama el café</i> ("They spill the coffee over")
Aspectual <i>se</i>	Adds aspectual meaning	<i>Se duerme</i> ("(S)he falls asleep")
Passive <i>se</i>	Places the influence on the action rather than on the doer/actor	<i>Se diseñó la catedral</i> ("The cathedral was designed")
Impersonal <i>se</i>	Makes no specific reference (refers to people in general)	<i>En este restaurante se come bien</i> ("You eat well in this restaurant")

Table 2: *Se* usage with examples according to Lozano (2005)

In most cases, the problem that we encounter is not knowing who or what is the direct or the indirect object. For example:

- (33) *Se rompió anoche.*
 (It) broke last night.
- (34) *Se lava los dientes.*
 ((S)he) brushes her/his teeth.

Here, we simply do not know who brushes their teeth or what is broken. Only context could tell us who or what performs – or receives – the action but AMR focuses on the analysis of single sentences. At this point, we are left with no other choice but to include something to cover semantic underspecification. To do that, we follow the aforementioned clitic annotation procedure.

(35)

Mándaselo ahora

(m / mandar
:modo imperativo
:ARG0 (t / tú)
:ARG1 (e / ente
:sinespecificar (l / lo))
:ARG2 (e2 / ente
:sinespecificar (s / se))
:tiempo (a / ahora))

Send it to her now

(s / send-01
:mode imperative
:ARG0 (y / you)
:ARG1 (i / it)
:ARG2 (s2 / she)
:time (n / now))

Send it to him now

(s / send-01
:mode imperative
:ARG0 (y / you)
:ARG1 (i / it)
:ARG2 (h / he)
:time (n / now))

Send it to them now

(s / send-01
:mode imperative
:ARG0 (y / you)
:ARG1 (i / it)
:ARG2 (t / they)
:time (n / now))

Doubts also come to mind when it comes to annotate reciprocal (Example 36) and reflexive (Examples 37 and 38) sentences in third person, or when we attempt to annotate impersonal or passive *se* (Examples 39 and 40, respectively). Therefore, we need something to make up for the lack of information.

(36) *Se gustan.*

They like each other.

(37) *Se gustan así mismos.*

They like themselves.

(38) *Se gusta así mismo.*

He likes himself.

(39) *Aquí se habla español, inglés y alemán.*

Spanish, English, and German are spoken here.

(40) *Se vende casa rural.*

Rural house for sale.

Thus, when *se* is reflexive, we use reentrancy – as we can see in the AMRs below.

(41)

Se gusta

(g / gustar
:ARG0 (e / ente
:sinespecificar (s / se))
:ARG1 e)

(S)he likes her/himself

(l / like-01
:ARG0 (X / she | he)
:ARG1 X)

Se gusta a sí misma

(g / gustar
 :ARG0 (e / ella)
 :ARG1 e)

She likes herself

(1 / like-01
 :ARG0 (s / she)
 :ARG1 s)

Se gustan a sí mismos

(g / gustar
 :ARG0 (e / ellos)
 :ARG1 e)

They like themselves

(1 / like-01
 :ARG0 (t / they)
 :ARG1 t)

Gustarse a sí/uno mismo

(g / gustar
 :ARG0 (u / uno)
 :ARG1 u)

To like oneself

(1 / like-01
 :ARG0 (o / one)
 :ARG1 o)

The criterion to annotate reciprocal *se* is slightly different. We also use reentrancy, but we add a concept named *se-recíproco* (“reciprocal-*se*”).

(42)

Se gustan

(g / gustar
 :ARG0 (e / ente
 :sinespecificar (s / se-recíproco))
 :ARG1 e)

They like each other

(1 / like-01
 :ARG1 (e / each
 :mod (o / other)))

Because representing reciprocity in Spanish AMRs is not only a problem that has to do with omitted entities, below we propose a way to annotate reciprocal *se* when it addresses an entity that is present in the sentence.

(43)

Ellos se gustan

(g / gustar
 :ARG1 (u / uno
 :mod (a / al-otro)))

Then, if we consider impersonal *se* and passive *se*, a different issue applies. Look at examples 39 and 40. Once again, we do not know who or what performs a given action. However, in these cases there is no subject that is explicitly stated in the sentence. Therefore, there is no need for us to include one.

(44)

Aquí se habla español, inglés y alemán

(h / hablar
 :ubicación (a / aquí)
 :ARG3 (y / y
 :op1 (e / español)
 :op2 (i / inglés)
 :op3 (a2 / alemán)))

Se vende casa rural

(v / vender
 :ARG1 (c / casa
 :mod (r / rural)))

5.1.5 Gender

For grammatical purposes, Spanish nouns can be either masculine or feminine. Even if the grammatical gender in Spanish often corresponds to the sex of the object when referring to animate beings, it must not be equated with sex.

When a noun describes an animate being, there are many cases in which we use the same word form to refer to entities of both sexes as English speakers do (Example 45) together with a precedent word that indicates the grammatical gender.

- (45) *Ella es una profesional / Él es un profesional.*
(S)he is a professional.

Exceptionally, some words are always the same gender, whether they refer to a male or a female being (Example 46). That is to say, there are Spanish nouns which are not gender-specific.

- (46) *Ella/él es una persona maravillosa.*
(S)he is a wonderful person.

However, usually there are pairs of words that correspond to the biological distinction between the sexes, these pairs either vary in their ending (Examples 47 and 48) or are completely different (Example 49).

- (47) *Ella es mi amiga / Él es mi amigo.*
(S)he is my friend.
- (48) *Ella es mi novia / Él es mi novio.*
She is my girlfriend / He is my boyfriend.
- (49) *Ella es mi nuera / Él es mi yerno.*
She is my daughter-in-law / He is my son-in-law.

In addition, when a noun describes an inanimate object, nouns tend to be either masculine (Example 50), or feminine (Example 51).

- (50) *El sillón.*
The armchair.
- (51) *La silla.*
The chair.

Yet, there is a group of these nouns that is ambiguous in the sense that they correspond to both genders (Example 52).

- (52) *El/la mar.*
The sea.

Gender is vital to Spanish grammar because, in some cases, the gender of a given noun can affect the meaning of a word and, in turn, the meaning of a sentence. Knowing the proper usage can mean the difference between finding a cure (*la cura*) and finding a priest (*el cura*). Thus, even though AMR omits articles and does not represent certain inflectional morphology, we consider that it is important to annotate gender whenever it is needed.

Imagine if a system performed the AMR-to-Spanish conversion of the following annotation:

- (53)
- Comemos la papa**
- (c / comer
:ARG0 (n / nosotros)
:ARG1 (p / papa))

If that was the case, then some possible translations would include:

- (54) *Comemos las papas.*
We eat the potatoes. ✓
- (55) *Comeremos la papa.*
We will eat the potato. ✓
- (56) *Comeremos al papa.*
We will eat the pope. ✗

Because our purpose is to provide an accurate semantic representation, we propose to include the gender roles *:fem* and *:masc* to the AMR guidelines in order to indicate whether a noun is feminine or masculine. The idea would be to always annotate the word in either its masculine

form or its unique form and to use a given gender role whenever applicable. Consider the following example:

(57)

Los niños y las niñas juegan al fútbol en el parque

Boys and girls play soccer in the park

(j / jugar

(p / play-01

:ARG0 (y / y
:op1 (n / niño :masc n)
:op2 (n2 / niño :fem n2))
:ARG1 (f / fútbol)
:ubicación (p / parque))

:ARG0 (a / and
:op1 (b / boy)
:op2 (g / girl))
:ARG1 (s / soccer)
:location (p2 / park))

As shown above, the nouns *niños* and *niñas* are annotated as their singular, masculine form *niño* together with a role that specifies their given gender. At the same time, the nouns *fútbol* and *parque* are not accompanied by any relation. This is simply because both words have only one gender: masculine. If we know that there is no room for ambiguity, including a role would just be redundant. In the table below we describe the annotation criteria adopted for using gender roles.

Type of noun	Subtype	Examples	Annotation procedure
Have the same form for both genders	A word that comes before determines the gender	Ese/esa violinista; el/la psiquiatra	(f / form :masc f :fem f)
	There is only one gender of the preceding word and it refers to both sexes	El tiburón; la víctima	(f / form)
	The gender of the preceding word does not add meaning	El/la vodka; el/la calor	(f / form)
Vary in form according to the gender	Same root with two different endings	Doctor/doctora, poeta/poetisa	(f / masculine-form :masc f :fem f)
	Two completely different words	Caballero/dama, hombre/mujer	(f / form)

Table 3: Rules for gender annotation of Spanish entities¹³

Because an adjective describes a noun or a pronoun, its gender should conform to the noun – or pronoun – involved. Thus, we never annotate a gender role with a variable that corresponds to an adjective. This role, on the contrary, is always connected to the word it describes. Consider the following example. The idea would be to adopt a sequential conversion process by which the word *amigo* is transformed into *amiga* because of the role *:fem* represents the inherent property of the noun and, in turn, the word *guapo* into *guapa* because the adjective is assigned via the gender of the noun it qualifies.

13 Generic masculine nouns and pronouns should be annotated in their masculine form without any non-core role whereas personal pronouns keep their original form as they are considered to be unique in meaning.

(58)

Tu amiga es muy guapa

(g / guapo
:grado (m / muy)
:campo (a / amigo)
:fem a
:posee (t / tú))

Your friend is very pretty

(p / pretty
:degree (v / very)
:domain (f / friend)
:poss (y / you))

There are some occasions in which a third-person NP is omitted. In such cases, adjectives can help us to detect whether an entity is masculine or feminine. But still, the corresponding role would be assigned to the omitted concept – as can be seen in the next example:

(59)

Es muy guapa

(g / guapo
:grado (m / muy)
:campo (e / ente
:sinnombre (s / sinnombre)
:fem e)))

(She) is very pretty

(p / pretty
:degree (v / very)
:domain (s / she))

5.1.6 Verbal Periphrases and Locutions

Verbal locutions and verbal periphrases are two types of verbal constructions in Spanish. The first one (Example 60) is a fixed combination of two or more words whose meaning is not the sum of the meaning of all their parts. The second one (Example 61) is made of two verb forms: a conjugated form of an auxiliary verb followed by the impersonal form of the main verb – which can be either infinitive, gerund, or participle – that mostly determines the meaning of the verbal lexical unit. It is worth mentioning that there is a set of verbs that are prone to be auxiliary verbs, some of them may be even equivalent to a few English modals or semimodals.

(60) *Echar de menos (algo/a alguien)*

Miss (something/someone)

(61) *Dejar de fumar / Seguir fumando*

Quit smoking / Keep smoking

As long as the verbal periphrases do not fall into the category of modal-like expressions, the AMR annotation should be rather simple. The auxiliary verb would be annotated as the

higher – but not necessarily the top – node, whereas the main verb would act as a core role.

(62)

Dejaste de fumar

(d / dejar
:ARG0 (t / tú)
:ARG1 (f / fumar))

Sigo fumando

(s / seguir
:ARG0 (y / yo)
:ARG1 (f / fumar))

The annotation of modality follows a similar approach. The only difference is that, just as English AMR annotates syntactic modals, Spanish AMR represents these auxiliary verbs with the equivalent modal concepts. An example is included below:

(63)

Debes fumar

(r / recomendar
:ARG1 (f / fumar
:ARG0 t)
:ARG2 (t / tú))

You should smoke

(r / recommend-01
:ARG1 (s / smoke-02
:ARG0 t)
:ARG2 (y / you))

Annotating verbal locutions, on the other hand, is a different story since there is no compositionality. Because of that, we need to be creative when it comes to include the semantic meaning representations of these. To do so, we propose the sequential procedure illustrated in Figure 10.

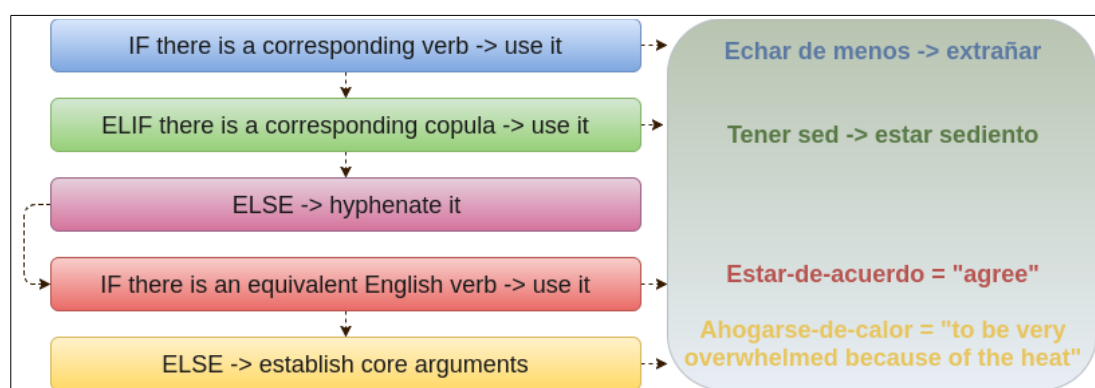


Figure 10: Step-by-step procedure on how to annotate verbal locutions

There are occasions in which a verbal locution has a verb-form synonym. If that is the case, then we need to use the corresponding verb to annotate our AMRs. If this is not the case, but it is still possible to convert it into a copulative construction, then that is what we do. Needless to say, we would remove the copulative verb and we would annotate the resulting construction

with *:campo* (“:domain”). Then, if we cannot rephrase a verbal locution in any of these ways, we hyphenate it. Whenever there is an equivalent English verb, we borrow the core argument of its particular frameset. If no equivalence exists, core arguments should be defined. Examples are presented below.

(64)

Te echo de menos

(e / extrañar
:ARG0 (y / yo)
:ARG1 (t / tú))

Tengo sed

(s / sediento
:campo (y / yo))

Estoy de acuerdo

(e / estar-de-acuerdo
:ARG0 (y / yo))

I miss you

(m / miss-01
:ARG0 (i / I)
:ARG1 (t / tú))

I am thirsty

(t / thirst-01
:ARG0 (i / I))

I agree

(a / agree-01
:ARG0 (i / I))

5.1.7 Double Negatives

As shown in the next pair of AMRs, unlike standard English, Spanish does allow double negatives:

(65)

No sabes nada, Jon Nieve

(d / decir
:ARG0 (y / yo)
:ARG1 (s / saber
:ARG0 p
:ARG1 (n / nada))
:ARG2 (p / persona
:nombre (n2 / nombre
:op1 “Jon”
:op2 “Nieve”)))

You know nothing, Jon Snow

(s / say-01
:ARG0 (i / I)
:ARG1 (k / know-01
:ARG0 p
:ARG1 (n / nothing))
:ARG2 (p / person
:wiki “Jon_Snow_(character)”
:name (n2 / name
:op1 “Jon”
:op2 “Snow”)))

Even if single negatives are allowed, such as in examples 66 and 67, there are certain constructions that make no sense as a single negative. This is the case of *saber nada* (“know nothing”) – which asks for a preceding word that expresses negation like adverbs *no* (“not”) or *nunca* (“never”). For this reason, we believe double negatives should not be put aside.

(66) *Nada es lo que parece.*
Nothing is what it seems.

(67) *No me gusta.*
I don't like it.

The solution that we came up with is as simple as adding a polarity role:

(68)

No sabes nada, Jon Nieve

(d / decir
:ARG0 (y / yo)
:ARG1 (s / saber :polaridad -
:ARG0 p
:ARG1 (n / nada))
:ARG2 (p / persona
:nombre (n2 / nombre
:op1 "Jon"
:op2 "Nieve"))))

If we swapt the word “no” for “nunca,” then the resulting AMR would be as the one below. In such case, it is easy to see that they bear the same structure in terms of the annotation of the verb “know-01.” This is because *nada* can mean either “nothing” or “anything” and *nunca* can mean “never” or “ever.”

(69)

Nunca sabes nada, Jon Nieve	You never know anything, Jon Snow
(d / decir	(s / say-01
:ARG0 (y / yo)	:ARG0 (i / I)
:ARG1 (s / saber :polaridad -	:ARG1 (k / know-01 :polarity -
:ARG0 p	:ARG0 p
:ARG1 (n / nada)	:ARG1 (a / anything)
:tiempo (n / nunca))	:time (e / ever))
:ARG2 (p / persona	:ARG2 (p / person
:nombre (n2 / nombre	:wiki "Jon_Snow_(character)"
:op1 "Jon"	:name (n / name
:op2 "Nieve")))	:op1 "Jon"
	:op2 "Snow")))

5.2 Quantitative Results

In this section, we present the results of our quantitative analysis of the degree of structural similarity that pairs of English-Spanish AMR graphs display. In other words, what we have called Degree of AMR Similarity (DAS). At the same time, we report on the correlation

between such similarity and the Degree of Syntax Similarity (DSS) that was introduced in the previous chapter. We also provide an interpretation of these results.

5.2.1 Results

As we have explained earlier, we counted the number of parallel AMR pairs that display a Substantial Structural Similarity as well as the number of the pairs that display a Partial Structural Similarity that either share the same top node (PSS 1) or not (PSS 2). Whenever a pair showed a PSS, we marked the differences regarding framesets, non-role concepts, and non-core roles. Moreover, we counted the number of the pairs that show a Total Structural Difference (TSD). The results are shown in Table 4.

Degree of AMR Similarity (DAS)	26% show a Substantial Structural Similarity (SSS)
23 different framesets	44% show a Partial Structural Similarity with the same top node (PSS 1)
26 different non-role concepts	
43 different non-core roles	
	12% show a Partial Structural Similarity with different top node (PSS 2)
	18% show a Total Structural Difference (TSD)

Table 4: Degrees of AMR Similarity (expressed in percentages)

Referring to Table 4, we can see that pairs of English-Spanish AMRs seem to be similar to a certain extent in the majority of the cases (82%¹⁴). Although, in around 68.3%¹⁵ of these cases, they seem to be simply partially similar. In addition, from these partly similar AMR pairs, it can be seen that around 78.5%¹⁶ of these pairs share the same top node. At the same time, it was found that there were 23 framesets, 26 non-role concepts, and 43 non-core roles that were different in the AMR pairs that show a Partial Structural Similarity.

Additionally, we thought it would be interesting to study any possible correlation between the two types of degrees that we have considered: the degrees of AMR and syntax similarity. In order to do that, first we counted the number of times that bilingual sentence pairs present some some sort of similarity – or not – in structure, as seen in Table 5.

14 SSS + PSS 1 + PSS 2

15 ((PSS 1 + PSS 2) / (SSS + PSS 1 + PSS 2)) * 100

16 (PSS 1 / (PSS 1 + PSS 2)) * 100

Degree of Syntax Similarity (DSS)	66% are of degree 1 (extremely or very similar)
	26% are of degree 2 (somewhat or slightly similar)
	8% are of degree 3 (not at all similar)

Table 5: Degrees of Syntax Similarity (expressed in percentages)

From Table 5, it has been found that 92%¹⁷ of the sentence pairs were similar to some extent. Out of these sentences, 71.7%¹⁸ are of degree 1. Surprisingly, the number of sentence pairs of degree 3 (expressed in percentages) is lower than the number of AMR pairs that show to be different in Table 4.

Then, we calculated the percentage of times that a type of DAS was found to correspond with a particular type of Degree of Syntax Similarity (DSS). The results can be found in Figure 11. It is worth noting that the percentages that go together with the types of DAS represent the number of pairs of AMRs that correspond to that particular type.

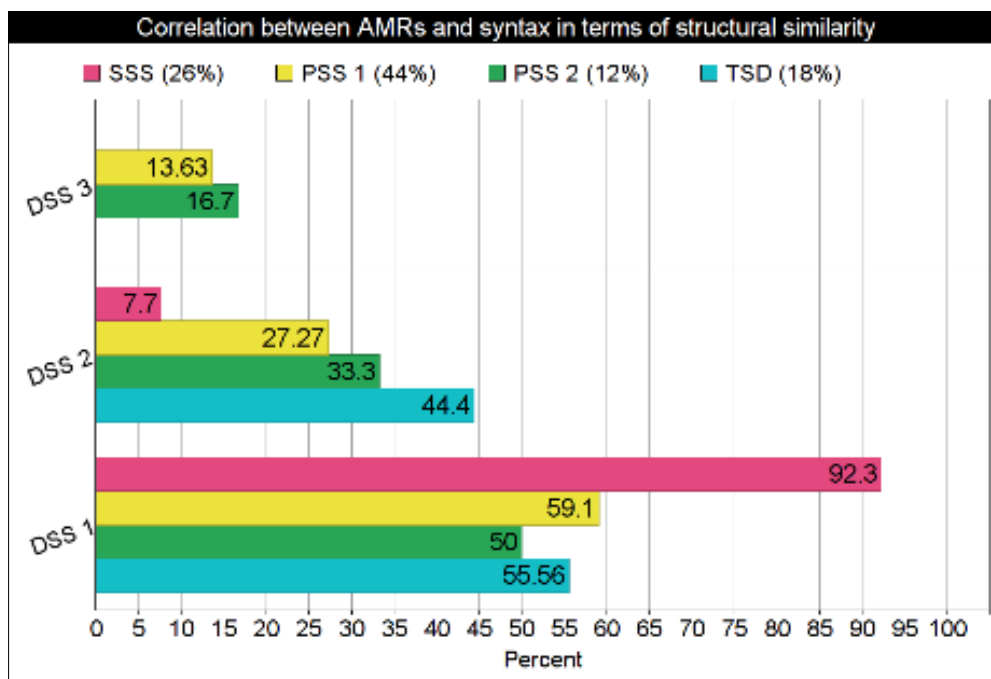


Figure 11: Correlation between DAS and DSS (expressed in percentages)

As is evident from Figure 11, except for the pairs of AMRs that display a Total Structural Difference (TSD), the more similar the structure of two AMRs is, the closer the structure of

17 DSS 1 + DSS 2

18 $(DSS\ 1 / (DSS\ 1 + DSS\ 2)) * 100$

its corresponding parallel sentences is presumed to be. Thus, the results show that most AMR pairs that display a Substantial Structural Similarity (SSS) are aligned to sentence pairs that are, at least, very similar whereas the remaining AMR pairs are aligned to pairs of sentences that show to be somewhat or slightly similar. Furthermore, those pairs of AMRs that show a Partial Structural Similarity (PSS) and share the same top node correspond to sentence pairs of DSS 1 in 59.1% of the cases, to sentence pairs of DSS 2 in 27.27% of the cases, and to sentence pairs of DSS 3 in 13.63% of the cases. The pairs that are partially similar but do not share the same top node correspond to sentence pairs of DSS 1 in half of the cases, to sentence pairs of DSS 2 in 33.3% of the cases, and to DSS 3 in 13.63% of the cases. Interestingly, more than half of the AMR pairs that show a TSD correspond to parallel sentences that are quite similar in structure. And, what is more, none of them correspond to sentence pairs that are not at all similar.

5.2.2 Conclusions

The results from Table 4 indicate that most of the bilingual pairs of AMRs seem to be similar, even if they merely display a partial similarity. But still, the fact that just 26% of these pairs show an SSS and that 18% of them show a TSD appear to suggest that there is still room for improvement.

Differences in the 28 AMR pairs that are categorized as partially similar revealed that non-core roles were the most frequent mismatch type. The visual comparison of the differences listed above led to the conclusion that these can differ when an event is annotated through the use of a concept or relation of the same kind, through the use of a concept or relation of a different kind, or through no use of any concept or relation at all.

For instance, consider the following pair of AMRs, where the non-role concept (*p / persona :mod (t2 / todo)*) of the Spanish AMR is simply annotated as “(e / everyone)” in the English AMR. Also the non-core role *:mod* does not appear in the English AMR. If the role *:polaridad* - of the Spanish AMR was reified as “have-polarity-91” in the English AMR, then that would have been considered a difference regarding the use of a concept or relation of a different kind.

(70)

No todos han tenido un amigo.

(t / tener :polaridad -
:ARG0 (p / persona
:mod (t2 / todo))
:ARG1 (p2 / persona
:ARG0-de (t3 / tener-rol-rel-91
:ARG1 p
:ARG2 (a / amigo))))

Not everyone has had a friend.

(h / have-03 :polarity -
:ARG0 (e / everyone)
:ARG1 (p / person
:ARG0-of (h2 / have-rel-role-91
:ARG1 e
:ARG2 (f / friend))))

All these differences in the annotation of parallel texts, together with the fact that only a small number of the pairs of bilingual AMRs was substantially similar, suggest that the level of compatibility between English and Spanish AMRs is not very high. A more detailed inspection of differences in the structure of parallel AMRs suggests that there are a few reasons for them to appear in the annotation.

On the one hand, translation and idiomacity play an important role. We performed either a semantic or an idiomatic translation rather than a word-for-word one. And, based on the correlation between the structure of parallel AMRs and the structure of its corresponding sentences displayed in Figure 11, we can observe how the number of AMR pairs that are similar decreases depending on their DSS. The more similar a pair of sentences is, the more similar their AMRs are.

On the other hand, the annotation scheme is also involved. Because AMR annotates sentences independently of context, there are occasions in which a sentence can be represented in more than one way. Consider the following example:

(71)

¿De qué planeta eres?

(s / ser-de-91
:ARG1 (t / tú)
:ARG2 (p / planeta
:campo (a / amr-desconocido)))

Which is your planet?

(p / planet
:poss (y / you)
:domain (a / amr-unknown))

Here we find two different meaning representations for two sentences that are constructed differently. Yet, semantically speaking, they are the same. Because we have translated the sentence “Which is your planet?” within a context, we know that this question does not ask about the planet that you possess, but about your planet of origin. Thus, the sentence *¿De qué planeta eres?* literally means “Which planet are you from?” Because AMR does not take

context into account, the representation of the question “Which is your planet?” is quite literal. Even if the reified concept “be-from-91” is not used, this representation is still acceptable. The problem here would be ambiguity.

Also, there are times in which certain rules cannot apply for both languages. For instance, in some cases in which inverse roles are used to represent “-er” nouns. E.g. The word “lamplighter” is represented as a person that light lamps, whereas the Spanish word *farolero* is represented with a concept of the same name simply because the meaning of such noun is significantly different from the verbal form *farolear* (“to brag”).

(72)

Lamplighter	Farolero
(p / person :ARG0-of (l / light-04 :ARG1 (l3 / lamp)))	(f / farolero)

Another instance includes some cases in which a single term in a language corresponds to more than one word in the other (“businessman” vs *hombre de negocios*), although there are other cases in which this difference does not matter (“little prince” vs *principito*).

(73)

Businessman	Hombre de negocios
(b / businessman)	(h / hombre :mod (n / negocio))
Little prince	Principito
(p / prince :mod (l / little))	(p / príncipe :mod (p2 / pequeño))

Then, it is also worth noting that the changes that we proposed to expand the guidelines also present a difference in the annotation of English-Spanish AMRs. But, we believe that the newly added information needed to be represented.

(74)

Era un monarca absoluto	He was an absolute monarch
(m / monarca :masc m :campo (e / entidad :sinnombre (s / sinnombre)) :mod (a / absoluto))	(m / monarch :mod (a / absolute) :domain (h / he))

From Figure 11, it can be inferred that the similarity of structure between a pair of sentences seems to be correlated with the similarity between its corresponding parallel AMRs for the most part. However, the fact that more than half of the AMRs that show a Total Structural Difference are of DSS 1 and that the rest are of DSS 2 suggests that there is not much correlation between the aforementioned degrees. This figure also shows that the correlation of bilingual AMR pairs that display a PSS with its corresponding sentences do not vary much whether the PSS is of one type or another. But, on the bright side, pairs of AMRs show to be structurally similar – or at least to be similar to some extent – in spite of showing major differences in the structures of their corresponding sentences.

More observations of our results are presented in the following chapter.

6 DISCUSSION

In this chapter, we discuss both the qualitative and quantitative results reported in the previous chapter. To start with, we answer our research questions. Then, we proceed to evaluate our approach to annotate AMRs and we compare it with an alternative approach. Finally, we outline open issues and limitations.

6.1 Research Questions

This research project is the first step towards the construction of a sizable sembank of Spanish sentences paired with their sentential, logical meanings. Our main goal was to study how to create Spanish AMRs. To address this objective, we examined if it was possible to use the current guidelines to annotate Spanish AMRs through trial and error, we designed an extension of the annotation scheme so that it could cover Abstract Meaning Representations for Spanish, and we performed a visual comparison to evaluate the similarity between English and Spanish AMRs. Thus, we proceed to answer our research questions.

- Is it possible to follow the current guidelines to annotate Spanish AMRs? And if not, how can the guidelines be refined in order to annotate Spanish AMRs? Also, what resources do we need to carry out such task?

As we have seen, the current guidelines, as they are, fail to represent Spanish semantic representations fully. This is no surprise, since it is clearly stated in the guidelines that AMR is not an interlingua. We have demonstrated that it is, in fact, possible to adjust the guidelines accordingly to cover their lack of certain meaning aspects in Spanish that cannot be ignored. To annotate Spanish AMRs, we need the guidelines to be adapted for this task, and we need an editor that is connected to a refined and updated version of AncoraNet.

- How similar are English and Spanish AMRs?

Based on our results, we deduce that English and Spanish AMRs are similar to some extent in most of the cases. Even if parallel AMRs are only completely different in 18% of the cases that we have studied, about 68.3% of the remaining 82% differ in some way or another. These results lead us to think that there are cases in which representations cannot be unified, in spite of changing the guidelines. We think, however, that some conversion or equivalence rules could be applied in the future to make up for these differences. But further studies need to be performed to address this subject. Below are two examples.

(75) (g / girl) == (n / niño :fem n)

(76) (b / businesswoman) == (m / mujer :mod (n / negocio))

- What can be learned from the gathered information for future annotation efforts?

Although a substantial amount of work remains to be done, the information that we have obtained serves as the foundation for future work. Because of this study, we now know what is needed to take the next step in this ongoing effort to build a Spanish semantic bank. In short, based on our work, we know how to cope with linguistic phenomena that did not have a way to be represented before. And, what is more, based on the limitations faced during the annotation process, we know the resources that are needed to achieve this goal.

6.2. Comparison with Other Works

With the exception of the online AMR parser¹⁹ that outputs the NLPwin Logical Form (Vanderwende, 2015) to AMR conversion of Vanderwende et al. (2015), this is the only publicly available study of Abstract Meaning Representation for Spanish. Because neither their gold AMR annotations in Spanish are available to the public nor we have gold annotations, in this work we did not aim to compare both methods in-depth. We did not even consider the accuracy of the system. Instead, we simply paid attention to how they represent

19 Available at: <https://www.microsoft.com/en-us/research/project/msr-splat/>

certain linguistic phenomena. However, we still think that making a fair comparison between the two styles is an interesting topic for future research.

With that said, at first glance, we noticed that the extension of the annotation scheme that we propose seems to be more complete than the annotation decisions taken by their parser. For instance, in the first sentence that includes a clitic construction, MSR SPLAT represents *lo* as (*x* / “*él*”).

- (77) *El reptil se volteó, quitándoselo de encima.*
The crocodile rolled over, throwing it off.

MSR SPLAT

(v / voltear
:ARG (r / reptil)
:manner (quitar
:ARG0 r
:ARG1 (x / “él”)
:prep-de (e / encima)))

Our extension

(v / voltear
:ARG (r / reptil)
:manera (quitar
:ARG0 r
:ARG1 (e / entidad
:sinespecificar (l / lo))
:prep-de (e / encima)))

According to the Royal Spanish Academy (RAE)²⁰, when the third person personal pronouns *él*, *ella*, *ellos*, and *ellas* function as a subject, they can only refer to people. Otherwise, no personal pronoun is explicitly present in a sentence.

- (78) *He leído tus últimos informes. Enhorabuena: son claros y ofrecen numerosos datos.* ✓
He leído tus últimos informes. Enhorabuena: ellos son claros y ofrecen numerosos datos. ✗

In that sense, our extension provides a more accurate representation by including the clitic pronoun *lo* since it cannot only refer to people but it can also refer to other beings, things, and concepts.

Another example has to do with the treatment of NP ellipses. Consider the AMRs from Example 79.

- (79) *Representaba una serpiente boa que se tragaba a una fiera.*
It depicted a boa constrictor swallowing a wild animal.

20 Available at: <http://lema.rae.es/dpd/srv/search?id=seEVswKc5D6y2K5WFZ>

MSR SPLAT

(r / representar
:ARG0 (p / pron)
:ARG1 (s / serpiente
:mod (b / boa
:ARG1-of (t / tragar)))
:prep-a (f / fiera))

Our extension

(r / representar
:ARG0 (e / entidad
:sinnombre (s / sinnombre))
:ARG1 (s / serpiente
:mod (b / boa)
:ARG0-de (t / tragar
:ARG1 (f / fiera))))

As can be seen, MSR SPLAT uses the concept (*p / pron*) or “pronoun” to annotate an entity that performs the action of depicting a boa constrictor or *serpiente boa*. And this choice seems to contradict the essence of AMR. After all, it is a semantic representation language that represents events – not word classes.

6.3 Limitations and Open Issues

In this section, we discuss the technical limitations of our approach, together with remaining issues in annotating AMRs in Spanish.

Since this is a preliminary study, we did not aim to build a new editor to annotate Spanish AMRs, in the same way that we did not attempt to update – and improve the quality of – the AnCora corpus. Due to the same reason, there was only one annotator and our dataset contained only 50 parallel bilingual AMR-annotated sentences. We understand that these choices may present some limitations and that is why our results should be interpreted with caution.

On a different note, in this thesis we provide solutions to annotate meaning representations in cases where we found sources of disagreement during the annotation procedure of our methodology. Indeed, we cannot ignore the possibility that there might be additional problems that an annotator could encounter in the future. It is also important to point out that we made the decision not to cover wikification. But we believe this should not necessarily present a big problem in the future.

Because we did not deal with every single aspect of meaning that AMR covers, our contribution should be considered as the first step in the development of guidelines for

annotating Spanish AMRs and not otherwise. Further prospective studies involving the use of larger data, the development of new resources, and a large number of annotators are necessary to confirm our hypothesis.

7 CONCLUSIONS AND FUTURE WORK

This chapter reviews our achievements and lists our contributions. Moreover, it outlines a list of possible directions for future work that either emerged during the study or are motivated by weaknesses of the proposed methodology.

7.1 Summary and Contributions

To our knowledge, this study is the first work towards the development of Abstract Meaning Representation for Spanish, for which no corpus currently exists. We first localized the missing aspects of meaning that AMR cannot portray and that should not be ignored. After we identified various linguistic phenomena that needed to be represented, we developed specifications for Spanish AMR and we applied these when annotating a selection of 50 sentences of the Spanish translation of *The Little Prince*. Finally, we compared our annotated AMRs with the English version ones to study their similarity and the viability of a unified AMR. The work reported in this thesis is part of our ongoing research effort to develop a large Spanish sembank. Our main contributions are summarized as follows:

- We demonstrate the possibility of annotating Spanish AMRs. We show that the idea of using AMR to annotate Spanish sentences is viable by means of adjusting the current annotation scheme.
- We design and implement an extension of the AMR guidelines to perform the annotation of Spanish meaning representations. Our implementation and some conversions can be found in Appendices 1 and 2, respectively.
- We introduce new metrics to compare bilingual AMRs manually. We compared pairs of AMRs in terms of both their structural similarity and the correlation between this similarity and the syntax similarity of their corresponding sentences.

7.2 Future Work

There are several lines of research arising from this work which should be pursued in order to build a large sembank of Spanish meanings. These are enumerated below.

- Refinement and update of the AnCora corpus. This corpus not only needs to be refined so that the mappings of senses are more accurate but also needs to be updated so that it is connected to an up-to-date version of the PropBank's inventory. At the same time, we think nominal and adjectival relations should be included as well.
- Development of a tool for Spanish AMR annotation. To facilitate AMR construction, an annotation tool is indispensable. Either an adaptation of the current editor or the creation of a brand new tool from scratch would work. Presumably, it would be connected to words that have PropBank frames.
- Inclusion of wikification. In order to avoid certain differences in reference annotation, wikification could be used to perform Named Entity Disambiguation (NED) – in the same way that it is used in the English version of AMR.
- Construction of a corpus by multiple annotators. We think it would be interesting to annotate *The Little Prince* with Spanish AMRs in order to compare the resulting AMRs to the English and Chinese versions.
- Improvement and automation of the metrics designed. We believe the incorporation of new criteria would lead to even more interesting results. Similarly, the automation of the metrics would make the evaluation more objective.

7.3 Closing Remarks

The broad underlying purpose of this thesis was to explore whether – and if so how – AMR could be applied to annotate Spanish semantic representations. In that sense, we have succeeded but, clearly, a substantial amount of work remains to be done. Our proposal to

extend the AMR guidelines is, without a doubt, the main contribution of our work. We believe it is the initial step to the development of a large semantic bank of Spanish sentences paired with their whole-sentence, logical meanings. We also hope that our work will be of use to those who are interested in using AMR to annotate other languages as well as to those who would like to perform cross-lingual research for different purposes like the unification of AMR or Machine Translation (MT).

REFERENCES

- Abend, O., Rappoport, A. (2013a). Universal Conceptual Cognitive Annotation (UCCA). In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, 228–238. Sofia, Bulgaria: Association for Computational Linguistics.
- Abend, O. and A. Rappoport. (2013b). UCCA: A semantics-based grammatical annotation scheme. In Proceedings of the 10th International Conference on Computational Semantics, 1–11. Potsdam, Germany: Association for Computational Linguistics.
- Artzi, Y., Lee, K. and Zettlemoyer, L. (2015). Broad-coverage CCG semantic parsing with AMR. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 1699–1710. Lisbon, Portugal: Association for Computational Linguistics.
- Banarescu L., Bonial C., Cai S., Georgescu M., Griffitt K., Hermjakob U., Knight K., Koehn P., Palmer M., and Schneider N. (2013). Abstract Meaning Representation for Sembanking. In Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse, 178–186, Sofia, Bulgaria: Association for Computational Linguistics.
- Basile, V., Bos, J., Evang, K. and Venhuizen, N.J. (2012): Developing a large semantically annotated corpus. In Proceedings of the 8th International Conference on Language Resources and Evaluation, 3196–3200. Istanbul, Turkey: European Language Resources Association.
- Bates, M. (1995). Models of natural language understanding. Proceedings of the National Academy of Sciences of the United States of America, 92(22), 9977–9982.
- Blackburn, P. and Bos, J. (2003). Computational semantics. *Theoria*, 18(1), 27–45.
- Bos, J. (2008). Wide-coverage semantic analysis with boxer. In Proceedings of the 2008 Conference on Semantics in Text Processing, 277–286. Venice, Italy: Association for Computational Linguistics.
- Bos, J. (2016). Expressive Power of Abstract Meaning Representations. *Computational Linguistics*, 42(3), 527-535.
- Butler, A. (2010). The Semantics of Grammatical Dependencies. *Current Research in the Semantics/Pragmatics Interface*, 23. Bingley, UK: Emerald.
- Butler, A., & Yoshimoto, K. (2012). Banking Meaning Representations from Treebanks. *Linguistic Issues in Language Technology*, 7(1), 1–22. CSLI Publications.

- Cai, S. and Knight, K. (2013). Smatch: an Evaluation Metric for Semantic Feature Structures. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Vol. 2, 748–752, Sofia, Bulgaria: Association for Computational Linguistics.
- Charniak, E. and Johnson, M. (2005). Coarse-to-fine n-best parsing and MaxEnt discriminative reranking. In Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, 173–180, Ann Harbor, MI, USA: Association for Computational Linguistics.
- Chiang, D., Andreas, J., Bauer, D., Hermann, K. M., Jones, B., Knight, K. (2013). Parsing Graphs with Hyperedge Replacement Grammars. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, 924–932. Sofia, Bulgaria: Association for Computational Linguistics.
- Chomsky, N. (1996). *Janua Linguarum. Series Minor: Studies on Semantics in Generative Grammar* (4). Berlin/Boston, DE: De Gruyter Mouton.
- Copestake, A. (2009). Slacker semantics: Why superficiality, dependency and avoidance of commitment can be the right way to go. In Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, 1–9. Athens, Greece: Association for Computational Linguistics.
- Crystal, D. (2008). *A Dictionary of Linguistics and Phonetics* 6th Edition. New York: Basil Blackwell.
- Curran, J., S. Clark, and J. Bos (2007). Linguistically Motivated Large-Scale NLP with C&C and Boxer. In Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions, 33–36. Prague, Czech Republic: Association for Computational Linguistics.
- Damonte, M., Cohen, S. B. and Satta, G. (2017, January 17). An Incremental Parser for Abstract Meaning Representation. In Proceedings of European Chapter of the Association for Computational Linguistics 2017. Valencia, Spain. Retrieved March 26, 2017, from the arXiv database.
- Davidson, D. (1969). The individuation of events. In N. Rescher (ed.), *Essays in Honor of Carl G. Hempel*. 216–234. Dordrecht, Netherlands: D. Reidel.
- Dixon, R. M. W. (2005). *A Semantic Approach To English Grammar*. New York, USA: Oxford University Press.
- Flanigan, J., Thomson, S., Carbonell, J., Dyer, C. and Smith, N. A. (2014). A discriminative graph-based parser for the Abstract Meaning Representation. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, 1426–1436, Baltimore, Maryland: Association for Computational Linguistics.
- Flanigan, J., Dyer, C., Smith N. A. and Carbonell, J. (2016) Generation from Abstract Meaning Representation using Tree Transducers. In Proceedings of NAACL-HLT 2016, 731–739. San Diego, California: Association for Computational Linguistics.

- Goodman, J., Vlachos, A. and Naradowsky, J. (2016). Noise reduction and targeted exploration in imitation learning for Abstract Meaning Representation parsing. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 1–11. Berlin, Germany: Association for Computational Linguistics.
- Kamp, H. and Reyle, U. (1993). From Discourse to Logic; An Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and DRT. Dordrecht: Kluwer Academic Publishers.
- Koller, A. (2015). Semantic Construction with Graph Grammars. In Proceedings of the 11th International Conference on Computational Semantics, 228–238. London, UK: Association for Computational Linguistics.
- Langacker, R.W. (2008). Cognitive Grammar: A Basic Introduction. New York, USA: Oxford University Press.
- Langkilde, I. and Knight, K. (1998). Generation that exploits corpus-based statistical knowledge. In Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and the 17th International Conference on Computational Linguistics, 704–710. Montreal, Quebec, Canada: Association for Computational Linguistics.
- Li, B., Wen, Y., Bu, L., Qu, W. and Xue, N. (2016). Annotating the Little Prince with Chinese AMRs. In Proceedings of LAW X – The 10th Linguistic Annotation Workshop, 7–15. Berlin, Germany: Association for Computational Linguistics.
- Li, X., Nguyen, T. H., Cao, K. and Grishman, R. (2015). Improving Event Detection with Abstract Meaning Representation. In Proceedings of the First Workshop on Computing News Storylines, 11–15. Beijing, China: Association for Computational Linguistics and The Asian Federation of Natural Language Processing.
- Lozano, L. (2005). Los diferentes tipos de “se” en ELE. In redELE, 3, doi: 10.4438/1697-9346-BV Retrieved from: <https://sede.educacion.gob.es/publiventa/d/20267/19/0>
- Marcus, M.P., Santorini B., and Marcinkiewicz M.A. (1993). Building a large annotated corpus of English: The Penn Treebank. Computational Linguistics, 19(2): 313–330. Association for Computational Linguistics. Cambridge, MA: MIT Press.
- Matthiessen, C.M.I.M. and J.A. Bateman. 1991. Text Generation and Systemic-Functional Linguistics. London, UK: Pinter.
- Oepen, S., Kuhlmann, M., Miyao, Y., Zeman, D., Cinkova, S., Flickinger, D., Hajic, J. and Uresova, Z. (2015). Semeval 2015 task 18: Broad-coverage semantic dependency parsing. In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), 915–926. Denver, Colorado: Association for Computational Linguistics.
- Palmer, M., Gildea, D. and Kingsbury, P. (2005). The proposition bank: An annotated corpus of semantic roles. Computational linguistics, 31(1): 71-106. Cambridge, MA: MIT Press.

- Pan, X., Cassidy, T., Hermjakob, U., Ji, H. and Knight, K. Unsupervised Entity Linking with Abstract Meaning Representation. (2015). In Proceedings of Human Language Technologies: The 2015 Annual Conference of the North American Chapter of the ACL, 1130–1139. Denver, Colorado: Association for Computational Linguistics.
- Sampson, G. (2002). Empirical Linguistics. London: Continuum.
- Saphra, N. and Lopez, A. (2015). AMRICA: an AMR Inspector for Cross-language Alignments. In Proceedings of NAACL-HLT 2015, 36–40. Denver, Colorado: Association for Computational Linguistics.
- Song, L., Zhang, Y., Peng, X., Wang, Z. and Gildea D. (2016). AMR-to-text generation as a Traveling Salesman Problem. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2084–2089. Austin, Texas: Association for Computational Linguistics.
- Taulé, M., Martí, M.A., Recasens, M. (2008) Ancora: Multilevel Annotated Corpora for Catalan and Spanish. In Proceedings of 6th International Conference on Language Resources and Evaluation, pp. 96-101, Marrakesh, Morocco: European Language Resources Association.
- Thomas, J.J., Cook, K.A. (Eds). (2005). Illuminating the Path. The Research and development Agenda for Visual Analytics. Los Alamitos, CA: IEEE Computer Society.
- Uresova, Z., Hajic, J. and Bojar O. (2014). Comparing Czech and English AMRs. In Proceedings of the Workshop on Lexical and Grammatical Resources for Language Processing, 55–64, Dublin, Ireland: Association for Computational Linguistics.
- Vanderwende, L. (2015) NLPwin – an introduction. Microsoft Research tech report no. MSR-TR-2015-23. Retrieved from:
<https://www.microsoft.com/en-us/research/project/nlpwin-parses-amr/>
- Vanderwende, L., Menezes, A., Quirk, C. (2015). An AMR parser for English, French, German, Spanish and Japanese and a new AMR-annotated corpus. In Proceedings of NAACL-HLT 2015, 26–30. Denver, Colorado: Association for Computational Linguistics.
- Venhuizen, N.J., Basile, V., Evang, L., Bos, J. (2013). Gamification for Word Sense Labeling. In Proceedings of the 10th International Conference on Computational Semantics, 397–403, Potsdam, Germany: Association for Computational Linguistics.
- Venhuizen, N.J., Bos, J., Hendriks, P., & Brouwer, H. (2014). How and why conventional implicatures project. In Proceedings of Semantics and Linguistic Theory 24, 63-83. New York, USA: CLC Publications.
- Wang, C., Xue, N., Pradhan, S., Pan, X. and Ji, H. (2016). CAMR at SemEval-2016 task 8: An extended transition-based AMR parser. In Proceedings of the 10th International Workshop on Semantic Evaluation, 1173–1178. San Diego, California: Association for Computational Linguistics.

- Weck, B. (2015). Assessing the impact of manual corrections in annotated corpora: A case study on the Groningen Meaning Bank. (Master's thesis, University of Groningen, Groningen, The Netherlands and University of Charles University in Prague, Prague, Czech Republic). Retrieved from https://lct-master.org/contents_2014/theses.php
- Xue, N., Bojar, O., Hajic, J., Palmer, M., Uresova, Z. and Zhang, X. (2014). Not an Interlingua, But Close: Comparison of English AMRs to Chinese and Czech. In Proceedings of the 9th International Conference on Language Resources and Evaluation, 1765–1772. Reykjavík, Iceland: European Language Resources Association.

APPENDIX 1

A.1.1 Spanish AMRs

This appendix contains the 50 Spanish sentences that we have manually annotated. It includes the following metadata:

- `::id` → `llp_es.N`, where “es” indicates that is a Spanish sentence and “N”²¹ its ID number.
- `::annotator` → The nickname of the annotator.
- `::tok` → The sentence to be annotated.

21 The ID number corresponds to the ID number of their corresponding English sentence.

::id lpp_es.1 ::annotator NMA
::tok Por tanto , puedes imaginarte mi asombro cuando una extraña vocecita me despertó al amanecer .
(c / causar
:ARG1 (p / posible
:ARG1 (i / imaginar
:ARG0 (t / tú)
:ARG1 (a / asombrar
:ARG1 (y / yo)
:tiempo (a2 / amanecer
:tiempo-de (d / despertar
:ARG1 y
:ARG0 (v / voz
:mod (e / extraño)
:mod (p2 / pequeño))))))

::id lpp_es.2 ::annotator NMA
::tok Él me contestó : “ Eso no importa ” .
(c / contestar
:ARG0 (e / él)
:ARG1 (y / yo)
:ARG2 (i / importar :polaridad -
:ARG1 (e2 / eso)))

::id lpp_es.3 ::annotator NMA
::tok Dibújame una oveja ...
(d / dibujar
:ARG0 (t / tú)
:ARG1 (o / oveja)
:ARG2 (y / yo)
:modo imperativo)

::id lpp_es.4 ::annotator NMA
::tok Todo es muy pequeño donde vivo .
(p / pequeño
:campo (t / todo)
:grado (m / muy)
:ubicación (v / vivir
:ARG0 (y / yo)))

::id lpp_es.5 ::annotator NMA
::tok ¡ Qué gracioso !
(g / gracioso
:grado (t / tan))

::id lpp_es.6 ::annotator NMA
::tok ¿ De qué planeta eres ?
(s / ser-de-91
:ARG1 (t / tú)
:ARG2 (p / planeta

:campo (a / amr-desconocido)))

::id lpp_es.7 ::annotator NMA

::tok ¿ De dónde vienes hombrecito mío ?

(d / decir

:ARG0 y

:ARG1 (v / venir

:ARG1 (h / hombre

:mod (p / pequeño)

:posee (y / yo))

:ARG3 (a / amr-desconocido))

:ARG2 h)

::id lpp_es.8 ::annotator NMA

::tok Cuando un astrónomo descubre uno de estos , no le asigna un nombre sino solamente un número .

(e / en-vez-de-91

:ARG1 (a3 / asignar

:ARG0 a

:ARG1 (n2 / número)

:ARG2 c

:manera (s / solamente))

:ARG2 (a2 / asignar

:ARG0 a

:ARG1 (n / nombre)

:ARG2 c)

:tiempo (d / descubrir

:ARG0 (a / astrónomo)

:ARG1 (c / cosa

:cant 1

:ARG1-de (i / incluir-91

:ARG2 (e2 / esto :masc e2))))))

::id lpp_es.9 ::annotator NMA

::tok Los niños siempre deberían mostrarse muy pacientes con los mayores .

(r / recomendar

:ARG1 (m / mostrar

:ARG0 (n / niño)

:ARG1 (p / paciencia

:grado (m2 / mucho))

:ARG2 (p2 / persona

:mod (e / edad

:grado (m3 / más)))

:tiempo (s / siempre)))

::id lpp_es.10 ::annotator NMA

::tok Pero sin duda , para los que entendemos la vida , las cifras son asuntos de indiferencia .

(c / contrastar

:ARG2 (o / opinar

:ARG0 (p / persona
:ARG0-de (e / entender
:ARG1 (v / vida)))
:ARG1 (a / asunto
:mod (i / indiferencia)
:campo (c / cifra))
:mod (d / dudar :polaridad -))

::id lpp_es.11 ::annotator NMA
::tok Olvidar a un amigo es triste .
(t / triste
:campo (o / olvidar
:ARG1 (p / persona
:ARG0-de (t2 / tener-rol-rel-91
:ARG2 (a / amigo))))

::id lpp_es.12 ::annotator NMA
::tok No todos han tenido un amigo .
(t / tener :polaridad -
:ARG0 (p / persona
:mod (t2 / todo))
:ARG1 (p2 / persona
:ARG0-de (t3 / tener-rol-rel-91
:ARG1 p
:ARG2 (a / amigo))))

::id lpp_es.13 ::annotator NMA
::tok Pero no tengo el éxito asegurado para nada .
(c / contrastar
:ARG2 (t / tener :polaridad -
:ARG0 (y / yo)
:ARG1 (e / éxito)
:mod (s / seguro)
:grado (p / para-nada)))

::id lpp_es.14 ::annotator NMA
::tok Pero el principito no respondió .
(c / contrastar
:ARG2 (r / responder :polaridad -
:ARG0 (p / príncipe
:mod (p2 / pequeño))))

::id lpp_es.15 ::annotator NMA
::tok Y me quedaba tan poca agua potable que me temía lo peor .
(y2 / y
:op1 (c / causar
:ARG0 (q / quedar
:ARG1 (a / agua
:mod (p / potable)
:mod (p2 / poco

:grado (t / tan)))
:ARG2 (y / yo))
:ARG1 (t / temer
:ARG0 y
:ARG1 (m / malo
:grado (m2 / máximo))))))

::id lpp_es.16 ::annotator NMA
::tok “Creo que es la hora de desayunar”, añadió ella al instante .
(a / añadir

:ARG0 (e / ella)
:ARG1 (c / creer
:ARG0 e
:ARG1 (h / hora
:propósito (d / desayunar)))
:tiempo (a2 / al-instante))

::id lpp_es.17 ::annotator NMA
::tok Pero , como dijo él , “ ¡ Nunca se sabe ! ” .
(c / contrastar

:ARG2 (d / decir
:ARG0 (e / él)
:ARG1 (s / saber :polaridad -
:tiempo (n / nunca))))

::id lpp_es.18 ::annotator NMA
::tok Las erupciones volcánicas son como el fuego de una chimenea .
(p / parecer

:ARG0 (e / erupción
:mod (v / volcánico))
:ARG1 (s / ser-de-91
:ARG1 (f / fuego)
:ARG2 (c / chimenea)))

::id lpp_es.19 ::annotator NMA
::tok El creyó que nunca querría volver .
(c / creer

:ARG0 (e / él)
:ARG1 (q / querer :polaridad -
:ARG0 e
:ARG1 (v / volver
:ARG1 e)
:tiempo (n / nunca)))

::id lpp_es.20 ::annotator NMA
::tok Te pido perdón .
(p / pedir

:ARG0 (y / yo)
:ARG1 (p2 / perdonar
:ARG0 t

:ARG1 y)
:ARG2 (t / tú))

::id lpp_es.21 ::annotator NMA

::tok ¡ Ahora vete !

(i / ir :modo imperativo

:ARG0 (t / tú)

:tiempo (a / ahora))

::id lpp_es.22 ::annotator NMA

::tok Él comenzó , por lo tanto , a visitarlos para ampliar su conocimiento .

(c / causar

:ARG1 (c2 / comenzar

:ARG0 (e / él)

:ARG1 (v / visitar

:ARG0 e

:ARG1 (e2 / entidad :sinespecificar (l/lo)))

:propósito (a / ampliar

:ARG0 e

:ARG1 (c3 / conocimiento

:posee e))))

::id lpp_es.23 ::annotator NMA

::tok Para ellos , todos los hombres son súbditos .

(o / opinar

:ARG0 (e / ellos)

:ARG1 (s / súbdito

:campo (h / hombre

:mod (t / todo))))

::id lpp_es.24 ::annotator NMA

::tok El principito buscó un sitio para sentarse por todos lados , pero todo el planeta estaba cubierto y obstruido por el magnífico manto de armiño del rey .

(c / contrastar

:ARG0 (b / buscar

:ARG0 (p / príncipe

:mod (p2 / pequeño))

:ARG2 (s / sitio

:propósito (s / sentar

:ARG1 p))

:ARG1 (l / lado

:mod (t / todo)))

:ARG1 (y / y

:op1 (c2 / cubrir

:ARG1 (p3 / planeta

:mod (t2 / todo))

:ARG2 (m / manto

:mod (m2 / magnífico)

:consistir-en (a / armiño)

:posee (r / rey)))

:op2 (o / obstruir
:ARG0 m
:ARG1 p3)))

::id lpp_es.25 ::annotator NMA
::tok “ Bostezar en presencia del rey contradice el protocolo ” , le dijo el monarca .
(d / decir
:ARG0 (m / monarca)
:ARG1 (c / contradecir
:ARG0 (b / bostezar
:ubicación (p / presencia
:posee (r / rey)))
:ARG1 (p2 / protocolo))
:ARG2 (e / entidad
:sinespecificar (l / le)))

::id lpp_es.26 ::annotator NMA
::tok “ No puedo evitarlo ” , respondió el principito , verdaderamente avergonzado .
(r / responder
:ARG0 (p / príncipe
:mod (p2 / pequeño))
:ARG2 (p3/ posible :polaridad -
:ARG1 (e / evitar
:ARG0 p
:ARG1 (e / entidad
:sinespecificar (l / lo))))
:mod (a / avergonzado
:manera (v / verdadero)))

::id lpp_es.27 ::annotator NMA
::tok He tenido un largo viaje y no he dormido nada ...
(y / y
:op1 (t / tener
:ARG0 (y2 / yo)
:ARG1 (v / viaje
:mod (l / largo)))
:op2 (d / dormir :polaridad -
:ARG0 y2
:mod (n / nada)))

::id lpp_es.28 ::annotator NMA
::tok Hace años que no veo a nadie bostezar .
(v / ver :polaridad -
:ARG0 (y / yo)
:ARG1 (b / bostezar
:ARG0 (n / nadie))
:tiempo (d / desde
:op1 (c / cantidad-temporal
:cant (m / más-de
:op1 1)

:unidad (a / año))))

::id lpp_es.29 ::annotator NMA

::tok Era un monarca absoluto .

(m / monarca

:masc m

:campo (e / entidad

:sinnombre (s / sinnombre))

:mod (a / absoluto))

::id lpp_es.30 ::annotator NMA

::tok ¿ Me puedo sentar ? , preguntó tímidamente el principito .

(p / preguntar

:ARG0 (p2/ príncipe

:mod (p3 / pequeño))

:ARG1 (p4 / permitir :modo imperativo

:ARG1 (s / sentar

:ARG1 p2))

:manera (t / tímido))

::id lpp_es.31 ::annotator NMA

::tok No te vayas .

(i / ir :polaridad -

:modo imperativo

:ARG0 (t / tú))

::id lpp_es.32 ::annotator NMA

::tok ¡ Pero aquí no hay nadie a quien juzgar !

(c / contrastar

:ARG2 (n / nadie

:polaridad -

:ARG1-de (j / juzgar)

:ubicación (a / aquí))

::id lpp_es.33 ::annotator NMA

::tok Es mucho más difícil juzgarse a sí mismo que juzgar a otros .

(d / difícil

:grado (m / más

:grado (m2 / mucho))

:campo (j / juzgar

:ARG0 (u / uno)

:ARG1 u)

:comparado-con (j2 / juzgar

:ARG1 (o / otro))

::id lpp_es.34 ::annotator NMA

::tok ¡ Qué sombrero tan raro llevas !

(s / sombrero

:mod (r / raro

:grado (t / tan))

:ARG1-de (l / llevar
:ARG0 (t2 / tú)))

::id lpp_es.35 ::annotator NMA
::tok Los vanidosos nunca escuchan nada que no sea elogios .
(e / escuchar :polaridad -
:ARG0 (p / persona
:mod (v / vanidoso))
:ARG1 (n / nada)
:tiempo (n2 / nunca)
:condición (e2 / elogio :polaridad -
:campo n))

::id lpp_es.36 ::annotator NMA
::tok Admirar significa que me consideras el hombre más guapo , el mejor vestido , el
más rico y el más inteligente del planeta .
(s / significar
:ARG1 (a / admirar)
:ARG0 (c / considerar
:ARG0 (t / tú)
:ARG1 (y / yo)
:ARG2 (h / hombre
:mod (g / guapo
:grado (m / máximo))
:ARG1-de (v / vestir
:manera (b / bien
:grado (m2 / máximo)))
:mod (r / rico
:grado (m3 / máximo))
:mod (i2 / inteligente
:grado (m4 / máximo))
:ubicación (p / planeta)))

::id lpp_es.37 ::annotator NMA
::tok “ Estoy bebiendo ” , respondió el bebedor , con aire lúgubre .
(r / responder
:ARG0 (p / persona
:ARG0-de (b / beber))
:ARG2 (b2 / beber
:ARG0 p)
:manera (l / lúgubre))

::id lpp_es.38 ::annotator NMA
::tok El cuarto planeta era de un hombre de negocios .
(p / pertenecer
:ARG0 (p2 / planeta
:ord (e / entidad-ordinal :valor 4))
:ARG1 (h / hombre
:mod (n / negocio)))

::id lpp_es.39 ::annotator NMA

::tok ¡ Tengo tanto que hacer !

(o / obligar

:ARG1 (y / yo)

:ARG2 (h / hacer

:ARG1 (t / tanto)))

::id lpp_es.40 ::annotator NMA

::tok De repente , el hombre de negocios se dio cuenta que no tenía ninguna esperanza de que lo dejaran en paz hasta que contestase esta pregunta .

(p / percatar

:ARG0 (h / hombre

:mod (n / negocio))

:ARG1 (e / esperar :polaridad -

:ARG0 h

:ARG1 (d2 / dejar

:ARG0 (e2 / entidad

:sinnombre (s / sinnombre))

:ARG1 (p2 / paz

:campo h))

:time (h2 / hasta

:op1 (c / contestar

:ARG0 h

:ARG2 (p3 / pregunta

:mod (e2 / esto))))))

:manera (d / de-repente))

::id lpp_es.41 ::annotator NMA

::tok Cuando encuentras un diamante que no le pertenece a nadie , es tuyo .

(p2 / pertenecer

:ARG0 d

:ARG1 t

:condición (e / encontrar

:ARG1 (d / diamante

:ARG0-de (p / pertenecer :polaridad -

:ARG1 (n / nadie)))

:ARG0 (t / tú)))

::id lpp_es.42 ::annotator NMA

::tok Tenía el espacio justo para una farola y un farolero .

(t / tener

:ARG0 (e / entidad

:sinnombre (s / sinnombre))

:ARG1 (e2 / espacio

:mod (j / justo)

:ARG2-de (c / caber

:ARG1 (y / y

:op1 (f / farola)

:op2 (f2 / farolero))))))

::id lpp_es.43 ::annotator NMA

::tok Las órdenes son órdenes .

(o / orden

:campo (o2 / orden))

::id lpp_es.44 ::annotator NMA

::tok Caminarás cuando quieras descansar y el día durará tanto como quieras .

(d / durar

:ARG1 (d2 / día)

:ARG2 (t / tanto

:grado (i / igual)

:comparado-con (q / querer

:ARG0 (t2 / tú)))

:condición (c / caminar

:ARG0 t2

:tiempo (q2 / querer

:ARG0 t2

:ARG1 (d3 / descansar))))

::id lpp_es.45 ::annotator NMA

::tok “ Soy desafortunado ” , dijo el farolero .

(d / decir

:ARG0 (f / farolero)

:ARG1 (a / afortunado

:campo (y / yo)

:polaridad -))

::id lpp_es.46 ::annotator NMA

::tok Es innecesario .

(n / necesitar

:ARG1 (e / entidad

:sinnombre (s / sinnombre)

:masc e)

:polaridad -)

::id lpp_es.47 ::annotator NMA

::tok “ Estoy aquí ” , dijo la voz , “ debajo del manzano ” .

(d / decir

:ARG0 (v / voz)

:ARG1 (e / estar-ubicado-en-91

:ARG1 (y / yo)

:ARG2 (a / aquí

:ubicación (d2 / debajo

:op1 (m / manzano))))))

::id lpp_es.48 ::annotator NMA

::tok “ Tienen mucha prisa ” , dijo el principito .

(d / decir

:ARG0 (p / príncipe

:mod (p2 / pequeño))
:ARG1 (a / apresurado
:campo (e / entidad
:sinnombre (s / sinnombre))
:grado (m / muy)))

::id lpp_es.49 ::annotator NMA
::tok Los expertos han hecho los cálculos .
(c / calcular
:ARG0 (p / persona
:mod (e / experto)))

::id lpp_es.50 ::annotator NMA
::tok Dame algo de beber ...
(d / dar :modo imperativo
:ARG0 (t / tú)
:ARG2 (y / yo)
:ARG1 (a / algo
:propósito (b / beber)))

A.1.2 English AMRs

This appendix contains the 50 English sentences from *The Little Prince Corpus*²². It includes the following metadata:

- `::id` → `llp_en.N`, where “en” indicates that is an English sentence and “N”²³ its ID number.
- `::annotator` → `lpp_1943.N`, where “N” indicates its ID number within *The Little Prince Corpus*.
- `::tok` → The sentence to be annotated.

²² Available at: <https://amr.isi.edu/download/amr-bank-v1.6.txt>

²³ The ID number corresponds to the ID number of their corresponding Spanish sentence.

```

# ::id lpp_en.1 ::annotator lpp_1943.43
# ::tok Thus you can imagine my amazement , at sunrise , when I was awakened by an
odd little voice .
(c / cause-01
  :ARG1 (p / possible-01
    :ARG1 (i2 / imagine-01
      :ARG0 (y / you)
      :ARG1 (a / amaze-01
        :ARG1 (i / i)
        :time (s / sunrise
          :time-of (w / wake-01
            :ARG0 (v / voice
              :mod (o / odd)
              :mod (l / little))
            :ARG1 i))))))

# ::id lpp_en.2 ::annotator lpp_1943.64
# ::tok He answered me : " That does n't matter . "
(a / answer-01
  :ARG0 (h / he)
  :ARG1 (i / i)
  :ARG2 (m2 / matter-01
    :ARG1 (t / that)
    :polarity -))

# ::id lpp_en.3 ::annotator lpp_1943.65
# ::tok Draw me a sheep ...
(d / draw-01
  :ARG0 (y / you)
  :ARG1 (s / sheep)
  :ARG2 (i / i)
  :mode imperative)

# ::id lpp_en.4 ::annotator lpp_1943.72
# ::tok Where I live , everything is very small .
(s / small
  :degree (v / very)
  :domain (e / everything)
  :location (l2 / live-01
    :ARG0 (i / i)))

# ::id lpp_en.5 ::annotator lpp_1943.118
# ::tok That is funny !
(f2 / funny
  :domain (t2 / that))

# ::id lpp_en.6 ::annotator lpp_1943.122
# ::tok Which is your planet ?
(p / planet
  :poss (y / you))

```

:domain (a / amr-unknown))

::id lpp_en.7 ::annotator lpp_1943.130
::tok My little man , where do you come from ?
(s / say-01

:ARG1 (c / come-01
:ARG1 (y / you)
:ARG3 (a / amr-unknown))
:ARG2 (m / man
:mod (l / little)
:poss (i / I)))

::id lpp_en.8 ::annotator lpp_1943.150
::tok When an astronomer discovers one of these he does not give it a name , but only a number .

(g / give-01
:ARG0 (a / astronomer)
:ARG1 (n2 / number)
:ARG2 (t / thing :quant 1
:ARG1-of (i / include-91
:ARG2 (t2 / this)))
:mod (o2 / only)
:time (d / discover-01
:ARG0 a
:ARG1 t)
:ARG1-of (i2 / instead-of-91
:ARG2 (n3 / name-01
:ARG0 a
:ARG1 t)))

::id lpp_en.9 ::annotator lpp_1943.181
::tok Children should always show great forbearance toward grown - up people .
(r / recommend-01

:ARG1 (s / show-01
:ARG0 (c / child)
:ARG1 (f / forbearance
:mod (g / great))
:ARG2 (g2 / grown-up)
:time (a / always)))

::id lpp_en.10 ::annotator lpp_1943.182
::tok But certainly , for us who understand life , figures are a matter of indifference .
(c2 / contrast-01

:ARG2 (c / certain
:domain (i2 / indifferent-01
:ARG1 (w2 / we
:ARG0-of (u / understand-01
:ARG1 (l / life)))
:ARG2 (f2 / figure))))

::id lpp_en.11 ::annotator lpp_1943.190
::tok To forget a friend is sad .
(s / sad-02
:ARG0 (f3 / forget-01
:ARG1 (p / person
:ARG0-of (h / have-rel-role-91
:ARG2 (f / friend))))))

::id lpp_en.12 ::annotator lpp_1943.191
::tok Not every one has had a friend .
(h / have-03 :polarity -
:ARG0 (e / everyone)
:ARG1 (p / person
:ARG0-of (h2 / have-rel-role-91
:ARG1 e
:ARG2 (f / friend))))))

::id lpp_en.13 ::annotator lpp_1943.196
::tok But I am not at all sure of success .
(c / contrast-01
:ARG2 (s / sure-02 :polarity -
:ARG0 (i / i)
:degree (a / at
:op1 (a2 / all))
:ARG1-of (s2 / succeed-01
:ARG0 i)))

::id lpp_en.14 ::annotator lpp_1943.288
::tok But the little prince made no reply .
(c / contrast-01
:ARG2 (r / reply-01 :polarity -
:ARG0 (p / prince
:mod (l / little))))

::id lpp_en.15 ::annotator lpp_1943.300
::tok And I had so little drinking-water left that I had to fear for the worst .
(a / and
:op1 (h / have-03
:ARG0 (i / i)
:ARG1 (w / water
:purpose (d / drink-01)
:mod (l / little
:degree (s / so))
:ARG1-of (l2 / leave-17))
:ARG0-of (o / obligate-01
:ARG1 i
:ARG2 (f / fear-01
:ARG0 i
:ARG1 (b / bad-07
:degree (m / most))))))

::id lpp_en.16 ::annotator lpp_1943.394
::tok " I think it is time for breakfast , " she added an instant later .
(a / add-01

:ARG0 (s / she)
:ARG1 (t / think-01
:ARG0 s
:ARG1 (t2 / time
:purpose (b / breakfast-01)))
:time (l / late
:degree (m / more
:quant (i / instant))))

::id lpp_en.17 ::annotator lpp_1943.438
::tok But , as he said , " One never knows ! "
(c / contrast-01

:ARG2 (s / say-01
:ARG0 (h / he)
:ARG1 (k / know-01 :polarity -
:ARG0 (o / one))))

::id lpp_en.18 ::annotator lpp_1943.441
::tok Volcanic eruptions are like fires in a chimney .
(e / erupt-0

:ARG1 (v / volcano)
:ARG1-of (r / resemble-01
:ARG2 (f / fire
:location (c / chimney))))

::id lpp_en.19 ::annotator lpp_1943.445
::tok He believed that he would never want to return .
(b / believe-01

:ARG0 (h / he)
:ARG1 (w / want-01 :polarity -
:ARG0 h
:ARG1 (r / return-01
:ARG1 h)
:time (e / ever)))

::id lpp_en.20 ::annotator lpp_1943.454
::tok I ask your forgiveness .
(a / ask-02

:ARG0 (i / i)
:ARG1 (f / forgive-01
:ARG0 y
:ARG1 i)
:ARG2 (y / you))

::id lpp_en.21 ::annotator lpp_1943.477
::tok Now go !

(g / go-02 :mode imperative
:ARG0 (y / you)
:time (n / now))

::id lpp_en.22 ::annotator lpp_1943.482
::tok He began , therefore , by visiting them , in order to add to his knowledge .

(b / begin-01
:ARG0 (h / he)
:ARG2 (v / visit-01
:ARG0 h
:ARG1 (t2 / they))
:mod (t / therefore)
:purpose (a / add-02
:ARG0 v
:ARG2 (k / knowledge
:poss h)))

::id lpp_en.23 ::annotator lpp_1943.489
::tok To them , all men are subjects .

(o / opine-01
:ARG0 (t2 / they)
:ARG1 (s2 / subject
:domain (m2 / man
:mod (a2 / all))))

::id lpp_en.24 ::annotator lpp_1943.491
::tok The little prince looked everywhere to find a place to sit down ; but the entire planet was crammed and obstructed by the king 's magnificent ermine robe .

(a / and
:op1 (c / cram-01
:ARG1 (r2 / robe
:mod (e2 / ermine)
:mod (m / magnificent)
:poss (k / king))
:ARG2 (p3 / planet
:extent (e3 / entire)))
:op2 (o / obstruct-01
:ARG0 r2
:ARG1 p3)
:concession (l / look-01
:ARG0 (p / prince
:mod (l2 / little))
:ARG1 (p2 / place
:purpose (s / sit-down-02
:ARG1 p))
:location (e / everywhere)))

::id lpp_en.25 ::annotator lpp_1943.493
::tok " It is contrary to etiquette to yawn in the presence of a king , " the monarch said to him .

(s / say-01

:ARG0 (m / monarch)
:ARG1 (c / contrary-01
:ARG1 (y / yawn-01
:location (k / king))
:ARG2 (e / etiquette))
:ARG2 (h / he))

::id lpp_en.26 ::annotator lpp_1943.496

::tok " I can 't stop myself , " replied the little prince , thoroughly embarrassed .

(r / reply-01

:ARG0 (p / prince
:mod (l / little))
:ARG2 (p2 / possible-01 :polarity -
:ARG1 (s / stop-01
:ARG0 p
:ARG1 p))
:manner (e / embarrass-01
:ARG1 p
:degree (t / thorough)))

::id lpp_en.27 ::annotator lpp_1943.497

::tok I have come on a long journey , and I have had no sleep ...

(a / and

:op1 (c / come-01
:ARG1 (i / i)
:ARG2 (j / journey-01
:ARG0 i
:ARG1-of (l / long-03)))
:op2 (s / sleep-01 :polarity -
:ARG0 i))

::id lpp_en.28 ::annotator lpp_1943.500

::tok It is years since I have seen anyone yawning .

(p / pass-03

:ARG1 (y3 / year)
:time (s3 / since
:op1 (s / see-01
:ARG0 (i / i)
:ARG1 (a / anyone
:ARG0-of (y / yawn-01))))))

::id lpp_en.29 ::annotator lpp_1943.513

::tok He was an absolute monarch .

(m / monarch

:mod (a / absolute)
:domain (h2 / he))

::id lpp_en.30 ::annotator lpp_1943.517

::tok " May I sit down ? " came now a timid inquiry from the little prince .

(i / inquire-01

:ARG0 (p / prince
:mod (l / little))
:ARG1 (p2 / permit-01 :mode interrogative
:ARG1 (s / sit-down-02
:ARG1 p))
:manner (t / timid)
:time (n / now))

::id lpp_en.31 ::annotator lpp_1943.564

::tok Do not go .

(g / go-02 :polarity - :mode imperative
:ARG0 (y / you))

::id lpp_en.32 ::annotator lpp_1943.568

::tok But there is nobody here to judge !

(n / nobody
:location (h / here)
:ARG1-of (j / judge-01))

::id lpp_en.33 ::annotator lpp_1943.578

::tok It is much more difficult to judge oneself than to judge others .

(d / difficult
:degree (m / more
:degree (m2 / much))
:domain (j / judge-01
:ARG0 (o2 / one)
:ARG1 o2)
:compared-to (j2 / judge-01
:ARG1 (o / other)))

::id lpp_en.34 ::annotator lpp_1943.610

::tok That is a queer hat you are wearing .

(h2 / hat
:mod (q2 / queer)
:domain (t2 / that)
:ARG1-of (w2 / wear-01))

::id lpp_en.35 ::annotator lpp_1943.624

::tok Conceited people never hear anything but praise .

(h / hear-01 :polarity -
:ARG0 (p / person
:mod (c / conceit))
:ARG1 (a / anything
:concession (p2 / praise-01))
:time (e / ever))

::id lpp_en.36 ::annotator lpp_1943.627

::tok To admire mean that you regard me as the handsomest , the best - dressed , the richest , and the most intelligent man on this planet .

(m / mean-01
 :ARG1 (a / admire-01)
 :ARG2 (r / regard-01
 :ARG0 (y / you)
 :ARG1 (i / i)
 :ARG2 (m6 / man
 :mod (h / handsome
 :degree (m2 / most))
 :ARG1-of (d / dress-01
 :manner (w / well
 :degree (m3 / most)))
 :mod (r2 / rich
 :degree (m4 / most))
 :ARG1-of (i2 / intelligent-01
 :degree (m5 / most))
 :location (p2 / planet
 :mod (t / this))))

::id lpp_en.37 ::annotator lpp_1943.638
 # ::tok " I am drinking , " replied the tippler , with a lugubrious air .
 (r / reply-01
 :ARG0 (p / person
 :ARG0-of (t / tipple-01))
 :ARG2 (d / drink-01
 :ARG0 p)
 :manner (l / lugubrious))

::id lpp_en.38 ::annotator lpp_1943.649
 # ::tok The fourth planet belonged to a businessman .
 (b / belong-01
 :ARG0 (p / planet
 :ord (o / ordinal-entity :value 4))
 :ARG1 (b3 / businessman))

::id lpp_en.39 ::annotator lpp_1943.667
 # ::tok I have so much to do !
 (o / obligate-01
 :ARG1 (i / i)
 :ARG2 (d / do-02
 :ARG1 (m / much
 :degree (s / so))))

::id lpp_en.40 ::annotator lpp_1943.682
 # ::tok The businessman suddenly realized that there was no hope of being left in peace until he answered this question .
 (r / realize-01
 :ARG0 (b / businessman)
 :ARG1 (h / hopeful-03 :polarity -
 :ARG1 (l / leave-14
 :ARG1 (p / peace

```

                                :domain b))
:time (u / until
      :op1 (a / answer-01
            :ARG0 b
            :ARG1 (q / question-01
                   :ARG1 (t / this))))))
:manner (s / sudden))

# ::id lpp_en.41 ::annotator lpp_1943.722
# ::tok When you find a diamond that belongs to nobody , it is yours .
(b / belong-01
  :ARG0 d
  :ARG1 (y / you)
  :condition (f / find-01
              :ARG0 y
              :ARG1 (d / diamond
                     :ARG0-of (b2 / belong-01
                               :ARG1 (n / nobody))))))

# ::id lpp_en.42 ::annotator lpp_1943.757
# ::tok There was just enough room on it for a street lamp and a lamplighter .
(a2 / accommodate-01
  :ARG0 (i / it)
  :ARG1 (a / and
        :op1 (l / lamp
              :mod (s / street))
        :op2 (p / person
              :ARG0-of (l2 / light-04
                       :ARG1 (l3 / lamp))))))
:extent (e / enough
        :mod (j / just)))

# ::id lpp_en.43 ::annotator lpp_1943.779
# ::tok Orders are orders .
(o / order
  :domain (o2 / order))

# ::id lpp_en.44 ::annotator lpp_1943.811
# ::tok When you want to rest , you will walk - - and the day will last as long as you like .
(w / walk-01
  :ARG0 (y / you)
  :time (w2 / want-01
        :ARG0 y
        :ARG1 (r / rest-01
               :ARG1 y))
  :ARG0-of (c / cause-01
           :ARG1 (l / last-01
                 :ARG1 (d / day)
                 :ARG2 (t / temporal-quantity
                       :degree (e / equal)

```

:compared-to (l2 / like-02
:ARG0 y))))))

::id lpp_en.45 ::annotator lpp_1943.815
::tok " I am unlucky , " said the lamplighter .
(s / say-01
:ARG0 (p / person
:ARG0-of (l / light-04
:ARG1 (l2 / lamp)))
:ARG1 (l3 / lucky
:domain p
:polarity -))

::id lpp_en.46 ::annotator lpp_1943.944
::tok It is unnecessary .
(n / need-01
:ARG1 (i / it)
:polarity -)

::id lpp_en.47 ::annotator lpp_1943.1046
::tok " I am right here , " the voice said , " under the apple tree . "
(s / say-01
:ARG0 (v / voice)
:ARG1 (i / i
:location (h / here
:location (u / under
:op1 (t / tree
:mod (a / apple)))
:mod (r / right))))

::id lpp_en.48 ::annotator lpp_1943.1189
::tok " They are in a great hurry , " said the little prince .
(s / say-01
:ARG0 (p / prince
:mod (l / little))
:ARG1 (h / hurry-01
:ARG1 (t / they)
:degree (g / great)))

::id lpp_en.49 ::annotator lpp_1943.1213
::tok Computations have been made by experts .
(m / make-01
:ARG0 (p / person
:ARG1-of (e / expert-01))
:ARG1 (t / thing
:ARG1-of (c / compute-01)))

::id lpp_en.50 ::annotator lpp_1943.1298
::tok Give me some of it to drink ...
(g / give-01

:ARG0 (y / you)
:ARG1 (t / thing
 :ARG1-of (i3 / include-91
 :ARG2 (i / it))
 :quant (s / some))
:ARG2 (i2 / i)
:mode imperative
:purpose (d / drink-01
 :ARG0 i2
 :ARG1 t))

APPENDIX 2²⁴

A.2.1 Conversion of Roles

CONVERSION OF ROLES			
TYPE OF ROLE	ENGLISH	SPANISH	REIFICATION (If applies)
Frame args	:ARGX	:ARGX	X
General semantic roles	:accompanier	:acompañante	acompañar
	:age	:edad	envejecer
	:beneficiary	:beneficiario	beneficiarse-de / recibir
	:cause	:causa	causar
	:compared-to	:comparado-con	ser-comparado-con-91
	:concession	:concesión	ser-concedido-91
	:condition	:condición	cumplir-condición-91
	:consist-of	:consistir-en	X
	:cost	:coste	costar
	:degree	:grado	X
	:destination	:destino	estar-destinado-en-91
	:direction	:dirección	X
	:domain	:campo	X
	:duration	:duración	durar
	:employed-by	:contratado-por	tener-rol-org-91
	:example	:ejemplo	ejemplificar
	:extent	:extensión	tener-extensión-91
	:frequency	:frecuencia	tener-frecuencia-91
	:instrument	:instrumento	tener-instrumento-91
	:li "list item"	:el "elemento de la lista"	tener-el-91
	:location	:ubicación	estar-ubicado-en-91
	:manner	:manera	tener-manera-91
	:meaning	:significado	significar
	:medium	:medio	X
	:mod "modifier"	:mod "modificador"	tener-mod-91
	:mode	:modo	X
	:mode interrogative	:modo interrogativo	X
	:mode exclamative	:modo exclamativo	X
	:mode expressive	:modo expresivo	X
	:name	:nombre	tener-nombre-91
	:ord "ordinals"	:ord "ordinales"	X
	:part	:parte	tener-parte-91
	:path	:camino	X
	:polarity	:polaridad	tener-polaridad-91
	:polite	:cortés	X
	:poss "possesed by"/"possessor"	:posee "poseedor"	tener / poseer
	:purpose	:propósito	tener-propósito-91
	:role	:rol	tener-rol-org-91
	:source	:fuente	ser-de-91
	:subevent	:subevento	tener-subevento-91
:subset	:subconjunto	incluir	
:superset	:superconjunto	incluir	
:time	:tiempo	estar-temporalmente-en-91	
:topic	:tema	referirse-a	
:value	:valor	X	
Quantities	:quant "quantity"	:cant "cantidad"	tener-cantidad-91
	:unit	:unidad	X
	:scale	:escala	X

24 This appendix includes some of the possible conversions. More conversions may need to be applied.

CONVERSION OF ROLES			
TYPE OF ROLE	ENGLISH	SPANISH	REIFICATION (If applies)
Date-entity	:day	:día	X
	:month	:mes	X
	:year	:año	X
	:weekday	:día-semanal	X
	:time	:hora	X
	:timezone	:zona-hora "zona horaria"	X
	:quarter	:cuarto	X
	:dayperiod	:periodo-día "periodo de días"	X
	:season	:estación	X
	:year2	:año2	X
	:decade	:década	X
	:century	:siglo	X
	:calendar	:calendario	X
:era	:era	X	
Ops	:opX	:opX	X
Multi-sentence	:sntX	:frsX	X
Prepositions	:prep-against	:prep-contra	X
	:prep-along-with	:prep-junto-con	X
	:prep-amid	:prep-entre	X
	:prep-among	:prep-entre	X
	:prep-as	:prep-como	X
	:prep-at	:prep-en / :prep-a	X
	:prep-by	:prep-por	X
	:prep-for	:prep-para / :prep-por	X
	:prep-from	:prep-desde / :prep-de	X
	:prep-in	:prep-en	X
	:prep-in-addition-to	:prep-además-de	X
	:prep-into	:prep-en	X
	:prep-on	:prep-en	X
	:prep-on-behalf-of	:prep-en-nombre-de	X
	:prep-out-of	:prep-fuera-de	X
	:prep-to	:prep-a / :prep-hacia	X
	:prep-toward	:prep-hacia	X
:prep-under	:prep-bajo	X	
:prep-with	:prep-con	X	
:prep-without	:prep-sin	X	
Conjunctions	:conj-as-if	:conj-como-si	X

A.2.2 Conversion of Named Entities

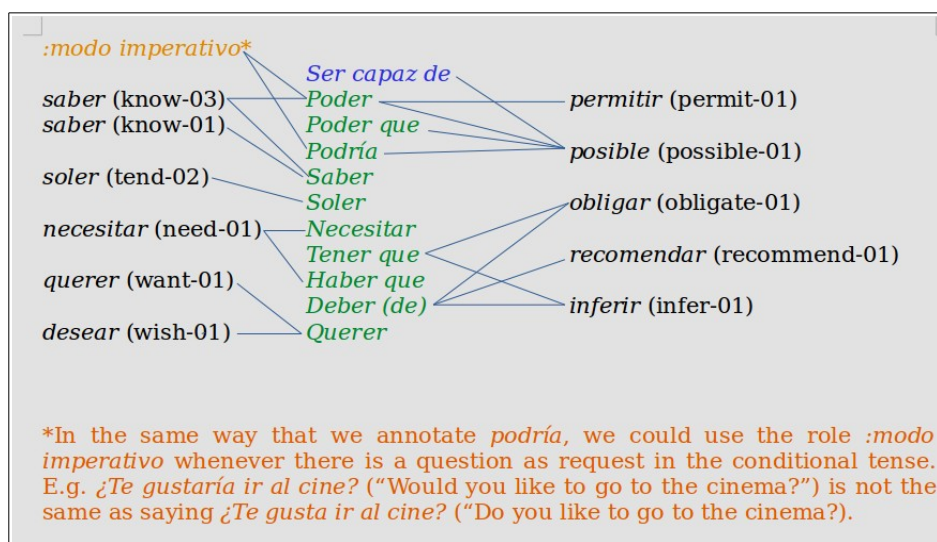
LIST OF NAMED ENTITY TYPES	
ENGLISH	SPANISH
person	persona
family	familia
animal	animal
language	idioma
nationality	racionalidad
ethnic-group	grupo-étnico
regional-group	grupo-regional
religious-group	grupo-religioso
political-movement	movimiento-político
organization	organización
company	empresa
government-organization	organización-gubernamental
military	ejército
criminal-organization	organización-criminal
political-party	partido-político
market-sector	sector-de-mercado
school	colegio
university	universidad
research-institute	centro-de-investigación
team	equipo
league	liga
location	ubicación
city	ciudad
city-district	barrio-de-ciudad
county	condado
state	estado
province	provincia
territory	territorio
country	país
local-region	región-local
country-region	región-nacional
world-region	región-mundial
continent	continente
ocean	océano
sea	mar
lake	lago
river	río
gulf	golfo
bay	bahía
strait	estrecho
canal	canal
peninsula	península
mountain	montaña
volcano	volcán
valley	valle
canyon	cañón
island	isla
desert	desierto
forest	bosque
moon	luna
planet	planeta
star	estrella
constellation	constelación

LIST OF NAMED ENTITY TYPES	
ENGLISH	SPANISH
facility	instalación
airport	aeropuerto
station	estación
port	puerto
tunnel	túnel
bridge	puente
road	carretera
railway-line	línea-ferroviaria
canal	canal
building	edificio
theater	teatro
museum	museo
palace	palacio
hotel	hotel
worship-place	lugar-de-culto
sports-facility	instalación-deportiva
market	mercado
park	parque
zoo	zoo
amusement-park	parque-de-atracciones
event	evento
incident	incidente
natural-disaster	desastre-natural
earthquake	terremoto
war	guerra
conference	conferencia
game	juego
festival	festival
product	producto
vehicle	vehículo
ship	barco
aircraft	aeronave
aircraft-type	tipo-de-aeronave
spaceship	nave-espacial
car-make	marca-de-coche
work-of-art	obra-de-arte
picture	imagen
music	música
show	espectáculo
broadcast-program	programa-de-radiodifusión
publication	publicación
book	libro
newspaper	periódico
magazine	revista
journal	revista-académica
natural-object	objeto-natural
law	ley
treaty	tratado
award	premio
food-dish	plato-de-comida
music-key	tono-musical
musical-note	nota-musical
variable	variable

A.2.3 Conversion of Quantities and Common Units

LIST OF QUANTITIES AND COMMON UNITS		
ENGLISH	SPANISH	COMMON UNITS AND SCALES
monetary-quantity	cantidad-monetaria	:unidad dólar, euro, libra
distance-quantity	cantidad-distancia	:unidad metro, kilómetro, pulgada,año-luz
area-quantity	cantidad-area	:unidad metro-cuadrado, acre, milla-cuadrada
volume-quantity	cantidad-volumen	:unidad litro liter, metro-cúbico, galón
temporal-quantity	cantidad-temporal	:unidad segundo, hora, semana, siglo
frequency-quantity	cantidad-frecuencia	:unidad hercio
speed-quantity	cantidad-velocidad	:unidad kilómetro-por-hora, metro-por-segundo
acceleration-quantity	cantidad-aceleración	:unidad metro-por-segundo-al-cuadrado
mass-quantity	cantidad-masa	:unidad gramo, onza, libra, tonelada
force-quantity	cantidad-fuerza	:unidad newton
pressure-quantity	cantidad-presión	:unidad pascal, bar, torr
energy-quantity	cantidad-energía	:unidad julio, megajulio, caloría
power-quantity	cantidad-potencia	:unidad vatio, caballo-de-potencia
charge-quantity	cantidad-carga	:unidad coulombio
potential-quantity	cantidad-voltaje	:unidad voltio
resistance-quantity	cantidad-resistencia	:unidad ohmio
inductance-quantity	cantidad-inductancia	:unidad henrio
magnetic-field-quantity	cantidad-campo-magnético	:unidad tesla, gauss
magnetic-flux-quantity	cantidad-flujo-magnético	:unidad máxwell, weber
radiation-quantity	cantidad-radiación	:unidad becquerel, curie, sievert
fuel-consumption-quantity	cantidad-consumo-combustible	:unidad litro-por-100-kilómetro
numerical-quantity	cantidad-numérica	:unidad punto, mol
information-quantity	cantidad-información	:unidad bit, byte, kilobyte
concentration-quantity	cantidad-concentración	:unidad molar, micromolar
catalytic-activity-quantity	cantidad-actividad-catalítica	:unidad katal, nanokatal
acidity-quantity	cantidad-acidez	:escala ph
seismic-quantity	cantidad-sísmica	:escala richter
temperature-quantity	cantidad-temperatura	:escala celsius, kelvin, fahrenheit

A.2.4 Conversion of Modals



A.2.5 Other Conversions

OTHER CONVERSIONS			
WORD FORM	ENGLISH	SPANISH	CONCEPT
Preposition	according (to)	según	decir
	instead (of)	en vez de	en-vez-de-91
	per	por (cada)	tasa-entidad-91
	regardless (of)	a pesar de	a-pesar-de
Conjunction	and	y	y
	because	porque	causar
	but	pero	contrastar
	or	o	o
Adverb	however	pero	contrastar
	therefore	por lo tanto	causar
OTHER	ENGLISH ROLE	SPANISH ROLE	COMMENT
Inverse roles	:X-of	:X-de	Where X is a role