OntoLingAnnot's Ontologies: Facilitating Interoperable Linguistic Annotations (Up to the Pragmatic Level)

Antonio Pareja-Lora

Abstract This paper presents the OntoLingAnnot annotation framework, already developed for the annotation of morphological, syntactic, semantic and discourse phenomena, and its extension to cover the annotation of pragmatic phenomena. This extension was considered the ideal test bed for the interoperability of the linguistic annotations performed by means of the platform, since (i) pragmatics itself deals with a real mix of different linguistic topics, such as speech acts, pragmatic coherence relations, deixis, presuppositions and implicatures; and (ii) it clearly interacts with the rest of levels, since (potentially) every linguistic unit at any level can have a pragmatic projection. In particular, it introduces the different pragmatic units that can be used to annotate texts and dialogues using the framework. These pragmatic units are included in the set of ontologies associated to OntoLingAnnot, whose design requirements and development process are also described here. Besides, this paper shows as well the main principles and properties of the OntoLingAnnot annotation framework that help its different annotations interoperate.

1 Introduction

Linguistics and linguistic annotation are very wide fields and, due to this, they have been traditionally partitioned somehow for their study and/or research. Therefore, the most usual criterion to partition them is based on the concept of level, which divides Linguistics into, for example, morphology, syntax, semantics, discourse and/or pragmatics. This partition of Linguistics and its applications has given rise to several good separate models of its different levels, which, nonetheless (and unfortunately), cannot interoperate and do not benefit from the advances of the others in most of the cases. This is due to the fact that this partition has also led to a some-

Antonio Pareja-Lora
ILSA-UCM / ATLAS-UNED, Facultad de Informática (UCM) – Profesor José García Santesmases s/n – 28040-Madrid. e-mail: apareja@sip.ucm.es

what poor communication between the resulting subareas and a lack of a global perspective of Linguistics and, in particular, of linguistic annotation.

Much on the contrary, the OntoLingAnnot framework has been developed following a comprehensive and complementary approach, which considers all these levels of annotation together, not separately. As commented in Pareja-Lora and Aguado de Cea (2010), this comprehensive approach allowed comparing these different levels and "finding the differences and similarities between them, so as to bear a general and uniform (level-independent) annotation framework across levels. In this comparison process, some regularities and uniformities across levels were found, which help structure and formalize all of them". These regularities and uniformities also help define more interoperable linguistic annotation schemes, since they rely on the common and level-independent properties of linguistic categories, and not so much on their common and level-dependent properties.[1]

This paper presents OntoLingAnnot and its extension in order to cover the annotation of pragmatic phenomena. In this extension, pragmatics was regarded as the ideal test bed for the interoperability of linguistic annotations. On the one hand, pragmatics itself deals with a real mix of different linguistic topics, such as (i) speech acts (Searle, 1979), (ii) pragmatic coherence relations (Asher and Lascarides, 2003); or (iii) deixis, presuppositions and implicatures (Levinson, 1983). These different pragmatic topics have been tackled traditionally following several fragmentary and/or partial approaches. Thus, developing an overall annotation scheme for pragmatics involved the interoperation of several separate (types of) linguistic annotations. On the other hand, the clear interaction of pragmatics with the rest of levels already included in OntoLingAnnot[2] (since, potentially, every linguistic unit at any level can have a pragmatic projection) entailed making the annotations of the remaining levels interoperate at least with the pragmatic ones.

Accordingly, this paper shows the main principles and properties of the OntoLingAnnot annotation framework that help its different annotations interoperate, together with an important part of the formalization of the pragmatic (annotation) level integrated into this framework. In particular, it introduces the different pragmatic units that can be used to annotate texts and dialogues using the framework. These pragmatic units are included in the set of ontologies (Gruber, 1993; Borst, 1997) associated to OntoLingAnnot, whose design requirements and development process are also described here.

This paper is organized as follows: Section 2 states the background and the main assumptions underlying the OntoLingAnnot annotation framework, distributed between Sect. 2.1, which summarizes OntoLingAnnot's basic principles and components, and Sect. 2.2, which discusses its improvements and own contributions. Then,

[1] Our efforts are related to other contributions in this volume, in particular Windhouwer and Wright (this vol.) and Chiarcos (this vol.). Whereas these, however, take a bottom-up perspective on linguistic annotations and register categories from existing annotation schemes (Windhouwer and Wright, this vol.) or generalize over them (Chiarcos, this vol.), OntoLingAnnot takes a top-down perspective in that it provides a formalization of linguistic phenomena which is then linked to annotation schemes.

[2] Morphology, syntax, semantics and discourse.

OntoLingAnnot's pragmatic units are presented in Sect. 3. The results and the conclusions drawn from the development and the evaluation of the framework are commented on in Sect. 4.

2 The OntoLingAnnot Framework

OntoLingAnnot (Pareja-Lora and Aguado de Cea, 2010; Pareja-Lora, in press) is a new annotation framework that applies some of the principles underlying the OntoTag annotation model of Aguado de Cea et al (2002, 2004) and also reuses some of its components. Hence, it is necessary to first introduce the OntoLingAnnot principles and components coming from OntoTag.

2.1 OntoLingAnnot's Principles and Components Coming from OntoTag

OntoTag basically consists of both (1) an annotation architecture and (2) an annotation scheme. On the one hand, OntoTag's annotation architecture allows for the collaboration and interoperation of several tools on the annotation of web pages. On the other hand, OntoTag's annotation scheme aims at an interoperable and joint annotation of morphosyntactic, syntactic and semantic[3] phenomena.

In particular, the principles and components of OntoTag that OntoLingAnnot applies and reuses come from its annotation scheme. They can be enumerated as follows:

1. A clear differentiation between the linguistic data categories (LDCs) used in the annotations and the format (or the way) in which these annotations are expressed. This differentiation contributes largely to enhancing the interoperability of linguistic annotations and is completely in line with the ISO standards being developed for linguistic annotation. LDCs are the object of a standard coming from the ISO/TC37/SC3, namely ISO/TC37/SC3 – Terminology and other language and content resources (2008), cf. Windhouwer and Wright (this vol.), linguistic annotation schemes are the objects of several other standards coming from the ISO/TC37/SC4, such as (1) ISO/TC37/SC 4 – Language resource management (2008) for morpho-syntactic annotation; (2) ISO/TC37/SC 4 – Language resource management (2009c) for syntactic annotation; (3) ISO/TC37/SC 4 – Language resource management (2009b), and ISO/TC37/SC 4 – Language resource management (2010d) for semantic annotation; (4) ISO/TC37/SC 4 – Language resource management (2010b) for discourse annotation; or (5) ISO/TC37/SC 4

[3] Restricted to senses and named entities.

– Language resource management (2010c) for discourse and pragmatic annotation.[4]
2. The formalization of LDCs as ontological terms. Since OntoTag's ontologies are implemented in OWL, this formalization enables identifying and referring to each LDC by means of its own Uniform Resource Identifier (URI), which is one of the requirements included in another ISO/TC37/SC4 standard (ISO/TC37/SC 4 – Language resource management, 2010a).
3. The basic classification of each LDC as a `Linguistic Unit`, or as a component of a linguistic feature, that is, as a `Linguistic Attribute` or as a `Linguistic Value`, in order to avoid redundancy and facilitate modularization.
4. The distribution of LDCs among three main ontologies (Aguado de Cea et al, 2004), originated by the classification described in the previous item: a Linguistic Unit Ontology (LUO), a Linguistic Attribute Ontology (LAO), and a Linguistic Value Ontology (LVO). These three ontologies were linked together by the Integration Ontology (IO), and they four, altogether, are the main components of OntoTag reused in OntoLingAnnot.
5. Texts units are annotated by means of triples `<LinguisticUnit, LinguisticAttribute, LinguisticValue>`, regardless of the type of annotation performed (morpho-syntactic, syntactic or semantic, in the case of OntoTag).
6. Text unit annotations, that is, the `<LinguisticSubject, LinguisticAttribute, LinguisticValue>` triples, are implemented by means of RDF triples `<Subject, Predicate, Object>`, in which the corresponding linguistic units (i.e., subjects), attributes (i.e., predicates), and values (i.e., objects), are conveniently formalized as classes or instances of one or more ontologies. This fulfils also one of the main requirements of the standard ISO/TC37/SC 4 – Language resource management (2009a), which is being developed within ISO/TC37/SC4 as well.

Nevertheless, these principles and components of OntoTag proved necessary but insufficient and/or too narrow for OntoLingAnnot from the beginning. The next subsection details how they were amended and extended in order to be generalized.

2.2 OntoLingAnnot's Improvements and Own Contributions

OntoLingAnnot's scope is wider and a bit more ambitious than OntoTag's. It is wider, since OntoLingAnnot tries not only to cover syntax and some phenomena of semantics or in the interface between syntax and morphology (i.e., morpho-syntax);

[4] Most surprisingly, ISO/TC37/SC4 has decided to develop these two discourse and pragmatics-related annotation standards in the ISO macro-project for the standards of other indisputable forms of semantic annotations, instead of creating a new project for them, as with ISO/TC37/SC 4 – Language resource management (2008) and ISO/TC37/SC 4 – Language resource management (2009c), for example.

instead, it seeks to cover these three levels completely, as well as discourse and pragmatics. It is more ambitious, since it aims at (a) achieving a real and full interoperability of its annotations and (b) being flexible and customizable, that is, both (b.1) extendable (to Phonology or Prosody, for example) and (b.2) scalable.[5] In other words, it should be possible to derive any particular annotation scheme from OntoLingAnnot by simply adding to the model and/or selecting from it the particular set of LDCs that refer to the phenomena being annotated. This is why the principles and components of OntoTag inherited by OntoLingAnnot proved insufficient (or too narrow) from the beginning and, hence, some of them had to be adapted as follows.

Firstly, the basic classification of LDCs in OntoTag neglected an important group of linguistic categories, that is, linguistic relationships. This is due to the fact that the few linguistic relationships contemplated in OntoTag (e.g., `Syntactic Function` or `Syntactic Dependency`) could be managed as linguistic attributes. However, when the set of LDCs of OntoTag was extended to cover morphology, discourse and pragmatics, it was clear that linguistic relationships deserve to be treated separately from linguistic attributes. It was also clear that (i) some of them share several properties (even across levels: for instance, the relative rank of the elements they interrelate), which could help distinguish several particular classes of linguistic relationships; (ii) most of them could be arranged into a taxonomy of linguistic relationships (e.g., dependencies, functions and coherence relations (see Hovy and Maier, 1995); and (iii) some of them possess their own attributes, which help characterize and subclassify them.[6] These three arguments made of linguistic relationships potential concepts of an ontology. This is how a fifth ontology of linguistic relationships (the Linguistic Relationship Ontology, LRO) came into being and joined immediately the ones present in OntoTag. Linguistic attributes were searched afterwards in order to tell real linguistic attributes from linguistic relationships and move the latter to the LRO. Then, the LRO was swelled with the relationship LDCs that were found when (re-)formalizing the five levels mentioned above.

Secondly, as soon as linguistic relationships were put into play in OntoLingAnnot, it was necessary to reconsider the types of linguistic triples defined in OntoTag. Thus, a new type of triple, `<LinguisticUnit, LinguisticRelation, LinguisticUnit>`, was added to the `<LinguisticUnit, LinguisticAttribute, LinguisticValue>` triples inherited by the framework.

Thirdly, the linguistic triples discussed in the previous paragraphs, when implemented, originated two types of RDF `<Subject, Predicate, Object>`

[5] Although no exhaustiveness can be claimed, there is a huge amount of linguistic terms (around 2000) already formalized in OntoLingAnnot's ontologies. Therefore, for certain linguistic phenomena and for some reasons (for instance, in order to guarantee a satisfactory inter-annotator agreement) it might be useful and recommended to reduce the set of linguistic categories chosen for their annotation. This can be easily done by ignoring some of the more fine-grained (though pertinent) terms in the ontologies, and choosing only the more coarse-grained.

[6] For instance, according to Hovy and Maier (1995), coherence relations are characterized by being established at the discourse or at the pragmatic level, which helps characterize and subclassify coherence relations into two different classes: discourse coherence relations and pragmatic coherence relations.

triples, namely (i) those in which the Predicate is a Linguistic Attribute and the Object is a Linguistic Value; and (ii) those in which the Predicate is a Linguistic Relationship and the Object is a Linguistic Unit.[7] Since all of them were already formalized as classes or instances of OntoLingAnnot's ontologies, this did not diminish the compliance of the model with the standard ISO/TC37/SC 4 – Language resource management (2009a).

2.2.1 OntoLingAnnot's Annotation Processes and/or Layers

Finally, up to this point, LDCs had been classified according to two different criteria or axes, that is, (1) the level to which they belong (morphology, syntax, semantics, discourse and/or pragmatics) and (2) the type of category they constitute (a Linguistic Unit, a Linguistic Attribute, a Linguistic Value or a Linguistic Relationship). Then, they were also classified according to a new criterion, i.e., the annotation process in which they are handled. In effect, some annotation processes were identified when studying the regularities and uniformities that share the different annotation models developed thus far (independently of their level). Following the terminology in Leech et al (1996), each of these annotation processes was referred to as an Annotation Layer, and each of the LDCs was linked to the Annotation Layer in which it is used for annotation. The resulting layers can be described as follows.

The first layer is the Segmentation Layer. In this layer, the linguistic units that are to be annotated are firstly identified and delimited, segmenting thus the text into its constituent units (according to the level considered).

The second layer is the Paradigmatic Labelling Layer, in which the units segmented in the previous layer can be further characterized by sub-classifying them and/or accompanying them with the particular features (i.e., the pairs <Attribute, Value>) of the level in question that they present in the text.

The third layer is the Syntagmatic Relation Identification Layer. To improve the annotation of the text, the linguistic relations holding between the linguistic units at the level considered can be identified as well. This layer is in charge of this type of annotation.

The fourth layer, the Syntagmatic Relation Labelling Layer, is responsible for further (and optionally) refining the annotation of these relations, as with units, sub-classifying and/or characterizing them by means of their corresponding features in the text.

The fifth layer is called the Resulting Unit Layer. A full annotation of a text at a given level includes, apart from the annotation of the layers mentioned above, an optional and complementary annotation of the higher-rank units that result from the composition or aggregation of other units, by means of one or more relations of that level (already identified and annotated). This is performed, at each level, within its particular Resulting Unit Layer. Since linguistic levels cannot be

[7] The Subject is a Linguistic Unit in both cases.

considered disjoint, the units that constitute the interface between two or more levels must be detailed in this layer too. As shown in the development of OntoLingAnnot, in most cases, the annotation of the units on the interface between two levels is a critical aspect as for the interoperability of linguistic annotations.

This classification of LDCs according to their layer helps scale and customize OntoLingAnnot to the needs of each particular linguistic annotation project. Obviously, depending on how deep the annotations need to be, they will include more or less layers of annotation, since they are fairly independent. Once decided which layers are to be annotated, the LDCs (i.e., the terms) available in OntoLingAnnot's ontologies for the annotation of these layers can be automatically selected and extracted. This frees the users of the platform from having to browse the whole set of OntoLingAnnot's LDCs in search of their own linguistic data category selection for a certain type of linguistic annotation.

All these principles, components and improvements constitute the backbone of the OntoLingAnnot annotation framework and its main contributions. The following section summarizes the pragmatic units that it includes. They represent the most relevant LDCs that can be used for the pragmatic annotation of texts according to this framework. The rest of classes and instances included in the pragmatic modules of OntoLingAnnot's ontologies (namely its pragmatic units, pragmatic attributes, pragmatic values and pragmatic relationships) are discussed in Pareja-Lora (in press). They are not included here for the sake of space.

3 The Pragmatic Units of OntoLingAnnot

The main (i.e., top-level) classes of the Linguistic Unit Ontology (LUO) that formalize the pragmatic units contemplated in OntoLingAnnot are `Macroproposition`, `Pragmateme` and `Pragmatic Functional Unit`.

A `Macroproposition` is both a `Pragmatic Unit` and a complex `Discourse Unit`[8] that serves as a unitary construction block at the `Pragmatic Level`. A `Speech Act`, for example, is a type of `Macroproposition`, as well as a `Trope` (e.g., a `Metaphor`). Macropropositions can be regarded as the linguistic units that result from the aggregation of some interrelated propositions from the `Discourse Level` (Dijk , editor). The `Apology` ('Excuse me'), the `Query` ('can you tell me where the nearest police station is, please?') and the `Begging Act` ('please') in Example 1 are instances of this type of units.

Person A: *Excuse me, can you tell me where the nearest police station is, please?*
Person B: *Go down the street and turn left at the traffic lights. I think it's on the right.*

Example 1: An excerpt of a short dialog

Macropropositions can be related to each other by means of pragmatic relations in order to build pragmatemes. The unit `Pragmateme`, hence, represents in On-

[8] That is, macropropositions stand on the discourse-pragmatics interface.

toLingAnnot the main resulting unit of a text pragmatic analysis. The role of this kind of `Pragmatic Unit` in pragmatic annotation can be better understood in the light of the units into which it can be sub-classified, such as `Macroproposition Aggregation Pragmateme` (`MAP`), `Pragmatic Transposition Unit` (`PTU`), `Emphasis-Related Unit` (`ERU`), `Saying` or `Set Phrase`. Example 1, as a whole, constitutes a particular type of `Pragmateme`, i.e., a `Macroproposition Aggregation Pragmateme`, which consists of a `Query` and its corresponding `Answer`, linked at the `Pragmatic Level` by a type of `Adjacency Pair Relation`.

Finally, a `Pragmatic Functional Unit` (`PFU`) signals a `Pragmatic Coherence Relation` (see Pareja-Lora, in press). Thus, a `PFU` is a linguistic unit that makes explicit a pragmatic relation that holds between two (adjacent) pragmatemes in text or in dialogue. This unit extends the concept of `Discourse Functional Unit` (`DFU`, see Romera, 2004) to the `Pragmatic Level`. For this reason, PFUs are to the `Pragmatic Level` and to pragmatic coherence relations as DFUs are to the `Discourse Level` and to discourse coherence relations. The change of `Turn` in Example 1 is an instance of an `Answer PFU`.

In total, OntoLingAnnot contains 192 pragmatic units (in the Linguistic Unit Ontology, LUO). As for the rest of ontological terms concerning pragmatics in this platform, briefly, OntoLingAnnot's ontologies contain

- 26 pragmatic attributes - 10 concepts and 16 instances (in the Linguistic Attribute Ontology, LAO);
- 81 pragmatic values - 27 concepts and 54 instances (in the Linguistic Value Ontology, LVO);
- 86 classes of pragmatic relations (in the Linguistic Relationship Ontology, LRO); and
- 24 pragmatic concepts relating the pragmatic level and its layers (in other OntoLingAnnot's ontologies).

They amount to 409 pragmatic terms: 339 concepts and 70 instances, apart from several other ontological terms (attributes, *SubclassOf*, *PartOf* and ad hoc relations, rules and axioms).

4 Results and Conclusions

This paper has introduced the OntoLingAnnot (linguistic) annotation framework and also its pragmatic units as a way to show its potential for pragmatic annotation and its interoperability with other levels of annotation. As shown in Pareja-Lora (in press), this is the first ontological (and, hence, computable) conceptualization of pragmatics thus far and, hence, it is an important contribution per se to the areas of Ontological Engineering, Pragmatics and Linguistic Annotation. Besides, no other pragmatic model or framework accounts globally and coherently for such a number of pragmatic phenomena and categories as those formalized and included

in OntoLingAnnot's ontologies, which is another important contribution to the areas aforementioned.

There remains the issue of the compliance of OntoLingAnnot with the ISO standards developed so far (namely ISO/TC37/SC 4 – Language resource management, 2008, ISO/TC37/SC 4 – Language resource management, 2009c, ISO/TC37/SC 4 – Language resource management, 2010a, and ISO/TC37/SC 4 – Language resource management, 2009a), which was sought and evaluated all throughout its development. The results of this continued evaluation were fairly satisfactory, in particular as far as the respective LDC coverage of each of its levels was concerned. As for the compliance of OntoLing's pragmatic annotations with ISO standards, the only standard dealing (tangentially) with pragmatics, as understood in this framework, is ISO/TC37/SC 4 – Language resource management (2010c). This standard draft contains a section dealing with speech acts (or dialogue acts, as they are termed in this document). When compared, OntoLingAnnot contains not only the categories for dialogue acts mentioned in ISO/TC37/SC 4 – Language resource management (2010c), but also some others collected mainly from the usual terminology of politics, law and religion (for the extension of commissives and declarations) as well as from several dictionaries (for the extension of directives and expressives). In addition, since speech acts are only a particular type of macropropositions and there are many other types of pragmatic units in OntoLingAnnot, at least its terminological coverage clearly exceeds the one of ISO/TC37/SC 4 – Language resource management (2010c).

Taking into account that, as shown previously, this approach can also be considered flexible, scalable, extensible and, thus, highly (re)usable, OntoLingAnnot can be viewed as an alternative reference model for the development of future linguistic and interoperable annotations.

Acknowledgements This research has been partially (i.e., financially) supported by Banco Santander and UCM, within the GR42/10-962022 project. The author also wants to thank, for their support when carrying out this research, the Engineering of Software Languages and Applications (Ingeniería de Lenguajes Software y Aplicaciones, ILSA) research group, from the Universidad Complutense de Madrid (UCM), and the Artificial Intelligence Techniques for Linguistic ApplicationS (ATLAS) research group, from the Universidad Nacional de Educación a Distancia (UNED).

References

Asher N, Lascarides A (2003) Logics of conversation. Cambridge University Press, Cambridge, UK

Borst WN (1997) Construction of engineering ontologies. PhD thesis, University of Twente, Enschede, Netherlands

Aguado de Cea G, Gómez-Pérez A, Álvarez de Mon I, Pareja-Lora A, Plaza-Arteche R (2002) OntoTag: A semantic web page linguistic annotation model. In: Semantic Web Meets Language Resources. AAAI Technical Report WS-02-16, AAAI Press, Menlo Park, California, USA, pp 20–29

Aguado de Cea G, Gómez-Pérez A, Álvarez de Mon I, Pareja-Lora A (2004) OntoTag's linguistic ontologies: Improving semantic web annotations for a better language understanding in machines. In: ITCC '04: Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'04), Volume 2, IEEE Computer Society, Washington, DC, USA, pp 124–128

Chiarcos C (this vol.) Interoperability of corpora and annotations. P. 161-179

Dijk (editor) T (1997) Discourse Studies (2 vols.). Sage, London, UK

Gruber TR (1993) A translation approach to portable ontologies. Knowledge Acquisition 5(2):199–220

Hovy E, Maier E (1995) Parsimonious or Profligate: How Many and Which Discourse Structure Relations? Tech. rep., Information Sciences Institute, University of Southern California, URL http://www.isi.edu/natural-language/people/hovy/papers/93discproc.pdf

ISO/TC37/SC 4 – Language resource management (2008) Morpho-syntactic annotation framework (MAF). International Standard Draft: ISO/DIS 24611, International Organization for Standardization (ISO)

ISO/TC37/SC 4 – Language resource management (2009a) Linguistic annotation framework (LAF). International Standard Draft: ISO/DIS 24612, International Organization for Standardization (ISO)

ISO/TC37/SC 4 – Language resource management (2009b) Semantic annotation framework (SemAF) – Part 1: Time & events. International Standard Draft: ISO/DIS 24617-1, International Organization for Standardization (ISO)

ISO/TC37/SC 4 – Language resource management (2009c) Syntactic annotation framework (SynAF). International Standard Draft: ISO/DIS 24615, International Organization for Standardization (ISO)

ISO/TC37/SC 4 – Language resource management (2010a) Persistent identification and sustainable access (PISA). International Standard, Final Draft: ISO/FDIS 24619, International Organization for Standardization (ISO)

ISO/TC37/SC 4 – Language resource management (2010b) Semantic annotation framework (SemAF) – Discourse structures. New Working Item: ISO/PWI 24617-5, International Organization for Standardization (ISO)

ISO/TC37/SC 4 – Language resource management (2010c) Semantic annotation framework (SemAF) – Part 2: Dialogue acts. International Standard Draft: ISO/DIS 24617-2, International Organization for Standardization (ISO)

ISO/TC37/SC 4 – Language resource management (2010d) Semantic annotation framework (SemAF) – Static spatial information. New Working Item: ISO/PWI 24617-6, International Organization for Standardization (ISO)

ISO/TC37/SC3 – Terminology and other language and content resources (2008) Specification of data categories and management of a Data Category Registry for language resources. International Standard Draft: ISO/DIS 12620.2, International Organization for Standardization (ISO)

Leech G, Barnett R, Kahrel P, Halteren Hv, Langé JM, Montemagni S, Voutilainen A (1996) Recommendations for the Syntactic Annotation of Corpora. European Project Deliverable: EAGLES Document EAG–TCWG–SASG/1.8, EAGLES Consortium, URL http://www.ilc.cnr.it/EAGLES96/segsasg1/segsasg1.html

Levinson SC (1983) Pragmatics. Cambridge University Press, Cambridge, UK, reprinted as Vol. A of *Computers & Typesetting*, 1986

Pareja-Lora A (in press) The pragmatic level of OntoLingAnnot's ontologies and their use in pragmatic annotation for language teaching. In: Bárcena E, Read T, Arús J (eds) Technological innovation in the teaching and processing of LSPS, Springer, Madrid, Spain, pp 547–574, to appear 2012

Pareja-Lora A, Aguado de Cea G (2010) Modelling discourse-related terminology in OntoLingAnnot's ontologies. In: Bhreathnach U, Barra-Cusack F (eds) Presenting terminology and knowledge engineering resources online: Models and challenges (TKE 2010), Dublin City University, Dublin, Ireland, pp 547–574

Romera M (2004) Discourse Functional Units: the Expression of Coherence Relations in Spoken Spanish. LINCOM, Munich, Germany

Searle J (1979) Expression and meaning: Studies in the theory of speech acts. Cambridge University Press, Cambridge, UK, (reprinted 1999)

Windhouwer M, Wright SE (this vol.) Linking to linguistic data categories in ISOcat. P. 99-107