

Bulletin de l'Institut de linguistique

30

Université de Lausanne

Comité de rédaction :

Cédric Margot

Michiel de Vaan

Pascal Singy

ISSN 1023-134X

©Université de Lausanne

Tous droits réservés

Digital Tools for Semantic Annotation: the WoPoss Use Case

Helena Bermúdez Sabel

&

The Semantics of Irish Determiner Phrases

Benjamin Storme

Université de Lausanne

Sommaire

Sommaire	7
Avant-propos	9
Digital Tools for Semantic Annotation: the WoPoss Use Case – Helena Bermúdez Sabel	12
Abstract.....	12
1. Rationale	12
2. The semantic annotation of modality	14
2.1. Project workflow	14
2.2. Specific requirements of the fine-grained annotation	17
3. Testing phase: an overview of annotation tools.....	19
4. Description of the annotation platform.....	23
5. Customization of the annotation platform	27
6. Final remarks.....	32
References	33
The Semantics of Irish Determiner Phrases – Benjamin Storme	38
Abstract.....	38
1. Introduction	38
2. The typology of DPs' semantic types	40
2.1. A working definition for DPs	40
2.2. DPs' semantic types in English	42
2.3. Type-shifting operator from $\langle\langle e, t \rangle, t \rangle$ to $\langle e, t \rangle$	46
3. Semantic tests	47
3.1. Higgins' tests	48
3.2. Partee's and Roy's tests	52
4. Irish DPs.....	54
4.1 Copular sentences.....	54
5. Analysis	73
Preposition <i>mar</i>	74
Preposition <i>ar</i>	76
Preposition <i>i</i>	77
6. Conclusion	81
References	82

Avant-propos

La section des sciences du langage et de l'information (SLI) de l'Université de Lausanne a accueilli plusieurs nouvelles personnalités entre 2018 et 2019. Helena Bermúdez Sabel, PhD en étude médiévale à l'université de Santiago de Compostela (2019) est membre du projet FNS *A World of Possibilities, Modal pathways over an extra-long period of time: the diachrony of modality in the Latin language* (WoPoss) basé à Lausanne. Ce projet, dirigé par Professeure Francesca Dell'Oro, porte sur l'étude diachronique de la modalité dans un corpus de langue latine sur une période allant de la préhistoire au VIIe siècle après JC.

Autre personnalité nouvelle, Benjamin Storme, PhD en linguistique au MIT (2017) est actif dans la recherche en typologie, phonétique, phonologie, syntaxe, morphologie et sémantique. Il enseigne depuis 2018 à l'Université de Lausanne comme Premier assistant et ses travaux actuels portent principalement sur l'évolution des systèmes sonores de différentes langues et les modèles grammaticaux des connaissances phonétiques et phonologiques des locuteurs.

Ce nouveau numéro des bulletins de linguistique (BIL) présente deux travaux issus de ces domaines de recherche en linguistique nouvellement représentés à la section des SLI l'Université de Lausanne. Le premier article de Helena Bermúdez Sabel est une présentation de la méthode et des outils de recherche développés dans le cadre du projet WoPoss. En deuxième partie, l'article de Benjamin Storme porte sur la sémantique des syntagmes déterminatifs de l'irlandais. Ces articles

permettent de rendre compte de l'orientation actuelle de certains projets de recherche présentés à Lausanne.

Cédric Margot

Digital Tools for Semantic Annotation: the WoPoss Use Case – Helena Bermúdez Sabel

Abstract

This paper examines the use of annotation platforms to perform semantic annotation of textual contents. It focuses on a specific tool called INCEpTION. This review stems from a project that studies modality in Latin from a diachronic perspective; thus, the analysis emanates from the development of an annotation pipeline for this particular use case. I briefly overview the role of semantic annotation in the project so as to delve into the specific requirements of the annotation process and how a customized tool assists in this procedure. After justifying the selection of INCEpTION over other annotation environments, a description of the functionalities of the tool is presented. The paper continues with a discussion of the tool's customization that was undertaken in order to meet the requirements of the project. This part draws attention to how the annotation challenges were tackled. To conclude, a general reflection on the use of annotation platforms is presented.

1. Rationale

This paper is developed in the framework of the FNS project *A World of Possibilities. Modal pathways over an extra-long period of time: the diachrony of modality in the Latin language* (WoPoss), led by Francesca Dell'Oro and whose members are Paola Marongiu and the present author.¹ This project studies the evolution of modal meanings in Latin, analysing modality mainly from a

¹ See <<http://woposs.unil.ch>> (accessed on 01/11/2019).

semantic perspective, although not exclusively. In the WoPoss project, modality is understood as the expression of possibility, necessity and probability. Modal meanings are empirically elicited by annotating modal passages in a diachronic corpus.

As pointed out by Nissim et al. (2013), many projects dealing with the annotation of modality entail a mere classification task in which annotators assign modality values to pre-selected markers or expressions (Nissim et al. 2013, 8). With regard to WoPoss, we have developed a complex annotation scheme (Dell’Oro 2019) that dissects a modal expression into its different components, that is, the modal marker, its scope, the state of affairs and the modal relation between marker and scope. These units are later described using various linguistic features.

The theoretical framework of WoPoss is largely based on Nuyts (2016). With respect to the annotation scheme, it was influenced by the work developed under the umbrella of the project *Modal – Modèles de l’annotation de la modalité à l’oral* (Ghia et al. 2016).² We also drew inspiration from the annotation parameters used by Jan Nuyts in his projects on the diachrony of the Dutch modal verbs.³

The complexity of our schema requires, on the one hand, a tool able to formalize the intricacy of this multifaceted linguistic phenomenon. On the other hand, it demands an annotation environment that makes it possible for people with different profiles to work collaboratively. We need a space for the inexperienced annotators to learn and practice and the experienced ones to guide them.

² The complete annotated corpus is available online (Pietrandrea et al. 2016). For more information about this project see <<https://modal.msh-vdl.fr/>> (accessed on 01/11/2019).

³ For a list of projects by Jan Nuyts concerning modality, please see the list available at <<http://woposs.unil.ch/credits.php>> (accessed on 01/11/2019).

The following section outlines the workflow of the project in order to contextualize the role of semantic annotations as part of the development (Section 2.1). In Section 2.2, I detail the specific requirements that an annotation platform should meet for the correct formalization of our annotation scheme. After analysing different annotation platforms and workflows (Section 3), the members of the WoPoss project concluded that INCEpTION⁴ was the most suitable annotation tool for our necessities. A brief description of this platform will be introduced in Section 4 and this review will focus on the functionalities that make this tool different from other annotation platforms. Section 5 sketches the customization of the tool that was developed in order to make of this platform a functional resource for our project. This section will pay special attention to the challenges presented in section 2.2 explaining how the customization works around them. The paper will conclude with some general remarks about the use of tools for semantic annotation.

2. The semantic annotation of modality

2.1. Project workflow

WoPoss has a corpus-based approach for the study of modality. Diversity was used as a determining factor for the selection of the texts to be included: we aim at a representative corpus in terms of diachronic, diatopic, diastratic and diaphasic parameters. The corpus spans the period from the 3rd BCE to the 7th century CE, and besides the different textual types and genres, we also took into consideration the various sources of transmission of ancient texts.⁵

4 <<https://inception-project.github.io>> (accessed on 01/11/2019).

5 About the importance of including both documentary and literary texts for the study of ancient languages, see Dell'Oro (2015).

Firstly, the selected works are gathered by retrieving them from different online resources that contain the texts under a free licence.⁶ These sources present the texts in different formats, so they are first converted into plain text. Pseudo-markup is added to the textual content in order to preserve the pertinent semantic information previously conveyed in the XML or HTML tags: this affects the chapter and book divisions, verse lines and foreign words, among other pieces of information that were contained in the source files.

Afterwards, these documents are automatically annotated using the StanfordNLP library for Python.⁷ Thus, lemmas, part of speech categorization, morphological features, and syntactic dependencies are added. The resulting CONLL-U⁸ files are uploaded to the annotation platform INCEpTION.

Through this platform, the annotators add the relevant semantic information. They read the complete text paying closer attention to any occurrence of the pre-selected modal markers (Dell'Oro 2019, 9-10). Every time a potential modal passage is found, they correct the automatic annotation (if needed) and then they annotate the passage according to the WoPoss scheme and guidelines. Inter-annotator agreement is frequently checked, especially when it involves less experienced annotators. Disagreements are discussed in order to evaluate whether the passage is ambiguous or if one of the annotators misinterpreted either the text or the (sub-)type of modality.

6 See <<http://woposs.unil.ch/credits.php>> (accessed on 13/11/2019) for a list of the digital libraries and sources employed.

7 <<https://stanfordnlp.github.io/stanfordnlp>> (accessed on 13/11/2019).

8 In the CONLL-U format, annotations are encoded in plain text files. Blank lines mark sentence boundaries, each line concerns the analysis of a word and each value of this analysis is separated by a single tab character. For a detailed explanation of this format see <<https://universaldependencies.org/format.html>> (accessed on 20/11/2019).

The revised annotated texts are then exported to the XMI format, one of the output formats available in INCEpTION, and one that is easy to transform into XML-TEI. Our annotated dataset will be preserved in TEI with the linguistic information encoded through stand-off annotation.

After this first transformation, a series of steps are implemented to cure the annotated documents and add more information.

The first one entails the automatic addition of linguistic features concerning the most ancient meaning of each modal marker. This meaning is elicited by reviewing and synthesizing lexicographical resources.⁹

The second step affects the pseudo mark-up that was added when the sources were first converted to plain text. These graphical conventions are then transformed into TEI elements. In a similar manner, miscellaneous information that was kept (unstructured) during the annotation process in a field labelled “note” is analysed and disambiguated, adding the pertinent XML elements when necessary.

The parameters that are relevant for the selection of the corpus – textual genre, type of transmission, chronology and origin of the author – are part of the metadata that will be added automatically using the Digital Humanities Toolkit (DHTK) (Picca and Egloff 2017).

At this point, the dataset is ready to be stored in a no-SQL database and to be published and exploited through a user-friendly interface.

⁹ The bases of this work are the entries of the markers in the *Thesaurus Linguae Latinae* (Thesaurusbüro München Internationale Thesaurus-Kommission, *n.d.*), and when this resource does not yet provide the description of the lemma, the *Oxford Latin Dictionary* is consulted (Glare 2012). In addition, current etymological dictionaries have also been consulted (Ernout and Meillet 2001; Meiser 2010; Vaan 2008).

As the description of the workflow suggests, WoPoss needed an annotation platform that, in the first place, could import the output format of natural language processing tools. It was important for the annotators to be able to access (and edit) the results of the linguistic automatic annotation so they could implement the pertinent corrections when necessary. In the second place, the support of rich semantic annotation schemes was required. It was important to have the ability to formalize large tagsets and create restrictions around them to facilitate the annotation process and to ensure the accuracy of the annotation. Finally, we needed to be able to export the annotated dataset to a format that allowed its transformation to different output formats. This guarantees the sustainability of the dataset and its efficient exploration and exploitation.

2.2. Specific requirements of the fine-grained annotation

In this section, I will detail the elements that needed to be formalized through the annotation platform from a generic point of view: the rationale of this section is to expound the technical functionalities that an annotation platform must have for the correct modelling of modality as understood in the WoPoss project.

- Interaction between multiple layers of annotation. Semantic interpretation is conditioned by other levels of linguistic analysis. Therefore, it is critical to discern between different layers of linguistic annotation.
- Annotation of relations. As briefly mentioned in Section 1, the theoretical approach to modality of WoPoss discriminates the different components of a modal passage. As understood in

this project, modality concerns the expression of the notions of possibility, necessity and probability. We identify the lexical elements that articulate these notions, that is, the modal markers. Modality concerns the stance of a speaker on a specific representation. This representation is the state of affairs.¹⁰ To analyse the state of affairs, we identify the scope, that is, the part of the clause to which the marker refers, and the participant or participants in the state of affairs, when pertinent. Finally, for each modal passage we examine the abstract relation between the marker and its scope. Therefore, a network of relations needs to be established between different linguistic components: the relation of the marker with its scope and, when relevant, the role of the participant with the scope.

- Annotation of linguistic contents below the word level. A word-based tokenization would not be granular enough to identify the linguistic units that comprise a modal expression. For instance, some of the modal markers selected for annotation concern morphological units smaller than the words, such as the adjectival suffixes *-bilis* or *-turus*. This means that the presence of these adjectives with a modal meaning required a segmentation of the word in which the suffix must be analysed as a modal marker, and the root as part of the scope (Dell’Oro 2020).
- Annotation of discontinuous elements. As the relevant linguistic elements for the annotation may not be contiguous, a method to identify tokens belonging to the same structure is needed.

¹⁰ There are special cases when, for example, the state of affairs is not explicit.

- Overlapping and stacking. Again, syntactic structures may determine the discontinuity and overlap of the segments that form a modal expression. In addition, various combinations of the modal units are possible so a flexible annotation system must be implemented: a marker might affect multiple scopes, or the same scope could be conditioned by more than one marker.
- Annotation of ambiguity. In order to understand modal shift, annotators take care to annotate the possibility of two (or more) modal readings.¹¹

3. Testing phase: an overview of annotation tools

Numerous benchmarks for the evaluation of software are available, including benchmarks that were specifically created for the recommendation of XML editors (van den Broek, Wiering, and van Zwol 2005) which could have been a starting point for the procedure of selecting an annotation tool. However, considering the specificities of the WoPoss project (briefly presented in the Introduction), it was decided to perform a hands-on experience with different annotation tools. Therefore, a mock-up for each evaluated tool was developed as a proof of concept.

For the selection of tools to be reviewed, we took into consideration the resources used by other projects.¹² I will briefly present the reasons why Analec and <Oxygen/> XML Editor were

¹¹ In addition, the discrimination of the meaning conveyed by natural language expressions requires a large amount and wide range of contextual information which is not always available in a project that analyses textual contents created thousands of years ago. Therefore, ambiguity is inevitable (Bunt 2017).

¹² After the proof of concept was finished, it came to my attention that the project *Portuguese Corpus Annotated for Modality - MODAL* (Hendrickx, Mendes, and Mencarelli 2012) used the tool MMAX which was not evaluated by the WoPoss team. This tool is very versatile but it does not support an installation as a service, which is especially useful for working collaboratively in a production environment. Moreover, it has not been updated since 2013. For a description of this tool see Müller and Strube (2006).

rejected in favour of WebAnno. Then, INCEpTION was used instead of WebAnno because the development of the later project merged with that of INCEpTION.

Analec¹³ is a specific tool for textual annotation that has a desktop version and also a plug-in as part of the modular platform for textometry, TXM.¹⁴ The main advantage of Analéc is its very intuitive interface (see Figure 1). It also provides additional functionalities thanks to the built-in analytical tools through which different calculations can be made: computation of frequencies, search of correlations or the establishment of the inter-annotator agreement. There is also the possibility to perform advanced queries of the annotations. Although the modifications of the scheme are easy to implement (and update), the definition of restrictions offers few possibilities. For instance, it is not possible to specify the cardinality of a feature, that is, to define whether a feature is optional or mandatory and whether it can be repeated. Moreover, restrictions conditioned by the value of a specific feature cannot be established, which is an important handicap considering, for instance, how much the description of an epistemic modal passage differs from that of a dynamic one in terms of pertinent features. Therefore, an annotator would have to read over non-pertinent features instead of having more guided annotation choices. The last disadvantage is that the same tokens cannot be analysed more than once. This makes the encoding of ambiguity especially convoluted, since the most straightforward annotation will entail the analysis of the same passage with the different meanings that are the source of the ambiguity.

13 For information about Analéc, see <<http://explorationdecorpus.corpusecrits.huma-num.fr/analéc-2>> (accessed on 04/11/2019) and Landragin et al. (2012).

14 <<http://textometrie.ens-lyon.fr>> (accessed on 04/11/2019).

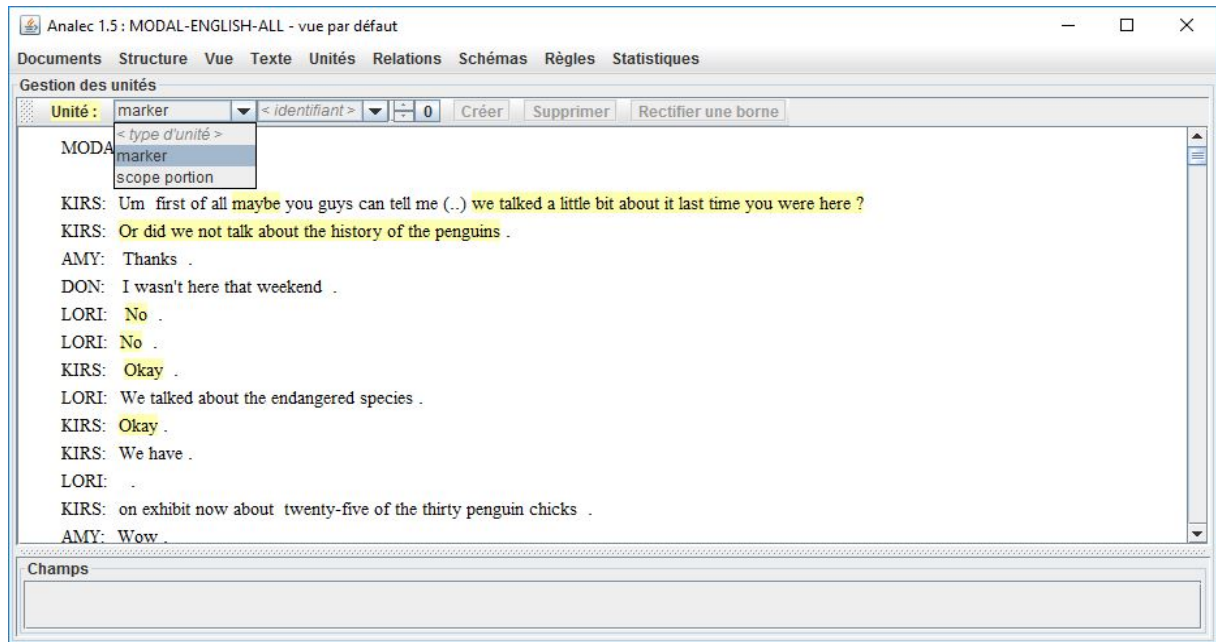


Figure 1. Screenshot of the annotation interface of Analec (*Modal project*)¹⁵

We also performed a proof of concept with the XML editor <oXygen/>¹⁶ using CSS and its Author Mode to customize the annotation experience and make it more user-friendly (Figure 2). The encoding strategy for this test implemented TEI (TEI Consortium 2019b) encoding in which manual annotations were added through stand-off methods.¹⁷ In contrast to Analec, the great advantage of directly editing the XML is the possibility to define a very complex scheme formalized through feature structures. A feature structure is a group of *attribute:value* pairs, where the values may either be atomic or nested feature structures (Witt and Stegmann 2009) so that complex hierarchies can be created, achieving a great level of granularity by describing a linguistic phenomenon as an accumulation of feature structures.

¹⁵ Taken from <<https://modal.msh-vdl.fr/index.php/2016/12/10/english-using-the-analec-tool>> (accessed on 19/11/2019).

¹⁶ <<https://www.oxygenxml.com>> (accessed on 04/11/2019).

¹⁷ More specifically, the mark-up technique detailed in Bermúdez Sabel (2018) built upon the TEI feature structures module (TEI Consortium 2019a).

By using this environment, no changes of format need to be done throughout the workflow since we would be working with XML and XML technologies from the source retrieval to the publication of the annotated dataset. The main disadvantage of this tool is that the annotation process comes off as tedious, especially for less experienced annotators, and specifically when dealing with discontinuous elements for which the boundaries of each segment need to be made explicit.

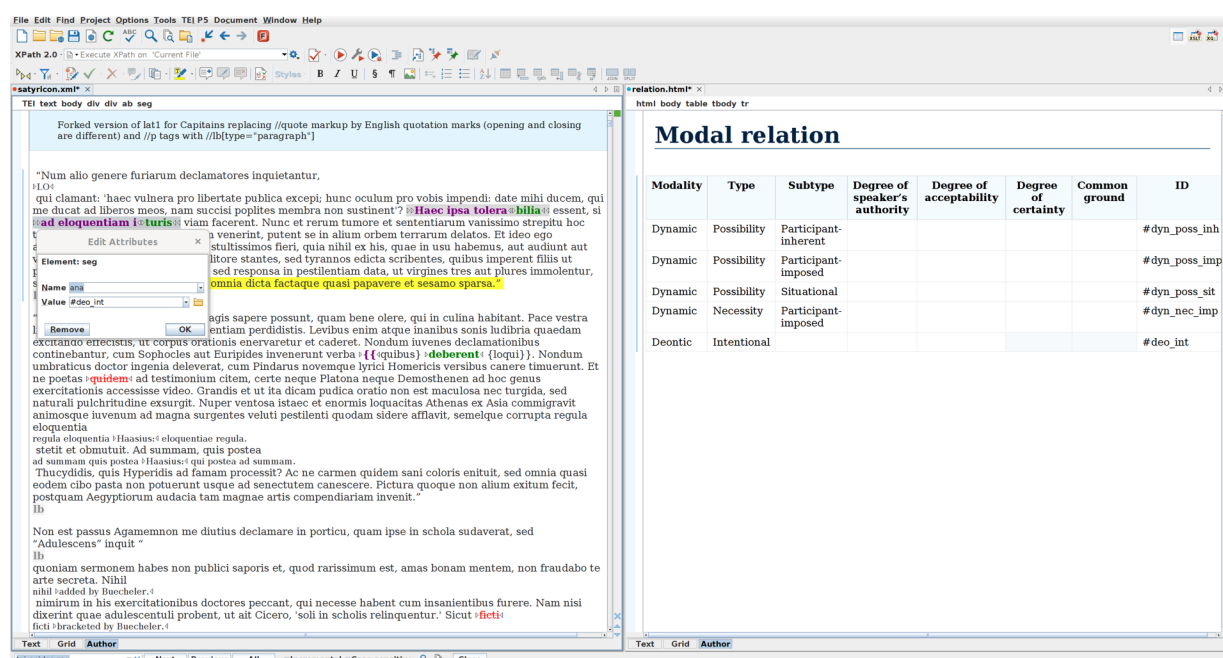


Figure 2. Screenshot of xOxygen/. In the left panel, green and magenta indicate the different type of modal units already annotated, grey highlight defines the modal relation. A dialog prompts the introduction of the code to describe the segment being analysed. These codes are searchable in the table displayed on the right panel.

WebAnno is a general purpose web-based annotation tool. Although conceived for linguistic annotation, it allows the

customization of any layer of annotation thus enabling its use even for non-linguistic annotation.

Since 2018, WebAnno entered a phase of development that is mainly prompted by the specifications of the INCEpTION project. Thus, any updates of WebAnno are done on the basis of how certain parts of this tool can be reused by the INCEpTION project. Therefore, instead of describing WebAnno, I proceed to the examination of INCEpTION in the following section.

4. Description of the annotation platform

INCEpTION is a multi-functional and multi-modular platform that enables the creation of corpora, the annotation of texts, and the management of knowledge.

INCEpTION is presented as a tool particularly adept at handling semantic annotations (Klie et al. 2018). One of the reasons behind this statement is the flexible multi-layer annotation support: different layers can be combined, all of them being implemented with freely configurable annotation schemes. In our case, the interaction and conditioning of semantics with other aspects of linguistic analysis, especially morphosyntactic features, needed to be made explicit so it was crucial to work with an annotation tool that supported multiple levels of annotation.

Besides providing the framework to develop tangential tasks directly related to text annotation, such as corpus management, INCEpTION also includes different features to improve the efficiency of the tasks themselves. Concerning the annotation procedure, it provides intelligent annotation assistance under the form of machine learning recommenders. These recommenders can be used during the annotation process to generate predictions

that the annotators may accept or reject. Through an active learning process, the evaluations by the users are employed to further improve the quality of the predictions (Klie 2018). Besides the built-in recommenders, users can train their own recommenders as their dataset is progressively annotated and validated. This means that the manual annotations can later be used for training and implementing an automatic annotation.

INCEpTION facilitates knowledge management thanks to the knowledge base module. This feature allows users to create their own knowledge base, to import one or to connect to remote knowledge bases, like DBpedia¹⁸ or YAGO.¹⁹ A knowledge base can be used, for instance, for linking entities. Besides adding more information about a particular entity, this step is especially useful for disambiguating mentions. Through a knowledge base, cross-document co-references can be introduced. Adding this type of references not only enriches the annotation of concepts or named entities, but it is also helpful for the addition of complex semantic information like, for example, taxonomic or meronymic relations (Eckart de Castilho et al. 2018).

In regard to the supervision of an annotation project, INCEpTION provides different functionalities to manage collaboration. Thanks to the various types of users available, members of a project with different profiles have a specific environment to perform their particular tasks: corpus management, customization of schemas, annotation, monitoring and curation (see Figure 3). All these functionalities are available depending on the type of user, so while an annotator may only access the annotation interface (see Figure 4), curators have access to a monitoring environment where

18 <<https://wiki.dbpedia.org/>> (accessed on 18/11/2019).

19 <<https://datahub.io/collections/yago>> (accessed on 18/11/2019).

they can check the progress of the annotation. In addition, inter-annotator agreement, that is, the degree of agreement between the annotators of the same text, can be automatically calculated according to three different types of measure.²⁰ Besides the statistical approach, the curator can easily compare the results of different annotators and validate (or reject) their annotations.

As a final remark about INCEpTION, it should be noted that the development of the tool is open: not only is the code freely available in a public repository,²¹ but the discussions and development tasks are also publicly managed via GitHub. It is also worth mentioning that the community of users of this resource has at its disposal an active mailing-list in which the developers are quick to offer their support by answering back to any problems or doubts posed by users.

²⁰ The available measurements are the Cohen's kappa, Fleiss' kappa and Krippendorff's alpha. For more information about the differences between this type of measures see Gwet (2014).

²¹ See <<https://github.com/inception-project/inception>> (accessed on 18/11/2019).

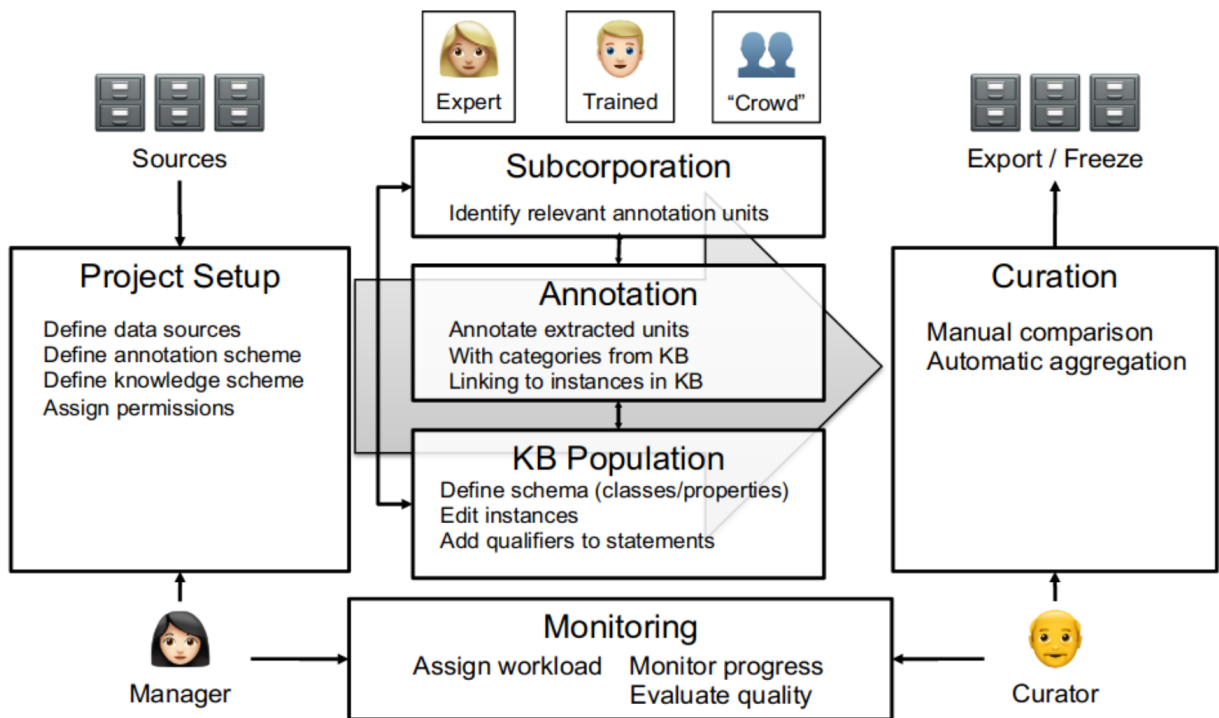


Figure 3. Workflow of a project managed through INCEpTION (Eckart de Castilho et al. 2019)

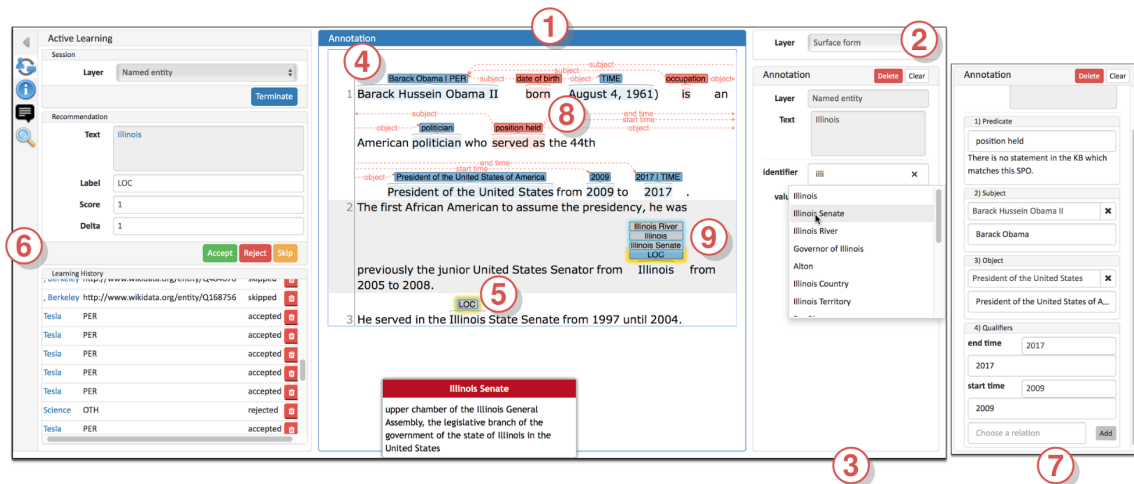


Figure 4. INCEpTION annotation editor: 1) annotation area, 2) annotation layer selection, 3) entity linking feature editor, 4) named entity linked to Wikidata, 5) entity mention suggestion, 6) active learning sidebar, 7) fact linking editor, 8) annotated fact, 9) entity linking recommendations (Klie et al. 2018, 8).

5. Customization of the annotation platform

This section will not delve into the details of user administration or document management. However, it will explain how the theoretical framework and the informational needs elicited during a first annotation test²² were formalized in INCEpTION. Thus, I will present the steps followed to create an annotation scheme in this tool.

The formalization of the annotation scheme entails the definition of different components, viz. the layers, their features and the tagsets that can control the values of those features. In addition, constraints can be declared in order to define the optional or mandatory nature of these elements as well as the relations between them.

The results of the automatic linguistic analysis are conveyed in fourth layers: *Lemma*, *Morphological Features*, *Part of speech*, and *Dependency*.

To encode the different elements that constitute a modal expression, the layer *Modal unit* was created. This is a layer of type “span”: it enables the annotation over a span of text. Span annotations can have any length, can overlap, can stack, can nest, and can cross sentence boundaries, but all these behaviours need to be configured. As explained in Section 2.2, both the marker and the scope do not respond to word boundaries, thus, this layer requires a configuration that stated that the level of granularity of the span is the character. Also, any type of overlap should be allowed. In regard to the features of this layer, they concern the

²² The WoPoss annotation guidelines (Dell’Oro 2019) were designed after the annotation of one text, the *Satyricon* by Petronius, by the three members of the project. The annotations were done independently and later, the results for each modal passage were put in common and thoroughly discussed.

definition of either the marker or the scope. There is a first feature to discriminate between the type of modal unit that is being analysed, and the value of this feature conditions the following elements that the annotator needs to define.

It was coherent to include both the marker and the scope in the same layer because, on the one hand, they share some linguistic features like the type of utterance or the polarity; and on the other hand, one can create a layer of type “relation” which enables the description of the relationship between spans that belong to the same layer. This last aspect was suitable for the creation of the layer *Modal relation*. Therefore, this layer is attached to the *Modal unit* one and it is used to define the abstract relationship between a marker and its scope (or scopes). Among other features, the different types and subtypes of modality are defined in this layer. Since the features to describe an epistemic passage are not pertinent to define, for instance, a dynamic one, different conditions are put into place so features appear in the annotation interface when they are really pertinent.

There are other linguistic elements that are relevant for the study of modality, therefore additional layers are created.

The state of affairs is the representation that is modalized in a modal passage. In general terms, it can be equivalent to the scope of the marker, but other contextual elements might be needed to reconstruct the state of affairs. This is the reason behind the annotation of a third layer named *Participant*. With this layer we identify the participant of the state of affairs even when it is made explicit at a large distance of the scope. Every participant needs to be linked to at least one scope.

Inspired by studies focused on the negation of modal expressions – e.g. van der Auwera (2001) – we decided to explicitly annotate the lexical element that provides a negative meaning to the marker, hence the existence of the layer *Negation*. In the same way that a *Participant* must be linked to a scope, it is mandatory to relate a negation with the relevant marker.

As exposed in section 2.2, both the marker and the scope can be discontinuous. To deal with this circumstance, a chain layer for each one of those elements was created. A chain layer includes both span and relation annotations into a single structural layer. This is an efficient way to deal with discontinuous elements. Of course, to avoid an annotation abuse, these layers are only pertinent when tackling segmented modal units.

Without entering into the functionalities related to the use of knowledge bases, INCEpTION presents two different types of features: link features and primitive ones.

Link features can be used to link one annotation to others. A link feature is the one which allows us to connect a participant with its scope or, for instance, a negative particle with the marker that it is affected by it.

The primitive feature types supported by INCEpTION are string, boolean, integer, and float. Boolean features are displayed in the user interface as a checkbox that can either be marked or unmarked. Integer and float features are displayed using a number field (although for short ranges radio buttons can be displayed instead). String features are filled in using a text field and they can be displayed as a single field or as a text area with multiple rows. However, if a string feature has a tagset associated with it, a drop-down menu appears instead.

Although there are elements in the annotation scheme of WoPoss that could be formalized as a boolean feature, they are defined instead as a string feature whose possible values are “true” or “false”. This is due to the fact that, at the time of writing,²³ the value of a boolean feature cannot be used in the second part of a conditional statement when defining restrictions.

In the WoPoss project, tagsets are created for any feature whose value is a string, except for the element “note”. As mentioned above, features defined with a tagset are displayed as a drop-down list which only allows the choices declared in the tagset to be selected (users cannot type a value).

The element “note” is an open feature in which the annotator can add any relevant information for the annotation that is not formalized in the other features. For instance, they can record any textual problems here for a latter use. During the curation of files done after the annotation these contents can be reviewed and the pertinent editorial modifications can be implemented.

In the previous paragraphs, the notion of constraints was mentioned when explaining the functionalities of the scheme declaration in INCEpTION. Constraints are used to establish conditional features, that is, features that only become available in the annotation interface if another feature has a specific value. Figure 5 shows the constraints that affect the layer *Modal unit*. The statements before the arrow are the conditions and the elements after the arrow are the features and values that appear if the conditions are met. As we can see, the syntax of the constraints is very straightforward.

²³ New versions of INCEpTION are released very frequently and they usually provide new functionalities.

Constraints may also be used for reordering the tags or restricting certain values in a given context.

To sum up, the use of constraints not only speeds up the annotation process, but it enables an annotation less prone to error.

```
Unit {
  Typeofmodalunit="(potential) marker" -> Pertinence="pertinent" |
    Pertinence="not pertinent - modal" | Pertinence="not pertinent - not
    modal";
  Pertinence="not pertinent - not modal" -> Diachrony="post-modal" | Diachrony="pre-modal";
  Pertinence="pertinent" -> Typeofutterance="interrogative" |
    Typeofutterance="non-interrogative" | Polarity="affirmative" |
    Polarity="negative";
  Typeofmodalunit="scope unit" -> Typeofutterance="interrogative" |
    Typeofutterance="non-interrogative" | Polarity="positive" |
    Polarity="negative" | SoAcontrol="+control" | SoAcontrol="-control" |
    SoAcontrol="+/-control" | SoAdynamicity="+dynamic" |
    SoAdynamicity="-dynamic" | SoAdynamicity="+/-dynamic";
}
```

Figure 5. Snippet of the constraint file

Besides the elaboration of the annotation scheme with the definition of constraints to aid (and validate) the annotation process, INCEpTION supports the display of additional documentation. The members and collaborators of WoPoss can access the annotation guidelines through the annotation interface at any moment. In addition, the tagsets employed in the layers created by the automatic analysis²⁴ are also available so these annotations can easily be reviewed (and corrected when needed).

In this section, all the challenges previously expounded in section 2.2 were addressed. Thereby, I showed that INCEpTION provides a suitable environment for the development of our project. This,

²⁴ The automatic annotation is done using a model trained with the Perseus treebank:
https://github.com/UniversalDependencies/UD_Latin-Perseus (accessed on 21/11/2019). For more
 information about the annotation of this resource see
https://universaldependencies.org/treebanks/la_perseus/index.html (accessed on 21/11/2019).

however, does not mean that some improvements would not be welcomed.²⁵

6. Final remarks

To conclude this paper, I underline the importance of working with the appropriate tools when tackling such a complex phenomenon as the semantic analysis of modality.

A tabular formalization, that is, the type of description that can be made in a spreadsheet, is hardly suitable for the definition of notions that are so intrinsic to the context. A correct semantic interpretation requires contextual information that includes the complete morphological and syntactical structure of the linguistic expression containing the modal passage. In this sense, a platform that supports the annotation directly on the text seems to be imperative. The WoPoss approach to modality requires the overview of the interaction between multiple linguistic elements, so a tool that enables the implementation of relations and links between those elements seems to be the most convenient resource. Moreover, it is very practical to use platforms that provide different standards as output formats. This guarantees the sustainability and interoperability of our data as well as their exploitation in different ways without depending on the tool in which the dataset was annotated.

Annotators need an environment in which they can work collaboratively. When evaluating the utility of a given tool, we must consider the learning curve. In general terms, great efforts are made to ensure annotation platforms are as intuitive as possible and made usable by people without a technical background.

²⁵ For instance, it would be useful if, through the constraint rules, one could define the order in which the list of features should appear, so features closely related would be displayed one after the other.

Although using a tool requires some practice, more time should be invested in the issues arisen from the complexity of the annotation scheme than in using the tool itself.

In this paper, I presented a review of the annotation procedure of a specific use case, the diachronic study of modality in the Latin language. Attention was paid to the particular challenges of this project and how a specific annotation platform, INCEpTION, was suitable for the formalization and implementation of a complex annotation scheme.

This annotation platform provides functionalities that we have not explored yet. Future steps of the project envision, on the one hand, the testing of the machine-assisted annotation, and on the other hand, the creation of a knowledge base that would formalize the theoretical framework in an ontology.

References

- Auwers, Johan van der. 2001. 'On the Typology of Negative Modals'. In *Perspectives on Negation and Polarity Items*, edited by J. Hoeksema, H. Rullmann, V. Sánchez-Valencia, and T. van der Wouden, 23–48. Amsterdam and Philadelphia: Benjamins.
- Bermúdez Sabel, Helena. 2018. 'Anotación Multicamada Externa e o Enriquecemento de Edicións Dixitais/Multi-Layered Stand-off Annotation and the Enrichment of Digital Scholarly Editions'. In *Humanidades Digitales*, edited by Déborah González and Helena Bermúdez Sabel, 4–17. Berlin, Boston: De Gruyter. <https://doi.org/10.1515/9783110585421-002>.
- Broek, Thijs van den, Frans Wiering, and Roelof van Zwol. 2005.

‘Backing the Right Horse: Benchmarking XML Editors for Text-Encoding’. *Humanities, Computers and Cultural Heritage*, 78.

Bunt, Harry. 2017. ‘The Semantics of Semantic Annotation’. In *Proceedings of the 21st Pacific Asia Conference on Language, Information and Computation (PACLIC21)*, 13–29. Korean Society for Language and Information.

Dell’Oro, Francesca. 2015. ‘What Role for Inscriptions in the Study of Syntax and Syntactic Change in the Old Indo-European Languages? The Pros and Cons of an Integration of Epigraphic Corpora’. In *Perspectives on Historical Syntax*, edited by Carlotta Viti, 271–90. Studies in Language Companion 169. Amsterdam: Benjamins.

———. 2019. ‘WoPoss Guidelines for Annotation’. Lausanne: Université de Lausanne. DOI: 10.5281/zenodo.3560951.

———. 2020. ‘L’expression de la modalité par des adjectifs : une comparaison entre l’adjectif grec ancien –υμος et l’adjectif latin en –bilis’. *ACME*. Manuscript submitted for publication.

Eckart de Castilho, Richard, Jan-Christoph Klie, Naveen Kumar, Beto Boullosa, and Iryna Gurevych. 2018. ‘Linking Text and Knowledge Using the INCEpTION Annotation Platform’. In *Proceedings of the 14th EScience IEEE International Conference*, 327–28. <http://tubiblio.ulb.tu-darmstadt.de/106983/>.

Eckart de Castilho, Richard, Jan-Christoph Klie, Ute Wichenbach, and Iryna Gurevych. 2019. ‘Beyond WebAnno: The INCEpTION Text Annotation Platform’. CLARIN Annual Conference, Leipzig.

- Ernout, Alfred, and Antoine Meillet. 2001. *Dictionnaire etymologique de la langue latine*. 4th ed. Paris: Klincksieck.
- Ghia, Elisa, Lennart Kloppenburg, Malvina Nissim, Paola Pietrandrea, and Valerio Cervoni. 2016. 'A Construction-Centered Approach to the Annotation of Modality'. In *Proceedings of the 12th Joint ACL-ISO Workshop on Interoperable Semantic Annotation*. Portoroz. https://www.academia.edu/32925215/A_Construction-centered_Approach_to_the_Annotation_of_Modality.
- Glare, P. G. W., ed. 2012. *Oxford Latin Dictionary*. Second Edition. 2 vols. Oxford, New York: Oxford University Press.
- Gwet, Kilem Li. 2014. *Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement among Raters*.
- Hendrickx, Iris, Amália Mendes, and Silvia Mencarelli. 2012. 'Modality in Text: A Proposal for Corpus Annotation'. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, 1805–1812. Istanbul, Turkey: European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2012/pdf/520_Paper.pdf.
- Klie, Jan-Christoph. 2018. 'INCEpTION: Interactive Machine-Assisted Annotation'. In *Proceedings of the First Biennial Conference on Design of Experimental Search & Information Retrieval Systems*, 105–105. <http://tubiblio.ulb.tu-darmstadt.de/106627/>.
- Klie, Jan-Christoph, Michael Bugert, Beto Boullosa, Richard Eckart de Castilho, and Iryna Gurevych. 2018. 'The INCEpTION

- Platform: Machine-Assisted and Knowledge-Oriented Interactive Annotation’. In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, 5–9. Association for Computational Linguistics. <http://tubiblio.ulb.tu-darmstadt.de/106270/>.
- Landragin, Frédéric, Thierry Poibeau, and Bernard Victorri. 2012. ‘ANALEC: A New Tool for the Dynamic Annotation of Textual Data’. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC’12)*, 357–362. Istanbul, Turkey: European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2012/pdf/638_Paper.pdf.
- Meiser, Gerhard. 2010. *Historische Laut- und Formenlehre der lateinischen Sprache*. 3rd ed. Darmstadt: wbg academic.
- Müller, Christoph, and Michael Strube. 2006. ‘Multi-Level Annotation of Linguistic Data with MMAX2’. In *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*, edited by Sabine Braun, Kurt Kohn, and Joybrato Mukherjee, 197–214. Frankfurt a.M., Germany: Peter Lang.
- Nissim, Malvina, Paola Pietrandrea, Andrea Sansò, and Caterina Mauri. 2013. ‘Cross-Linguistic Annotation of Modality: A Data-Driven Hierarchical Model’. In *Proceedings of the 9th Joint ISO - ACL SIGSEM Workshop on Interoperable Semantic Annotation*, 7–14. Potsdam. <https://www.aclweb.org/anthology/papers/W/W13/W13-0501/>.
- Pietrandrea, Paola and *varii auctores*. 2016. ‘Modal – Modèles de

*l'annotation de La Modalité à l'Oral [Corpus]'. ORTOLANG (Open Resources and TOols for LANGuage).
<https://hdl.handle.net/11403/modal>.*

TEI Consortium. 2019a. 'Feature Structures'. In *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/FS.html>.

———. 2019b. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. TEI Consortium. <http://www.tei-c.org/Guidelines/P5/>.

Thesaurusbüro München Internationale Thesaurus-Kommission (n.d.). *Thesaurus linguae Latinae. Editus iussu et auctoritate consilii ab academiis societatibusque diversarum nationum electi*. Berlin, Boston: De Gruyter.

Vaan, Michiel de. 2008. *Etymological Dictionary of Latin: And the Other Italic Languages*. Brill.

Witt, Andreas, and Jens Stegmann. 2009. 'TEI Feature Structures as a Representation Format for Multiple Annotation and Generic XML Documents'. In *Proceedings of Balisage: The Markup Conference 2009*. Montréal, Canada: Balisage Series on Markup Technologies.
<http://www.balisage.net/Proceedings/vol3/html/Stegmann01/BalisageVol3-Stegmann01.html>.

The Semantics of Irish Determiner Phrases – Benjamin Storme

Abstract

This paper studies the semantics of Irish determiner phrases (DPs). It is shown that, contrary to English DPs, Irish DPs cannot have predicative readings, unless they combine with additional morphological material. These results have implications for the crosslinguistic status of the silent type-shifting operators proposed in the literature to account for English DPs' readings.

1. Introduction

In English, singular determiner phrases (DPs) typically denote individuals (type e) or sets of properties (type $\langle\langle e, t \rangle, t \rangle$), as in (1a) and (1b) (Heim and Kratzer 1998). Some authors (Higgins 1973, Williams 1983, Partee 1987) have noted that they can also denote properties, as in (1c), where the DP *a good semanticist* is coordinated with the property-denoting adjective phrase (AP) *smart* (type $\langle e, t \rangle$). Taking types e and $\langle\langle e, t \rangle, t \rangle$ as the basic types of English DPs, Partee (1987) proposed that $\langle e, t \rangle$ readings are made available by string-vacuous type-shifting operations applied to the basic types (see also Winter 2000).

- (1) a. $[_{DP, e}$ This man] came.

- b. [_{DP}, <<e,t>,t> A man] came.
- c. Mary considers John [_{AP}, <e, t> smart] and [_{DP}, <e, t> a good semanticist].

The goal of this paper is to address the question of the status of these string-vacuous type-shifting operations that have been posited for English. Are they universal? If not, could it be because some languages need to express them overtly, by means of a morpheme? In this paper, these two questions will be addressed through an investigation of the semantics of Irish DPs. The Irish data were obtained through elicitation with an Irish native speaker. It will be shown that (i) no string-vacuous type-shifting operation from e or $\langle\langle e,t\rangle,t\rangle$ to $\langle e,t\rangle$ is available in Irish, (ii) DPs can be interpreted predicatively only if combined with additional morphological material, and (iii) among the three morphemes allowing for predicative readings of their DP complements, two can be considered straightforwardly as morphological exponents of a type-shifting operator à la Partee. The consequence will be that combining Partee's (1987) system with an overt/covert parameter constitutes a promising approach to the typology of DPs' interpretations. Additional data from French and Russian point in the same direction.

Section 2 introduces a brief typology of DPs' semantic types based on English. In section 3, semantic tests used to identify DPs' semantic types (in particular, predicative vs. referential) are critically reviewed. Section 4 applies some of these tests to Irish data and show that Irish DPs cannot denote properties unless some overt material is added. Section 5 tackles the question of how this overt material combines semantically with the basic meanings of DPs to yield predicative readings: the predicative *mar* + DP and *ar* + superlative DP constructions are shown to lend themselves to a compositional account compatible with Partee's type-shifting system, whereas the *ina* + DP construction requires a syncategorematic treatment. Section 6 concludes with the implications of this study for the typology of DP meanings.

2. The typology of DPs' semantic types

This section first introduces a working definition for DPs and then presents the range of semantic types they can have in English, as well as the relationships between those, based on Partee's (1987) work.

2.1. A working definition for DPs

In what follows, the working definition in (2) is assumed for DPs. The paper focuses on singular DPs.

- (2) Determiner Phrases (DPs) are constituents that necessarily have among their denotations individuals (type e DPs) or sets of sets of individuals (type $\langle\langle e, t \rangle, t \rangle$).

This definition has both a syntactic and a semantic component. Syntactically, a DP is a constituent and therefore must be identifiable as such by constituency tests. Semantically, a DP denotes either an individual or a set of sets of individuals and therefore must be identified by semantic tests. The need to treat DPs as being able to denote sets of sets of individuals (or equivalently sets of properties) dates back to Montague's work and arguments and references can be found in Heim and Kratzer's (1998) textbook.

As a consequence of the definition in (2), the presence of an overt determiner will not be a necessary criterion for a constituent to qualify as a DP. Languages like Russian do not have determiners, and yet would be treated under this analysis as languages having DPs. For these languages, some authors also report syntactic arguments to support the idea that they have null determiners (see Pereltsvaig 2007:21-23 on Russian), but here it will suffice to have a semantic argument, as long as the expression is a constituent. This question will turn out to be important for Irish. In particular, a

semantic argument will be used to establish that, although the two sentences in (3a) and (3b) seem to have the same syntactic structure (with different word orders) and the same meaning (leaving aside the difference between the two proper names), they actually have quite different logical forms (LFs) and give rise to different inferences: Irish *dochtúir* in (3a) is an entity-denoting DP (type *e*) whereas French *docteur* in (3b) is a property-denoting NP (type $\langle e, t \rangle$).

- (3) a. Is dochtúir Cathal. (Irish)
is doctor Cathal
Cathal is a doctor.
b. Jean est docteur. (French)
Jean is doctor
Jean is a doctor.

2.2. DPs' semantic types in English

The expressions that satisfy the definition of DPs in (2) in English include definite descriptions, like *the King of France*, pronouns, like *he*, proper names, like *John*, quantifier phrases, like *every man*, and wh-phrases, like *which man* (Heim and Kratzer 1998, Karttunen 1977). Some authors (Higgins 1973, Williams 1983, Partee 1987) have noted these expressions can also have additional semantic

types. In particular, they can denote properties, as the definite description *her husband* in (4a), the quantifier phrase *an authority on unicorns* in (4b), and the pronoun *that* in (4c).

- (4) a. Bill has become $[_{DP, \langle e, t \rangle} \text{her husband}]$. (Higgins 1973:225)
b. Mary considers John competent in semantics and $[_{DP, \langle e, t \rangle} \text{an authority on unicorns}]$. (Partee 1987:360)
c. They said she was beautiful and she was $[_{DP, \langle e, t \rangle} \text{that}]$. (Partee 1987:373)

Williams (1983) also notes that they can have even more complex semantic types when built with quantifiers quantifying over properties as *every* and *what* in (5a) and (5b). To make the availability of this reading clearer, sentence (5a) can be paraphrased as follows: at one time or another, every property that is relevant in the context of utterance (for instance being a doctor, being a student, being a linguist...) is such that John has had it. The quantifier *every* quantifies over properties (type $\langle e, t \rangle$) and not over individuals, hence *everything* must be construed with a semantic type $\langle \langle \langle e, t \rangle, t \rangle, t \rangle$. Similarly for sentence (5b), which can be paraphrased as follows: which of the properties that are relevant in the context are such that John has acquired them?

(5) a. At one time or another, John has been [DP, <<<e,t>,t>,t> everything].

(Williams 1983:426)

b. [DP, <<<e,t>,t>,t> What] has John become? (Williams 1983:426)

The range of semantic types available for DPs in English is summarized in Table 1.

	Simple	Complex
Basic types	e	<<e,t>,t>
Derived types	<e,t>	<<<e,t>,t>,t>

Table 1: Basic and derived semantic types for English DPs

The division between basic and derived types follows from the definition of DPs in (2), according to which sequences are identified as DPs if they have the basic types. They can also have additional derived types, but not necessarily. The motivation for distinguishing the derived from the basic types has to do with these readings' markedness within and across languages. Sequences that are traditionally identified as DPs based on their distributional properties denote individuals or sets of individuals in most of their uses in English.

Also, in some languages, the availability of the additional readings is more restricted. For instance, in French, the <e,t> reading of indefinite or definite DPs is available in complement position of *be* but not in small clauses (Roy 2005), as shown by the contrast between (6a) and (6b). Adjectives are available in complement position of small clauses (6c). The derived readings are then attested in French in a subset of the contexts where they are attested in English. Irish will provide further evidence for a distinction between basic and derived types, as derived readings will turn out not to be attested in this language, even in contexts like (6a).

- (6) a. Paul est une référence en matière de licornes.
Paul is a reference in matter of licorns
Paul is a reference on unicorns.
- b. *Marie le considère une référence en matière de licornes.
Mary him considers a reference in matter of licorns
Mary considers him an authority on unicorns.
- c. Marie le considère beau.
Marie him considers beautiful.
Marie finds him beautiful.

2.3. Type-shifting operator from $\langle\langle e,t\rangle,t\rangle$ to $\langle e,t\rangle$

How can these different semantic types be related together? Partee (1987) proposed a general generative mechanism producing type $\langle e,t\rangle$ expressions from type $\langle\langle e,t\rangle,t\rangle$ expressions.²⁶ Consider a domain of individuals D with only four individuals, a black square s_b , a white square s_w , a black circle c_b , and a white circle c_w . According to the classic analysis of determiner a as an existential quantifier, the expression a *square* denotes the set of properties that are true of at least one square. Now consider four basic predicates *square'*, *circle'*, *black'*, and *white'* and a conjunction $\&$. From those, eight predicates denoting non-empty subsets of the domain $D = \{s_b, s_w, c_b, c_w\}$ can be construed, namely *square'*, *circle'*, *black'*, *white'*, *square' & black'*, *square' & white'*, *circle' & black'*, and *circle' & white'*. Among those sets, the set denoted by a *square* will contain all the sets that contain at least one square. This set contains $\{s_b\}$, $\{s_w\}$, $\{s_b, s_w\}$, $\{s_b, c_b\}$, $\{s_w, c_w\}$, $\{s_b, c_b, c_w\}$, etc. The union of the singleton sets in the denotation of a *square*, $\{s_b\}$ and $\{s_w\}$, is the set $\{s_b, s_w\}$, namely the denotation of *square'*, the set of squares in D . The operation that takes the union of all the singleton sets in the denotation of quantifier is a mapping from $\langle\langle e,t\rangle,t\rangle$ denotations to $\langle e,t\rangle$ denotations. In the example above, it takes the meaning of a *square*, a set of sets of individuals, and returns a

²⁶ Readings of DPs with type $\langle\langle\langle e,t\rangle,t\rangle,t\rangle$ will not be further discussed in this paper. Type e DPs can always reanalyzed as type $\langle\langle e,t\rangle,t\rangle$ DPs, where the DP denotes all the properties that are true of an individual instead of the individual himself.

$\langle e, t \rangle$ type-shifted meaning for *a square*, equivalent to the meaning of *square*.

The same operation can be applied to the meaning of *the black circle*. Assume a quantificational analysis of singular definite determiner phrases, where the definite description *the black circle* denotes the set of properties that are true of exactly one black circle (the semantic type is $\langle \langle e, t \rangle, t \rangle$). Then, *the black circle* denotes the set $\{\{c_b\}\}$ (instead of c_b under a referential analysis of singular definite descriptions). Now, applying the same type-shifting operation from above, i.e. taking the union of the singleton sets in the set denoted by *the black circle*, results in the type-shifted meaning $\{c_b\}$ for the expression *the black circle* (with semantic type $\langle e, t \rangle$).

3. Semantic tests

Semantic tests have been proposed to distinguish predicative and nonpredicative readings of DPs. Higgins' (1973) tests to identify DPs' predicative readings are first reviewed and shown to be problematic (section 3.1). Section 3.2 presents a test to identify DPs' predicative readings from Partee (1987) and a test to identify DPs' referential readings from Roy (2005).

3.1. Higgins' tests

Higgins (1973) proposed two tests to identify DPs with predicative readings. The first test applies to definite DPs like *my sister* in (7) and goes as follows: if a definite DP does not convey a uniqueness presupposition in a given context, then it denotes a property in this context. Higgins (1973) used this test to support the claim that *my sister* denotes an individual in (7a) and a property in (7b).

- (7) a. That's [_{DP} my sister]. (type *e* according to Higgins 1973)
b. She's [_{DP} my sister]. (type $\langle e, t \rangle$ according to Higgins 1973)

This test is problematic for two reasons. First, it is unclear that (7a) but not (7b) comes with a uniqueness presupposition. Second, it is hard to use the intuition about the absence of uniqueness presuppositions to decide on the semantic type of the post-copular DP. The question is as follows: if there is no uniqueness presupposition, does it entail that the DP has a predicative reading? There are indeed three semantic types that are compatible with a DP coming without uniqueness presuppositions. It could be a type $\langle e, t \rangle$ DP. It could be a quantifier phrase of type $\langle \langle e, t \rangle, t \rangle$, where the determiner is an existential quantifier (e.g. *a sister of mine*). Or it could be a definite description of type *e* with a

uniqueness presupposition that got cancelled by global accommodation (von Stechow 2004). Cancellation of a uniqueness presupposition for a definite description is possible even in positions where it clearly denotes an individual, as in (8a). (8a) can be used felicitously even if the speaker has more than one sister. The fact that this presupposition is cancelled in this case rather than absent can be shown by the contrast between (8a) and (8b): if a speaker has just a single sister, she will generally prefer to say (8a) over (8b), preferring the expression that maximizes the presupposition (Heim 1991).

- (8) a. My sister came.
b. A sister of mine came.

The same preference for (9a) over (9b) is found, in case the speaker has just one sister. This suggests that the absence of uniqueness presupposition detected by Higgins in (7b) could come from accommodation of the presupposition of a type e expression.

- (9) a. She is my sister.
b. She is a sister of mine.

In Irish, the way sentences (7) and (8) are interpreted is also consistent with the presupposition cancellation account. The Irish informant reported that (10a) does not entail that the speaker has just one sister. She also reported that, among (10a) or (10b), she would prefer uttering (10a) if she had just one sister. One can conclude that Irish *mo* comes with a uniqueness presupposition that can be cancelled.

- (10) a. *Tháinig mo dheirfiúr*
come.past my L-sister
My sister came.
- a. *Tháinig deirfiúr liomsa*
come.past sister with.me-emph
A sister of mine came.

Going back to Higgins' original examples in (7), the same contrast was found in Irish as in English. The informant reported that (11a) does not entail that the speaker has just one sister. She also reported that she would prefer (11a) over (11b) in case she had just one sister. These results show that the absence of uniqueness presupposition does not need to be traced back to a <e, t> reading of *mo dheirfiúr* in (11a), but can be accounted for by an independent mechanism of presupposition cancellation.

- (11) a. Sín í mo dheirfiúr.
it she my L-sister
That's my sister.
- b. Is deirfiúr liomsa í sin.
cop sister with-me.emph she it
It's a sister of mine.

The second test proposed by Higgins (1973) is also problematic because it assumes that DPs denote individuals rigidly. Higgins (1973:226) argues that the fact that the syllogism in (12) does not follow is evidence for the complement of *become* requiring a <e,t> argument (*her* is assumed to have the same referent in both (12a) and (12b)). However, if DPs may denote individuals non rigidly, namely their denotation can change depending on the tense and world of evaluation (Musan 1995), then this problematic syllogism is no longer predicted to hold: the DP *her husband* in (12a) and in (12b) may refer to different people because the time of evaluation of the DP may differ in the two cases.

- (12) a. Bill has become her husband.
b. Her husband is tall.
c. Therefore, Bill has become tall.

A clearer example is provided by sentence (13). If the definite description was interpreted rigidly, with the evaluation time being the moment of utterance, then this sentence, uttered in 2014, would mean that John Kennedy became Obama. The fact that this is not the meaning of (13) can be predicted under a type *e* approach of the meaning of *the president of the USA* where the time of evaluation of the DP shifts with the time of evaluation of the sentence.

(13) John Kennedy became the president of the USA.

3.2. Partee's and Roy's tests

Partee (1987) provides what will be considered as the crucial test to identify predicative readings of DPs: whether they can be coordinated with an adjective phrase. This test is based on the assumption (i) that natural language coordination *and* denotes a function that takes expressions of the same semantic type as arguments and (ii) that adjectives can only be of type $\langle e, t \rangle$. This test establishes that the DP *a good semanticist* can have a type $\langle e, t \rangle$ interpretation in English (14a), but that the corresponding DP

un bon sémanticien ‘a good semanticist’ cannot in the parallel French sentence (14b).²⁷

- (14) a. Mary considers John smart and a good semanticist.
b. *Marie considère Jean intelligent et un bon sémanticien.

To identify referential readings of DPs, the identity question test will be used (a version of this test can be found in Roy 2005). A DP has a referential reading if it can be used to answer a question asking about someone’s identity. In (15), the question *Who is John?* can be answered with a post-copular DP (15b) but not with any verbal phrase (15a). This difference can be accounted for if the answer to an identity question needs to be an entity-denoting expression.

- (15) a. Who is John? - #He works at the hospital.
b. Who is John? - He is my doctor.

²⁷ The fact that the unacceptability of (14b) does not have to do only with the impossibility to conjoin phrases of different syntactic types is conformed by the observation that (ib) is not good either.

- (i) a. Marie considère Jean intelligent.
b. *Marie considère Jean un bon sémanticien.

4. Irish DPs

Section 4.1 looks at the interpretation of DPs in Irish copular sentences - a topic that has already gathered a lot of attention, in particular among syntacticians (McCloskey and Hale 1983, Carnie 1995, Doherty 1996). Section 4.2 focuses on the interpretation of DPs after *become* verbs and section 4.3 on the interpretation of DPs as complements of small clauses.

4.1 Copular sentences

Irish has two elements that correspond to the single *be* verb in English, *is* and *tá*. The first one is labeled identificational, and the second one predication by Doherty (1996). A further complication has to do with the fact that the *is*-copular sentence comes with two different word orders, depending on the nature of what would be the post-copular phrase in English. When it is an indefinite, this constituent comes before the subject. When it is a definite, it comes after the subject. Carnie (1995) calls the first one predication *is* and the second equative *is*. The three types are presented in (16), along with their names borrowed from Doherty (1996) and Carnie (1995). These names should just be taken as labels for now and do not entail a particular semantic analysis. Following Doherty (1996), *tá* is labelled as *be* and *is* as COP (for copula).

- (16) a. Tá Seán ina dhochtúir. (Predicational)
be Sean in.his doctor.
Sean is a doctor.
- b. Is é Seán an dochtúir. (Identificational equative)
COP he Sean the doctor
Sean is the doctor.
- c. Is dochtúir (é) Seán. (Identificational predicational)
COP doctor he Sean
Sean is a doctor.

The identity question test may be used to determine the semantic types of the expressions *ina dhochtúir* in (16a), *an dochtúir* in (16b), and *dochtúir* in (16c). The results are presented in (17): (17a) asks about the identity of John and three answers are considered in (17b)-(17d).

- (17) a. Cé hé Seán?
Who he Sean?
Who is Sean?
- b. #Tá Seán ina dhochtúir.
is Sean in-his doctor
Sean is a doctor.

c. Is é Seán an dochtúir.

cop he Sean the doctor

Sean is the doctor.

d. Is dochtúir (é) Seán.

cop doctor he Sean

Sean is a doctor.

Unsurprisingly, (17b) makes a bad answer and (17c) a good answer to the question in (17a): the expression *ina dhochtúir* has a nonreferential (or nonquantificational) meaning in (17a) and the expression *an dochtúir* has a referential (or quantificational) meaning in (17c). The case of (17d) is more interesting, given the somewhat contradictory label given to this case in (16c) (i.e. identificational predication). The identity question test shows that (17d) patterns with (17c) and thus that *dochtúir* is a referential (or quantificational) DP and not a predication NP. This result is not trivial given the fact that other languages like French treat bare nominals like *dochtúir* in (17d) as predication NPs, as shown by the results of the same test applied to French in (18). (18b) with a bare nominal in postcopular position is not a good answer to the question in (18a). French (18b) patterns like Irish (17b), with the *ina* construction, and French (18d), with an indefinite determiner,

patterns like Irish (17d), with a bare nominal. Unsurprisingly, (18c), with the definite DP, patterns like Irish (17c).

(18) a. Qui est Jean ?

Who is Jean

Who is Jean?

b. #Jean est docteur.

Jean is doctor

Jean is a doctor.

c. Jean est le docteur.

Jean is the doctor

Jean is the doctor.

d. Jean est un docteur.

Jean is a doctor

Jean is a doctor.

These results point to the following conclusion: Irish *is* and *tá* differ in the semantic type of their second argument, *e* or $\langle\langle e,t\rangle,t\rangle$ in the former case and $\langle e,t\rangle$ in the latter case. Doherty (1996) and Carnie (1995) report three counterexamples to this claim (19), where *is* clearly cooccurs with predicative expressions (type $\langle e,t\rangle$).

(19) a. Is as Inis Eoghain é.

cop out-of Inish Owen

He is from Inish Owen.

b. Is ó Bhaile Átha Cliath iad.

cop from Dublin they

They are from Dublin.

c. Is liomsa an t-Alfa Romeo sin.

cop with.me-emph the Alfa Romeo that

I own that Alfa Romeo. (lit. 'That Alfa Romeo is with me')

The informant also accepts the sentences in (19) as grammatical. Also, she judged the same sentences with *tá* instead of *is* in (20) as agrammatical.

(20) a. *Tá sé as Inis Eoghain

be he out-of Inish Owen

He is from Inish Owen.

b. *Tá siad ó Bhaile Átha Cliath.

be they out-of Dublin

They are from Dublin.

c. *Tá an t-Alfa Romeo sin liomsa.

be the Alfa Romeo that with.me-emph

I own that Alfa Romeo.

Also, she judged the sentences in (19) as being good answers to the questions in (21), but not to the questions in (22), confirming Doherty's (1996) and Carnie's (1995) claim that the sentences in (19) are predicative sentences.

(21) a. Cérbh as hé?

Where from he

Where is he from?

b. Cérbh as iad?

where from they

Where are they from?

c. Cé leis?

Who with.him

Whose is it?

(22) a. Cé hé?

Who he

Who is he?

b. Cé siad?

who they

Who are they?

c. Céard an t-Alfa Romeo?

what the Alfa Romeo sin

What is that Alfa Romeo?

Following Doherty (1996) and Carnie (1995), prepositions *as*, *ó*, and *le* are analyzed as lexically-specified exceptions that must combine syntactically with *is* but form type <e,t> expressions with their arguments.

These exceptions put aside, the results of the identity question test provide good evidence that

the basic split between *is* and *tá* is semantic and has to do with semantic type considerations rather than with more subtle semantic distinctions (see Carnie 1995 for arguments against the individual/stage level distinction proposed by Doherty 1996).

The coordination test also suggests that *an dochtúir* and *dochtúir* have the same semantic type and corroborates the results in (17).

The two constituents *ealaontoir* and *an dochtúir is fearr i bParaás* in (23a) and (23b) can be coordinated to yield the grammatical sentence in (23c).

(23) a. *Is ealaontoir é Seán.*

COP artist he Sean

Sean is an artist.

b. S' é Seán an dochtúir is fearr i bParaás.

COP he Sean the doctor cop preferable in Paris

Sean is the best doctor in Paris.

c. Is ealaontoir é Seán agus an dochtúir is fearr i bParaás.

COP artist he Sean and the doctor cop preferable in Paris

John is an artist and the best doctor in Paris.

There is evidence that DPs *an dochtúir* and *dochtúir* cannot be vacuously type-shifted to denote properties. For instance, the informant reported that these constituents cannot be coordinated with an adjective under *tá*. To make sure that this incompatibility was not due to a syntactic incompatibility, the informant was also asked whether it is possible to coordinate an indefinite DP like *dochtúir* with one of the prepositional phrases that are allowed to occur with *is* and with a predicative interpretation (see examples in (19)). This made it possible to keep the syntax constant as much as possible and just play with the semantic types. As expected, this is not possible: (24) is not acceptable.

(24) *Is dochtúir agus as Inis Eoghain é.

Cop doctor and out-of Inish Owen he

He is a doctor and from Inish Owen.

Can pronouns anaphorize a property like *go deas* ‘nice’ in (25), as they do in English (with the propredicate *that*) and in French (with the propredicate *le*)?

(25) Tá Seán go deas.

be Sean ptcl nice

Sean is nice.

The data in (26) and (27) further suggest that pronouns cannot serve as propredicates in Irish. The sentences that show that this option is out in Irish are sentences (26d) and (27d), where the neuter pronoun *sin* cannot anaphorize a property. The other sentences show how Irish speakers anaphorize a property denoted by an adjective: by ellipsis in (26a), (26b), (27b), and (27c) or by anaphorizing the whole sentence with the neuter pronoun *sin* in (26c) and (27a).

(26) a. Tá sé.

Be he

He is so.

b. Tá sé go deimhin.

Be he ptcl sure

He is so indeed.

c. Tá sin/*s'e amhlaidh.

Be it/he so

It/*He is so.

d. *Tá sé sin.

Be he it

He is that.

(27) a. Duirt siad go raibh sí go h'alainn agus b' fhíor sin.

Say-past they that be.past she ptcl beautiful and be.past L-true sin.

They said she was beautiful and that was true.

b. Duirt siad go raibh sí go h'alainn agus bhi sí.

Say-past they that be.past she ptcl beautiful and be.past she.

They said she was beautiful and she was.

c. Duirt siad go raibh sí go h'alainn agus bhi sí go deimhin.

Say-past they that be.past she ptcl beautiful and be.past she ptcl sure.

They said she was beautiful and she was for sure.

d. Duirt siad go raibh sí go h'alainn agus bhi sí sin.

Say-past they that be.past she ptcl beautiful and be.past she it.

They said she was beautiful and she was that.

However, pronouns can occur in the *is* copular sentence where they denote individuals. Pronoun *é* can occupy the same slot as the referential expression *Seán* in sentences (28) and (29).

(28) a. *sin é Seán*

it he Sean

That's Sean.

b. *sin é é*

it he he

That's him.

(29) a. *s' é sin Seán*

Cop he it Sean

That's Sean.

b. *s' é sin é.*

Cop he it he

That's him.

Also, the inanimate interrogative pronoun *céard* cannot quantify over a property in a *tá* copular sentence, as shown by the inacceptability of (30).

(30) **Céard ata Sean?*

What C-is Sean

What is Sean?

Can the universal quantifier phrase *gach rud* ‘everything’ be used to quantify over properties in Irish, as in English (31)?

(31) At one time or another, John has been everything.

To elicit this kind of sentence, the informant was given the following context. In a small village, Sean does all the different jobs that one can think of: he is a doctor, a professor, etc... The informant was asked how she would translate the sentence *Sean is everything* in this context. Sentence (32a) with the quantified DP in postcopular position is not an option. Instead, the informant indicated she would use either (32b) and (32c), which have both very different structures from the English target sentence in (31).

(32) a. *Is é Seán gach rud.

Cop he Sean everyt thing

Sean is everything.

b. Tá chuile cheird ag Seán.

be every occupation to Sean

Sean has every occupation.

- c. Tá láimh ag Seán i ngach rud.
be hand to Sean in every thing
Sean has his hand in everything.

The data seen so far suggest that DPs - definite descriptions, indefinite DPs, pronouns, wh-words, or universally quantified DPs - cannot denote properties (or sets of sets of properties for the quantified DPs) in copular sentences. However, DPs may receive a predicative interpretation when embedded in a larger constituent.

One of these constructions has already been introduced earlier: this is the *ina* construction, exemplified again in (33), for each of the three grammatical persons (1st, 2nd, 3^d) and two numbers (singular, plural). I am a doctor, You are a doctor, etc.).

- (33) a. Tá mé i mo dhochtúir.
I am a doctor.
b. Tá tú in do dhochtúir.
You are a doctor.
c. Tá sí/sé ina dochtúir/dhochtúir.
She/He is a doctor.
d. Tá muid ina ndochtúirí.
We are doctors.

e. Tá sibh in bhur ndochtúirí.

You are doctors.

f. Tá siad ina ndochtúirí.

They are doctors.

Syntactically, the *ina* construction comprises the preposition *i* ‘in’ followed by a possessive DP, as can be seen in the sentences in (34), which are parallel to the sentences in (33).

(34) a. Tá mé i mo theach.

I am in my house.

b. Tá tú in do theach.

You are in your house.

c. Tá sí/sé ina teach/theach.

She/He is in her/his house.

d. Tá muid inár dtithe.

We are in our houses.

e. Tá sibh in bhur dtithe.

You are in your houses.

f. Tá siad ina dtithe.

They are in their houses.

The second construction involving a DP and that can denote a property is the *ar* + superlative DP construction. The superlative DP cannot occur in *tá* copular sentence, unless it is the complement of the preposition *ar* ‘on’, as shown in (35).

(35) Tá sé *(ar) an dochtúir is fearr san áit.

Be he on the doctor cop preferable in-the place

He is the best doctor in the place.

The fact that these two expressions (*ina* + DP and *ar* + superlative DP) denote properties can be shown by their ability to be coordinated with an adjective, apparent in the sentences in (36).

(36) a. Tá sé cliste agus ina dhochtúir.

be he smart and in-his doctor

He is smart and a doctor.

b. Tá sé cliste agus ar an dochtúir is fearr san áit.

be he smart on the doctor cop preferable in-the place

He is smart and the best doctor in the place.

The data in (37) show how the acceptability of *become* sentences varies as a function of the nature of the postcopular constituent. *Become* sentences are good with adjectives, as in (37a), and

prepositional phrases, as in (37d) and (37e), but they are bad with DPs *dochtúir* and *an dochtúir* in (37b) and (37c). These results are compatible with the hypothesis according to which *d'éirigh* requires a type $\langle e, t \rangle$ argument and DPs *dochtúir* and *an dochtúir* cannot be vacuously type-shifted to $\langle e, t \rangle$.

(37) a. D'éirigh Seán bocht/feargach.

arise.past Sean poor/angry

Sean became poor/angry.

b. *D'éirigh Seán dochtúir.

arise.past Sean doctor

Sean became a doctor.

c. *D'éirigh Seán an dochtúir.

arise.past Sean the doctor

Sean became the doctor.

d. D'éirigh Seán ina dochtúir.

arise.past Sean in-his doctor

Sean became a doctor.

e. D'éirigh Seán mar dochtúir.

arise.past Sean as doctor

Sean became a doctor.

Other ways of saying *He became a doctor* are presented in (38). These data point to the same conclusion: a DP can occur in post-copular position when complement of *ina*, as in (38a), or *mar*, as in (38b). In (38c), where the DP *dochtúir* appears without any preposition, the syntactic structure is completely different: the third person pronoun is no longer subject of the sentence but complement of a preposition, and *dochtúir* occupies the subject position.

(40) a. Chuaigh sé ina dhochtúir.

go.past he in-his L-doctor

He became a doctor.

b. Chailigh Seán mar dochtúir.

Become.past Sean as doctor

Sean became a doctor.

c. Rinneadh dochtúir dé.

do.past.impers doctor of-him

He became a doctor (literally, A doctor was made of him).

The Irish verb *faighim* ‘I find’ can be construed with a small clause complement. The second constituent of a small clause can be an adjective (39a) or a prepositional phrase headed by *ina* (39b) or *ar* (39c), but not a DP (39c)-(39d).

(39) a. Faighim é (an-)cliste.

Find.pres.1sg him (ptcl-)smart

I find him (very) smart.

b. Faighim é ina dhea-dhochtúir.

Find.pres.1sg him in-his L-good doctor

I find him a good doctor.

c. *Faighim dea-dochtúir é

find.pres.1sg good-doctor him

I find him a good doctor.

d. Faighim é *(ar) an dochtúir is fearr

Find.pres.1sg him on the doctor cop preferable

I find him the best doctor.

The data in (40) confirm that DPs cannot denote properties (see (40a) and (40b) where DPs cannot be coordinated with adjectives), unless they are included in a larger prepositional phrase (see (40c) and (40d) where the *ina*- and *ar*-prepositional phrases can be coordinated with adjectives). Sentence (40e) shows that the *ina*- and *ar*-prepositional phrases have the same semantic type: the two constituents can be coordinated under *agus* ‘and’.

(40) a. *Faighim é cliste agus dea-dochtúir

find.pres.1sg him smart and good-doctor

I find him smart and a good doctor.

b. *Faighim é cliste agus an dochtúir is fearr.

find.pres.1sg him smart and the doctor cop preferable

I find him smart and the best doctor.

c. Faighim é cliste agus ina dhea-dhochtúir.

find.pres.1sg him smart and in-his good-doctor

I find him smart and a good doctor.

d. Faighim é cliste agus ar an dochtúir is fearr.

find.pres.1sg him smart and the doctor cop preferable

I find him smart and the best doctor.

e. Faighim é ina dhia-cheile agus *(ar) an dochtúir is fearr san áit.

find.pres.1sg him in-his good husband and the doctor cop
preferable in-the place

I find him a good husband and the best doctor in the place.

The same point can be made using another environment allowing small clauses in Irish, namely the complement position of *agus* when it is used as an adjunct to a sentence (Carnie 1995). The DP *dlíodóir* cannot occur in this environment, as shown in (41b), unless it is preceded by a preposition, as in (41c) and (41d). In sentence (41d), as well as in the other sentences with *mar*, there is no

similarity entailment: *mar dlíodóir* does not mean ‘similar to a lawyer’, but ‘being a lawyer’.

(41) a. agus é i gCalifóirnia (Carnie 1995)

and he in California

while being in California

b. *agus é dlíodóir

and he lawyer

while being a lawyer

c. agus é ina dhlíodóir

an he in-his L-lawyer

while being a lawyer.

d. agus é mar dlíodóir

and he as lawyer

while being a lawyer.

5. Analysis

The survey presented in the previous section suggests that DPs cannot denote properties in Irish, contrasting with English DPs. They can denote properties only when they form a complex with a preposition among *i*, *mar*, and *ar*. These differences are summarized in Table 2.

		e	<<e,t>,t>	<e,t>
--	--	---	-----------	-------

English	DP	ok	ok	ok
Irish	DP	ok	ok	*
	P DP	*	*	ok

Table 2: Semantic types available for DPs in English and Irish

An attractive way of accounting for the differences and similarities between the semantics of Irish and English DPs consists in adopting Partee's (1987) analysis and enriching it with an overt/covert parameter: the type-shifting operators posited by Partee would be covert in English and overt in Irish, morphologically realized by the prepositions *i*, *mar*, and *ar*.

In what follows the feasibility of this analysis is evaluated. For each preposition *i*, *mar*, and *ar*, the meaning of the complement DP will first be computed. Then the meaning of the complex [P DP] will be computed following the hypothesis that P is a type-shifting operator à la Partee. Third, the resulting $\langle e, t \rangle$ meaning will be compared to the actual meaning of the P(DP) constituent.

Preposition *mar*.

Assume that *mar* combines with an existentially quantified DP *dochtúir*_{DP} 'a doctor', formed from a noun *dochtúir*_N, whose meaning is defined in (42a). In set-theoretic terms, this DP denotes

the set of properties that are true of at least one doctor, formally defined in (42b). Now assume that *mar* is a Partee-style type shifting operator, namely it takes as argument a set of sets of individuals and returns the union of the singleton sets in this set. The meaning of *mar dochtúir_{DP}* resulting from applying the meaning of *dochtúir_{DP}* to that of *mar* is formally defined in (42c).

- (42) a. $[[\text{dochtúir}_N]] = \{x_e: x \text{ is a doctor}\}$
 b. $[[\text{dochtúir}_{DP}]] = \{P_{\langle e,t \rangle}: \exists x \in [[\text{dochtúir}_N]] [P(x)]\}$
 c. $[[\text{mar dochtúir}_{DP}]] = \cup \{a\} \in [[\text{dochtúir}_{DP}]] \text{ where } \{a\} \text{ is a variable over singleton sets}$

Now, it can be proven that *mar dochtúir_{DP}* denotes the set of doctors. For each individual in the set of doctors to be distinct from the others, there must be at least one property that is only true of him. Going back to the initial example with squares and circles, the black circle is different from all the other individuals by being the only individual that is a member of the *black'* & *circle'* set. This means that, for each doctor in $[[\text{dochtúir}_N]]$, there will be at least one singleton set containing him in the set of sets of individuals $[[\text{dochtúir}_{DP}]]$. As a consequence, the union of those singleton sets will be equal to the set of doctors. This result is represented in (43).

(43) $[[\text{mar } [\text{dochtúir}_{\text{DP}}]]] = [[\text{dochtúir}_{\text{N}}]] = \{x_e: x \text{ is a doctor}\}$

Does the predicted meaning for *mar dochtúir*_{DP} corresponds to its actual meaning? As far as semantic type goes, the prediction is borne out: *mar dochtúir*_{DP} is correctly predicted to be of type $\langle e, t \rangle$. Also, in the examples cited above, *mar dochtúir*_{DP} denotes the set of doctors.

Preposition *ar*.

Assume *ar* combines with a definite DP *an dochtúir is fearr*_{DP} ‘the best doctor’, interpreted as a set of properties, namely the set of properties that are true of the unique individual in the set $[[\text{dochtúir is fearr}_{\text{NP}}]]$. Now assume that *ar* has the same meaning as *mar*. Then, the meaning of $[\text{ar } [\text{an dochtúir is fearr}_{\text{DP}}]]$ is the meaning formalized in (44c).

- (44) a. $[[\text{dochtúir is fearr}_{\text{NP}}]] = \{x_e: x \text{ is the best doctor}\}$ if there is a single best doctor; \emptyset , otherwise
- b. $[[\text{an dochtúir is fearr}_{\text{DP}}]] = \{P_{\langle e, t \rangle} : \lambda x \in [[\text{dochtúir is fearr}_{\text{NP}}]] [P(x)]\}$
- c. $[[\text{ar } [\text{an dochtúir is fearr}_{\text{DP}}]]] = \cup \{a\} \in [[\text{an dochtúir is fearr}_{\text{DP}}]]$ where $\{a\}$ is a variable over singleton sets

Now, we can prove that *ar an dochtúir is fearr_{DP}* denotes the singleton set containing the best doctor. Indeed, all the singleton sets contained in $[[\text{an dochtúir is fearr}_{DP}]]$ will contain the best doctor and no singleton set will contain anyone else - otherwise, then they would not be members of $[[\text{an dochtúir is fearr}_{DP}]]$. Taking the union of those singleton sets will return the singleton set containing the best doctor.

$$(45) \quad [[\text{ar } [\text{an dochtúir is fearr}_{DP}]]] = [[\text{dochtúir is fearr}_{NP}]] = \{x_e: x \text{ is the best doctor}\}$$

The meaning predicted for *ar an dochtúir is fearr_{DP}* matches the actual meaning of this expression: *ar an dochtúir is fearr_{DP}* is correctly predicted to denote a set of individuals and therefore can be coordinated with an adjective; also, it is correctly predicted to come with a uniqueness presupposition and denotes the singleton set containing the best doctor.

Preposition *i*.

Applying the same approach for the preposition *i* will not work: the fact that *i* takes a possessive DP as complement yields an undesirable meaning for the whole construction. If the possessive

pronoun and the sentential subject do not match (say the possessive denotes

an individual y and the sentential subject an individual z), then *ina dhochtúir* should denote the singleton set containing the unique doctor of y , as shown in (46) (g is the assignment function that provides a referent for the occurrence a_1 of the possessive). If the possessive pronoun and the sentential subject corefer, then *ina dhochtúir* should denote the singleton set containing the unique doctor of individual y .

- (46) a. $[[\text{dhochtúir}_N]]^g = \{ \langle x_e, y_e \rangle : x \text{ is a doctor of } y \}$
 b. $[[a_1 \text{ dhochtúir}_{NP}]]^g = \{ x_e : x \text{ is a doctor of } g(1) \}$
 c. $[[a_1 \text{ dhochtúir}_{DP}]]^g = \{ P_{\langle e, t \rangle} : \lambda x \in [[a_1 \text{ dhochtúir}_{NP}]] [P(x)] \}$
 d. $[[\text{in-}[a_1 \text{ dhochtúir}_{DP}]]]^g = \bigcup \{ a \} \in [[a_1 \text{ dhochtúir}_{DP}]]^g \text{ where } \{ a \} \text{ is a variable over singleton sets}$

The meaning derived in (46d) is problematic in two respects. First, *dochtúir* is wrongly predicted to behave as a two-place predicate in this type of sentences. Second, the *ina* construction is wrongly predicted to come with a uniqueness presupposition. This is because the meaning of the whole expression is built from the meaning of its parts, and among them is the possessive pronoun

which comes with a uniqueness presupposition. The mechanism proposed by Partee does not cancel the presuppositions of the embedded DP and this is desirable for treating the *ar an dochtúir is fearr* ‘the best doctor’ case. The fact that possessive pronouns usually come with uniqueness presuppositions even when complement of preposition *i* is shown by the pair of sentences in (47). When a speaker has just one house, he will prefer to utter (47a) over (47b). This can be accounted for by assuming that the possessive *mo* in (47a) comes with a uniqueness presupposition, that the indefinite in (47b) does not, and that speakers follow a ‘Maximize presupposition’ principle.

- (47) a. Tá sé i mo theach
Be he in my L-house
He is in my house.
- b. Tá sé i dteach liomsa
be he in house with-me
He is in a house of mine.

These problematic predictions could be avoided by considering that the possessive pronoun is not interpretable in this construction. However, one would have to account for why possessive pronouns can be interpreted in some contexts and not

in others. Moreover, if the possessive pronoun turns out to be uninterpretable only in this type of construction, the analysis becomes equivalent to a syncategorematic treatment of the *ina* construction. The fact that this construction is a kind of quirk can be shown by the contrast in (48): only the definite DP *mo dhochtúir* ‘my doctor’ can occur in this construction; the indefinite DP *dochtúir liomsa* ‘a doctor of mine’ is ruled out.

- (48) a. Tá mé i mo dhochtúir.
be I in my L-doctor
I am a doctor.
- b. *Tá mé in dochtúir liomsa
be I in doctor with-me.EMPH
I am a doctor.

Also, the fact that the possessive does not get a meaning can be shown by sentence (49), where the argument of *dochtúir* ‘doctor’ is denoted by the prepositional phrase *ag mo dharthair*, and not by the possessive contained in the preposition *ina*.

- (49) Tá sé ina dhochtúir ag mo dharthair
Be he in.his doctor at my L-brother
He is a doctor of my brother.

6. Conclusion

Irish DPs cannot be interpreted predicatively and therefore have a more constrained distribution than their English counterparts. Predicative readings become available only when DPs combine with overt material and, in two out of three cases, this material can be treated as overt versions of Partee's string-vacuous type-shifting operators. Other languages have DPs whose distribution is more restricted than in English: for instance, in Russian and French, additional elements (instrumental case for Russian (50a), and the preposition *comme* in French (50b)) are required to combine with DPs in order to yield predicative readings in some environments.

(50) a. Čexov byl pisatel-em. (Peretsvaig 2007:1-2)

Chekhov was writer-INSTR

Chekhov was a writer.

b. Marie le considère comme une référence en matière de licornes.

Mary him considers as a reference in matter of licorns

Mary considers him an authority on unicorns.

The results of this paper suggest that combining Partee's (1987) system with an overt/covert parameter constitutes a promising approach to the typology of DPs' interpretations.

References

- Carnie, Andrew. 1995. Non-Verbal Predication and Head-Movement. Doctoral Dissertation, MIT.
- Doherty, Cathal. 1996. Clausal Structure and Modern Irish Copula. *Natural Language and Linguistic Theory* 14:1-46.
- von Fintel, Kai. 2004. Would you Believe it? The king of France is Back! Presuppositions and Truth-value Intuitions. Anne Bezuidenhout and Marga Reimer (eds.), *Descriptions and Beyond: An Interdisciplinary Collection of Essays on Definite and Indefinite Descriptions and other Related Phenomena*, 315-341. Oxford: Oxford University Press.
- Heim, Irene, and Angelica Kratzer. 1998. Semantics in Generative Grammar. Malden & Oxford: Blackwell.
- Higgins, Francis Roger. 1973. The Pseudo-cleft Construction in English. Doctoral Dissertation, MIT.
- Karttunen, Lauri. 1977. Syntax and Semantics of Questions. *Linguistics and Philosophy* 1:3-44.

- McCloskey, Jim, and Ken Hale. 1983. On the Syntax of Person-Number Inflection in Modern Irish. *Natural Language and Linguistic Theory* 1:487-533.
- Musan, Carnie. 1995. On the Temporal Interpretation of Noun Phrases. Doctoral Dissertation, MIT.
- Partee, Barbara H. 1987. Noun Phrase Interpretation and Type-shifting Principles. Groenendijk, Jeroen A. G., Dick de Jongh, Martin J. B. Stokhof (eds), *Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers*, 115-141. Foris Publications.
- Pereltsvaig, Asya. 2007. Copular Sentences in Russian: a Theory of Intra-clausal Relations. *Studies in natural language and linguistic theory*. Springer.
- Roy, Isabelle. 2005. Predicate Nominals in Eventive Copular Sentences. *International Journal of Basque Linguistics and Philology* 213-236.
- Williams, Edwin. 1983. Semantic vs. Syntactic Categories. *Linguistics and Philosophy* 6:423-446.
- Winter, Yoad. 2000. DP Structure and Flexible Semantics. Hirotani, M., Coetzee, N., Hall, N. & J-Y. Kim (eds), *Proceedings of the 30th Conference of the North Eastern Linguistic Society*, University of Massachusetts at Amherst.

