Irish Clefting - The LFG Perspective

Sebastian Sulger

A thesis submitted in partial fulfillment of the requirements for the award of

Master of Arts (M.A.)

to the

Universität Konstanz

Supervisors: Prof. Miriam Butt, Prof. Janet Grijzenhout

March 2009
Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Master of Arts (M.A.) is entirely my own work, that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Konstanz, March 5th

Sebastian Sulger

Student ID: 01/565075
Acknowledgments

I thank my supervisors Miriam Butt and Janet Grijzenhout. Thanks also go out to my colleagues in G 219, Tina Bögel and Melanie Seiss, and Veronika Walther. All have proofread major parts of this thesis and made many very valuable comments. Thanks to Melanie for providing help with \LaTeX, and Tina for discussing implementational issues.

I am indebted to Miriam Butt who evoked my interest in LFG and thereby also in syntax and computational linguistics. I am also thankful for the job she provided me with and for the possibility to travel around the world and attend conferences.

Last but certainly not least, I also say thank you to my family. My studies would not have been possible without the huge support I have received from them.
Abstract

This thesis presents an analysis of Irish clefting couched within Lexical-Functional Grammar. Cleft sentences are formed using two syntactic permutations. First, a copula is introduced, taking the clefted phrase or word as a predicate. Second, a relative clause is formed containing the remaining material of the original sentence. This basic pattern is valid across a variety of languages, including Irish. I present different approaches towards copula predication taken within Lexical-Functional Grammar and discuss my analysis of Irish copula constructions. Based on the insights from simple copula predication, I derive my analysis for Irish clefting, claiming that in principle, a parallel syntactic approach for both simple copula clauses and clefting can be assumed. The syntactic analysis of both copula clefting has been implemented using the XLE software; implementational issues are also discussed in the thesis. I proceed by discussing pragmatic aspects of clefting. Cleft sentences are not only interesting from a syntactic point of view, but also are used to separate new information from old information, which makes them a prime example for the application of Information Structure, a theory related to Discourse Pragmatics. In the modular architecture of Lexical-Functional Grammar, additional levels of representation may be added to allow for extra-syntactic annotation. I use the projection of i(nformation)-structure to map strings in the sentence to discourse functions. Applying this type of annotation to clefting, I arrive at a more complete analysis of the form and function of Irish clefts.
Contents

1 Introduction 1
   1.1 What this thesis is about . . . . . . . . . . . . . . . . 1
   1.1.1 A first glance at the data . . . . . . . . . . . . . . . 2
   1.2 Preliminaries . . . . . . . . . . . . . . . . . . . . . . . . 3
      1.2.1 Irish . . . . . . . . . . . . . . . . . . . . . . . . . . 3
      1.2.2 Irish and theoretical syntax . . . . . . . . . . . . . . 4
      1.2.3 Lexical-Functional Grammar . . . . . . . . . . . . . . 5
      1.2.4 Irish and LFG . . . . . . . . . . . . . . . . . . . . . . 8
      1.2.5 XLE . . . . . . . . . . . . . . . . . . . . . . . . . . . 8
      1.2.6 The ParGram project . . . . . . . . . . . . . . . . . . 9

2 The Irish Copula in LFG 12
   2.1 Copula in general . . . . . . . . . . . . . . . . . . . . . . 12
      2.1.1 Some examples . . . . . . . . . . . . . . . . . . . . . 14
   2.2 The Irish copula . . . . . . . . . . . . . . . . . . . . . . . 16
      2.2.1 Two Irish verbs: * is and * bí * . . . . . . . . . . . . . 17
      2.2.2 A systematic overview . . . . . . . . . . . . . . . . . 19
      2.2.3 Copula dropping in Irish . . . . . . . . . . . . . . . . 25
   2.3 Analyses of copula in LFG . . . . . . . . . . . . . . . . . . 26
      2.3.1 The single-tier analysis . . . . . . . . . . . . . . . . 27
      2.3.2 The open complement double-tier analysis . . . . . . 30
      2.3.3 The closed complement double-tier analysis . . . . . 36
2.4 Towards an analysis of Irish copula in LFG ............. 40
  2.4.1 Sample analyses ........................................ 42
  2.4.2 Some remarks on the semantics of Irish copula con-
       structions ................................................. 50

3 Irish Clefting ................................................. 52
  3.1 Clefting: a possible definition .......................... 52
  3.2 The Irish data ............................................. 53
  3.3 Copula constructions, questions, and simple clefts: some sim-
       similarites ................................................. 56
  3.4 Towards an analysis of simple Irish clefts in LFG .......... 61
       3.4.1 A tentative analysis ................................. 61

4 Information Structure and LFG ............................... 70
  4.1 Information Structure .................................... 70
       4.1.1 Information Structure: a possible definition ...... 71
       4.1.2 Main concepts of Information Structure .......... 72
  4.2 Information Structure and Irish Clefting ................. 77
  4.3 Projecting Information Structure into LFG ............... 79
       4.3.1 Previous work on interfacing IS and LFG ........ 79
  4.4 Towards a more complete analysis of Irish clefting ...... 84
       4.4.1 Consequences for a computational grammar ...... 90

5 Conclusion .................................................. 93
  5.1 Future work .............................................. 94

Bibliography .................................................. 95
List of Figures

1.1 Celtic language family tree ........................................ 4
1.2 simple c-structure ................................................. 6
1.3 simple f-structure .................................................. 7
1.4 ParGram c-structure for a transitive English sentence .... 10
1.5 ParGram f-structure for a transitive English sentence .... 11

2.1 Copula construction analyses in LFG ......................... 27
2.2 single-tier analysis in Japanese ................................. 28
2.3 Open double-tier analysis of French copula ................. 31
2.4 Closed double-tier analysis of French copula ............... 32
2.5 Open double-tier analysis of French raising verb .......... 34
2.6 Clashing open double-tier analysis with divergent subjects 35
2.7 Well-formed closed double-tier analysis with divergent subjects 36
2.8 Well-formed closed double-tier analysis of *She seems in a bad mood.* .................................................. 38
2.9 Closed double-tier analysis without surface copula in Arabic 39
2.10 PREDLINK analysis of *Is dochtúir é:* c-structure ........ 43
2.11 PREDLINK analysis of *Is dochtúir é:* f-structure ........ 43
2.12 null-copula analysis of *Dochtúir é:* c-structure .......... 44
2.13 null-copula analysis of *Dochtúir é:* f-structure .......... 44
2.14 ownership sentence: c-structure ............................... 46
2.15 ownership sentence: f-structure ............................... 46
2.16 ambiguous analysis of identification sentence (I) ........ 48
2.17 ambiguous analysis of identification sentence (II) . . . . . . . . 48

3.1 cleft sentence with adverbial in focus position: c-structure . . . 63
3.2 cleft sentence with adverbial in focus position: f-structure . . . 63
3.3 cleft sentence with dropped copula: c-structure . . . . . . . . . 64
3.4 cleft sentence with dropped copula: f-structure . . . . . . . . . 65
3.5 cleft sentence with adjective in focus position: c-structure . . . 66
3.6 cleft sentence with adjective in focus position: f-structure . . . 66
3.7 cleft sentence with NP in focus position: c-structure . . . . . . 67
3.8 cleft sentence with NP in focus position: f-structure . . . . . . 68

4.1 f-structure focus annotation: scope too wide (I) . . . . . . . . 81
4.2 f-structure focus annotation: intended focus . . . . . . . . . . 81
4.3 Russian c-structure: sub-constituent focus . . . . . . . . . . . 82
4.4 f-structure focus annotation: scope too wide (II) . . . . . . . . 82
4.5 i-structure projection in LFG architecture . . . . . . . . . . . 83
4.6 cleft sentence with NP in focus position: f-structure . . . . . . 85
4.7 cleft sentence with focus NP: i-structure . . . . . . . . . . . . . 85
4.8 cleft sentence with adjective in focus position: f-structure . . . 86
4.9 cleft sentence with adjective focus: i-structure . . . . . . . . . 86
4.10 cleft sentence with adverbial in focus position: f-structure . . . 87
4.11 cleft sentence with adverbial focus: i-structure . . . . . . . . . 87
4.12 cleft sentence with partial focus: f-structure . . . . . . . . . . 88
4.13 cleft sentence with partial focus: i-structure . . . . . . . . . . 88
4.14 cleft sentence with complex focus NP: i-structure (I) . . . . . 91
4.15 cleft sentence with complex focus NP: i-structure (II) . . . . . 91
4.16 cleft sentence with complex focus NP: i-structure (III) . . . . 92
**Glossing Conventions**

In this thesis, I use the following conventions for the glossing of examples.

<table>
<thead>
<tr>
<th>1P</th>
<th>First person</th>
<th>Indef</th>
<th>Indefinite</th>
</tr>
</thead>
<tbody>
<tr>
<td>2P</td>
<td>Second person</td>
<td>Masc</td>
<td>Masculine gender</td>
</tr>
<tr>
<td>3P</td>
<td>Third person</td>
<td>Neg</td>
<td>Negative</td>
</tr>
<tr>
<td>AGR</td>
<td>Agreement marker</td>
<td>NegQ</td>
<td>Negative question</td>
</tr>
<tr>
<td>ART</td>
<td>Article</td>
<td>Neut</td>
<td>Neuter gender</td>
</tr>
<tr>
<td>CL</td>
<td>Class prefix</td>
<td>Num</td>
<td>Numeral</td>
</tr>
<tr>
<td>COMP</td>
<td>Complementizer</td>
<td>PART</td>
<td>Particle</td>
</tr>
<tr>
<td>Compar</td>
<td>Comparative</td>
<td>Past</td>
<td>Past tense</td>
</tr>
<tr>
<td>Com</td>
<td>Common case</td>
<td>Perf</td>
<td>Perfect</td>
</tr>
<tr>
<td>CONJ</td>
<td>Conjunction</td>
<td>Pl</td>
<td>Plural</td>
</tr>
<tr>
<td>COP</td>
<td>Copular verb</td>
<td>Poss</td>
<td>Possessive</td>
</tr>
<tr>
<td>Dat</td>
<td>Dative case</td>
<td>Pres</td>
<td>Present tense</td>
</tr>
<tr>
<td>Decl</td>
<td>Declarative</td>
<td>PRON</td>
<td>Pronoun</td>
</tr>
<tr>
<td>Def</td>
<td>Definite</td>
<td>Rel</td>
<td>Relative</td>
</tr>
<tr>
<td>DET</td>
<td>Determiner</td>
<td>Sbj</td>
<td>Subject</td>
</tr>
<tr>
<td>Emph</td>
<td>Emphatic</td>
<td>Sg</td>
<td>Singular</td>
</tr>
<tr>
<td>Fem</td>
<td>Feminine gender</td>
<td>TOP</td>
<td>Topic marker</td>
</tr>
<tr>
<td>Fut</td>
<td>Future tense</td>
<td>VN</td>
<td>Verbal noun</td>
</tr>
<tr>
<td>Imp</td>
<td>Imperfect</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 What this thesis is about

Like other languages, Irish often uses cleft sentences as a means of focusing. Clefting has consequences for the Information Structure of a sentence, in that a particular portion of the sentence is focused and other parts of the sentence are presupposed. Clefting is also a special syntactic construction, involving multiple syntactic processes. Generally, across languages, a copula clause is used, taking the focused material as its predicate. Additionally, the remainder of the original sentence is wrapped in a relative clause following the copula clause; see (1b) for an example from English, which is a cleft sentence created by permutation of (1a).

(1) a. The market is going down.
   
b. It is the market that is going down.

Although there is some variation from this basic pattern, it holds valid for simple clefting constructions cross-linguistically. The thesis is concerned with two quite different aspects of Irish cleft constructions: the syntactic processes involved, and the Information Structure consequences of clefting.
Since the concepts of “copula verbs” and “Information Structure” are central to the notion of clefting, I discuss these issues in chapters 2 and 4. Chapter 4 also describes how the idea of Information Structure can be integrated into the framework of Lexical-Functional Grammar (LFG). In chapter 3, I present some Irish data showing the process of clefting and propose my own analysis and implementation. I explain the term “Information Structure” in chapter 4, and describe its integration into LFG with special emphasis on its application to Irish cleft sentences.

1.1.1 A first glance at the data

To give the reader an impression of the Irish data, I start by giving some simple examples for Irish copula sentences. Sentences (2) – (4) are simple Irish copula sentences. (2) is called a classification sentence, (3a) and (3b) are identification sentences, and (4) expresses ownership. I will discuss the Irish copula system in section 2.2 and present my analysis of Irish copula constructions in section 2.4.

(2) Is dochtúir í.
   COP.Pres doctor she
   ‘She is a doctor.’

(3) a. Is é mo dheartháir an dochtúir.
   COP.Pres AGR.Masc.Sg my brother ART.Def.Sg doctor
   ‘My brother is the doctor.’

   b. Is é an dochtúir an sagart.
      COP.Pres AGR.Masc.Sg ART.Def.Sg doctor ART.Def.Sg priest
      ‘The priest is the doctor.’

(4) Is le Pól an carr.
    COP.Pres with Paul ART.Def.Sg car
    ‘The car belongs to Paul.’ [lit. ‘The car is with Paul.’]
The sentence in (5b) contains a simple Irish cleft, the construction which is the focus of this thesis. It is a syntactic permutation of the sentence in (5a). A copula is used for fronting a certain constituent, and a relative clause is formed containing the remaining material of the original sentence.

\begin{exe}
\begin{tabular}{ll}
(5) & a. Léigh an múinteoir leabhar inné. \\
& read.Past ART.Def.Sg teacher book yesterday \\
& ‘The teacher read a book yesterday.’ \\
\hline
& b. Is é an múinteoir a léigh leabhar inné. \\
& COP.Pres AGR.Masc.Sg ART.Def.Sg teacher read.Past book yesterday \\
& ‘It is the teacher who read a book yesterday.’
\end{tabular}
\end{exe}

1.2 Preliminaries

This thesis discusses a specific structure within Irish syntax, namely the so-called cleft sentences, and how these can be modeled in an implementation couched within the framework of Lexical-Functional Grammar (LFG). Therefore, in this introduction, I make several basic remarks that might be useful to any reader who is not familiar with either LFG or Irish or both.

1.2.1 Irish

Irish is the language the people of Ireland originally spoke and is one of the oldest languages of Europe. Inscriptions on stones in Old Irish were found dating from the 4th century CE. Being an Insular Celtic language, it is closely related to Welsh, Breton, Cornish, and especially Scots Gaelic and Manx. Of these, Cornish and Manx are already regarded as extinct. The Celtic language family tree depicted in Figure 1.1 is adapted from Stenson (1981). The asterisk marks extinct languages.
Irish is the national and first official language of the Republic of Ireland. It was once spoken all over the island of Ireland; however, it currently has the status of being a minority vernacular. Today, it is mostly spoken in specific communities all around the island and particularly in the so-called “Gaeltachtai”, areas that exist mainly in the rural Western part of the island. Here, Irish is still in use on an everyday basis and also learned as a mother tongue. Irish is also an official minority language in the Northern part of the island.

Irish has been on the brink of extinction for a considerable amount of time. The government of Ireland has taken some measures to prevent the Irish language from dying out — the government has been promoting Irish television and radio broadcasting, and books about the language are quite affordable. Irish is also a complementary subject in school.

1.2.2 Irish and theoretical syntax

Despite the fact that there exist some very good sources for Irish (Dónaill, 2005, Ó Siadhail, 1989), work on Irish from a theoretical-syntactic point of view remains rare. Quite a large amount of work has been done by James McCloskey (McCloskey, 1983, 1991). This work is couched within transformational syntax. McCloskey assumes a VP in the deep structure, then con-
structs the surface structure by a set of syntactic rules. Work has also been
done by Nancy Stenson, who gives an impressive overview of the syntax of
Modern Irish (Stenson, 1981). Ó Siadhail (1989) is another, less formal ac-
count of Modern Irish grammar, including an exhaustive account of dialectal
variation. The work of Andrew Carnie has to be mentioned here as well,
who worked extensively on both verb initial languages in general (e.g. Carnie
et al., 2000, 2005) and Irish (e.g. Carnie, 2005, 1997).

1.2.3 Lexical-Functional Grammar

LFG is a syntactic theory that was developed in the late 70’s and early 80’s
by Joan Bresnan and Ronald M. Kaplan (Kaplan and Bresnan, 1982, Bres-
nan, 1978). LFG was invented following the tradition of Chomskyan theories,
developed in the late 50’s. Despite its close connection to this traditional ap-
proach, it was intended as an alternative to the ideas of transformational
grammar. LFG is based on strong mathematical foundations; it was made
clear from the beginning that the theory will have a mathematically well-
defined basis through constraints describing the relation between different
levels of representation. This well-defined mathematical basis makes it pos-
sible to implement computational LFG grammars, which is why LFG is a
formalism that has been and often is used for different applications within
the field of computational linguistics.

C- and F-structure

One important point about LFG is that it does not incorporate any syntactic
movement whatsoever; as a consequence, it does not assume a “deep” struc-
ture from which any actual syntactic order is derived, but only focuses on the
surface word order and the functional syntax of the language (Bresnan, 2001,
Kaplan and Zaenen, 1989). LFG uses two different kinds of representation
to display
1. linear precedence and constituency and

2. syntactic function.\(^1\)

These structures are called c(onsituent) and f(unctional) structure. The c-structure represents the syntactic tree structure. It encodes linear precedence and shows how constituents combine to form a sentence. Figure 1.2 is a very simple example of a c-structure.

![Figure 1.2: simple c-structure](image)

It is important to note that LFG c-structures do not contain any traces, as traces are only employed in theories that involve syntactic movement.

Besides constituency relations, which are encoded at c-structure, LFG also uses functional annotation. For this, attribute-value matrices are used; these are called f-structures in LFG. A simplified f-structure that corresponds to the c-structure in Figure 1.2 is given in Figure 1.3.

The relationship between c- and f-structure is the same as the relationship between form and function of syntactic elements. The different structures model different dimensions of grammar — c-structure shows overt form of expression, and the way phrase structure categories are used to build up this expression; f-structure models the abstract system of function, and how it relates to expression (Bresnan, 2001, Falk, 2001). To shed more light on

\(^{1}\)The term “syntactic function” is used in the context of LFG as a generic term for grammatical functions, such as subject or object; e.g., see Falk (2001).
this distinction between form and function within LFG, I show a simplified\(^2\) example of rule-writing in LFG in (6).

\[
\begin{align*}
\text{S} & \rightarrow \text{NP}: (\uparrow \text{SUBJ})=\downarrow; \text{VP}: \uparrow=\downarrow. \\
\text{VP} & \rightarrow \text{V}: \uparrow=\downarrow; \text{NP}: (\uparrow \text{OBJ})=\downarrow. \\
\text{NP} & \rightarrow \text{D}: \uparrow=\downarrow; \text{N}: \uparrow=\downarrow.
\end{align*}
\]

Three rules like the ones in (6) can account for simple transitive sentences like the one depicted in Figures 1.2 and 1.3. The rules not only provide the information necessary for building up a correct constituent structure, along valid syntactic insights; they also provide functional annotation, in the sense that they annotate the syntactic function for the constituents: the first noun phrase (NP) is given the role of the subject, the verb is annotated as being the head of the structure (as indicated by the “up-equals-down” annotation\(^3\)) and the second NP is assigned the object role. Dalrymple (2001) gives an overview of possible rule annotations in LFG.

\(^2\)This example is obviously not only simplified, but also incomplete, as lexical entries are missing from the example for the sake of descriptiveness.

\(^3\)This annotation basically means that any information provided by the lexical entry of the verb itself is given as it is to the mother node of the tree. In this case, this means that the PRED information, along with the subcategorization frame of the verb, is inserted in the top-level of the f-structure — as seen in Figure 1.3.
LFG is a modular grammar architecture

Another important feature of LFG, especially in the context of this thesis, is its modular design. A grammar composed within LFG is not restricted to c- and f-structure, but can include one or more projections that go beyond the level of syntax and use the input provided by c- or f-structure. Possible extra projections include an optimality-theory projection (o-structure, see Butt et al., 1999a, Bresnan, 2001, Frank et al., 1998), a prosodic projection (p-structure, see Butt and King, 1998, Bögel et al., 2008) and a projection for Information Structure (i-structure, see King, 1997, Butt and King, 1997, O’Connor, 2004). The latter is of particular interest in the context of this thesis; see chapter 4.

1.2.4 Irish and LFG

There is also some work on Irish syntax from the LFG point of view. Ash Asudeh (Asudeh, 2002) presents an analysis of preverbal particles and ad- junction in Irish couched within LFG. It is shown that five seemingly incompatible claims about the syntax of Irish may be unified using some of the basic principles formulated in LFG. Work has also been done by Andrew Carnie (Carnie, 2005) who argues that within LFG, one does not have to assume a VP for Irish, as it is a non-derivational framework and provides mechanisms that render the assumption of an Irish VP unnecessary.

1.2.5 XLE

XLE is a powerful natural language processing tool which is used to parse and generate text using computational grammars based on the LFG notation. The predecessor of the XLE program is called “Grammar Writer’s Workbench”; its latest version is 3.1, documented in Kaplan and Maxwell III (1996). This toolset is not being developed anymore, however, and XLE has pocketed all the functions of the workbench.
XLE is used for writing and testing small toy grammars which aim at modeling certain syntactic phenomena within LFG, but also for large-scale processing of larger amounts of text using broad-coverage grammars (Butt et al., 1999b). In particular, it forms the basis for the grammar development effort within the ParGram project (see section 1.2.6). It comes with a graphical user interface, is written in C and runs on Linux, Solaris and Mac OS X machines (Crouch et al., 2008). The latest XLE release dates January 21st, 2009. Many c- and f-structures that appear in the remaining parts of this thesis were produced by XLE. Crouch et al. (2008) is a documentation for XLE.

1.2.6 The ParGram project

The work presented in this thesis follows work that has been done within the ParGram ("Parallel Grammar") project. The ParGram project is a loose alliance among grammar writers around the world within the framework of LFG. The grammars developed in this project are parallel in the sense that they adhere to a commonly defined set of analyses and features and that they are guided by certain linguistic principles (Butt et al., 1999b,a, 2002). The grammars are also considered deep in the sense that they provide a deep parsing of sentences; i.e. they not only provide a phrase structure for a specific sentence, but also register grammatical functions and grammatical features from a commonly defined set (e.g. case, mood, noun type, voice, and so on).\footnote{As opposed to “deep parsing”, in “shallow parsing” grammars are used to merely annotate superficial phrase structure; an example for this type of annotation is finite-state chunking. For a differentiation, see e.g. Jurafsky and H. Martin (2000).}

ParGram is continuously growing; broad-coverage grammars exist for English, German, French, Japanese; smaller-scale grammars are under development for a variety of languages, including Chinese, Norwegian, Turkish, Urdu, Welsh, among others. The grammars are implemented using the XLE pars-
ing program. Figures 1.4 and 1.5 display a ParGram analysis of the transitive sentence in (7); i.e. the figures show how the broad-coverage English LFG grammar analyzes transitive sentences by means of c- and f-structure. These structures were produced by the XLE program. As can be seen easily, these structures contain much more detail than the simplified structures in Figures 1.2 and 1.3.

(7) The snowplow cleared the road.

Figure 1.4: ParGram c-structure for a transitive English sentence
Figure 1.5: ParGram f-structure for a transitive English sentence

In this introduction, I have discussed the main goals of the thesis and laid out some preliminaries. I have presented the key ideas of Lexical-Functional Grammar, in particular its conception as a modular grammar architecture. The next chapter deals with the Irish copula and its analysis in LFG.
Chapter 2

The Irish Copula in LFG

In this chapter, I first provide a definition of the term “copula”, then present some examples from different languages. I describe the Irish copula system, giving a systematic overview. Then, I present the approaches taken towards analyzing copula constructions within LFG. Then, I turn to my own analysis of the copula in Irish and describe my motivations for choosing this particular analysis.

2.1 Copula in general

Copula are widely considered to form a part of the lexical category “verb” in many languages (e.g. Pustet, 2003, Bußmann, 2002, Declerck, 1988, Stenson, 1981), although there are often several characteristics which differentiate them from the rest of the members of the category (i.e. auxiliaries and full verbs). In other languages, they share more properties with the members of other parts-of-speech, hence are analyzed as belonging to another category (see, for example, Pustet, 2003). Since the copula items are considered to belong to the category of verbs in the language of interest of this thesis, Irish (Ó Siadhaí, 1989, Stenson, 1981), and since we find most languages employ copula verbs, I confine myself to describing copula verbs within this study.
Below, I list some of the more obvious and general functions and characteristics of copula verbs.

- Copula verbs are used to link a subject to its predicate.

   The main function of copula verbs is the linking of a clause’s subject to its predicate (Bußmann, 2002, Declerck, 1988, Pustet, 2003), be it a nominal, adjectival or adverbial predicative. Another term sometimes found in the literature for “copula verb” is “linking verb” (Pustet, 2003), although the concept of linking verbs is much more complex; i.e. the idea of unifying the two classes “copula verb” and “linking verb” is wrong in that the class of linking verbs contains the class of copula verbs, but is much larger (den Dikken, 2006).

- Copula verbs often do not express an action or condition.

   Intuitively, this seems to be the most striking difference between full verbs and copulas. Copula verbs are “semantically empty”; they do not carry a lexical meaning (Bußmann, 2002, Pustet, 2003). They do not add to the compositional meaning of the sentence they occur in; hence, they can only be defined in terms of their syntactic function (Pustet, 2003, Hengeveld, 1992).

- Copula verbs are often morphologically and phonetically poor.

   Related to the characteristics above is the observation that copula verbs are poor, i.e. simple, as far as morphology and phonetics are concerned. In many languages, they may be dropped, i.e. omitted in certain contexts (Pustet, 2003). On the other hand, they are often irregular, i.e. they diverge from the default verbal inflectional paradigm in many languages. Based on the basic assumptions and observations, I give the following definition for copulas in (8) from Pustet (2003).
A copula is a linguistic element which co-occurs with certain lexemes in certain languages when they function as predicates. A copula does not add any semantic content to the predicate phrase it is contained in. (Pustet, 2003, p. 5)

I give some examples for copula verbs from several languages in the following subsection.

2.1.1 Some examples

In this subsection, I present some simple examples for copula constructions across three different languages and explain the function of the copula in the context of the respective sentence.

Spanish has two different copula verbs at its disposition: *ser* and *estar*. Both verbs are used to link a subject to its predicate. See the examples in (9) and (10). Both verbs have other functions as auxiliaries and full verbs as well, but even in their function as copula verbs, they differ in the type of predication they govern. In (9a) and (9b), *ser* as a copula verb is used to link an individual predicate to the subject of the clause. Note that the predication describes inherent properties of the subject. *Estar*, on the contrary, is used for situational predication; i.e. it takes as predicate arguments items which say something about the situation the subject is in (see also, for example, Vera-Morales, 2008). An example is given in (10). This difference in predication has been named “stage level vs. individual level predication” and is described in Ramchand (1997) and Kratzer (1989).

(9) a. *Yo soy* profesora.
   I COP.Pres.1P.Sg teacher.Fem
   ‘I am a teacher.’

b. *Yo soy* inteligente.
   I COP.Pres.1P.Sg intelligent.1P.Sg
   ‘I am intelligent.’
English also employs such a differentiation within the lexicon in terms of aspect predication; one cannot make out the differentiation on the surface, however, since the two copula verbs for the two types of predication are homonymous. Like German, for example, it uses two versions of a single verb, be, for both stage level and individual level predication (Carlsson, 1977); see, however, Jäger (1999), who argues for an unambiguous copula in such languages. See the examples in (11) and (12).

\[(11)\]
\[a. \quad \text{I am} \quad \text{a} \quad \text{teacher.}\]
\[\text{I COP.Pres.1P.Sg ART.Indef.Sg teacher}\]
\[‘\text{I am a teacher.’}\]

\[b. \quad \text{I am} \quad \text{intelligent}.\]
\[\text{I COP.Pres.1P.Sg intelligent}\]
\[‘\text{I am intelligent.’}\]

\[(12)\]
\[\text{I am} \quad \text{tired}.\]
\[\text{I COP.Pres.1P.Sg tired}\]
\[‘\text{I am tired.’}\]

In (11a) and (11b), we find individual level predication, and (12) shows an example of stage level predication in English. In both types of predication, the verbal copula be is used.

The last set of sample data comes from Swahili. The Swahili copula ni is generally regarded as a verbal element, although on its own, it can not carry any verbal inflectional affixes. The reason for this is that when it is used in combination with verbal inflectional categories, ni is replaced by the verbal copula kuwa, which in turn can carry inflectional affixes (Pustet, 2003; Marshad and Suleiman, 1991). Both of the copula combine only with
nominal and adjectival predicatives, not with verbal ones. Apparently, \textit{ni} and its suppletion \textit{kuwa} fulfill both stage level and individual level predication purposes, as is exemplified by (13a), (13b) and (14).

(13) a. h-uyu \textit{ni} n-dege.
    \begin{tabular}{lllllllll}
    h-uyu & \textit{ni} & n-dege. & \\
    this-CL.Sg & COP & Pres & CL.Sg-bird & \\
    ‘This is a bird.’ & \text{(Pustet, 2003, p. 41)} & \\
    
    
    b. Ali \textit{ni} m-wivu.
    \begin{tabular}{lllllllll}
    Ali & \textit{ni} & m-wivu. & \\
    COP & Pres & CL.Sg-jealous & \\
    ‘Ali is jealous.’ & \text{(Pustet, 2003, p. 41)} & \\
    
    
    (14) Ali a-li-kuwa m-wivu.
    \begin{tabular}{lllllllll}
    Ali & a-li-kuwa & m-wivu. & \\
    3P.Sg.Sbj-Past-COP & CL.Sg-jealous & \\
    ‘Ali was jealous.’ & \text{(Pustet, 2003, p. 42)} & \\
    
    
    One important point that is of special interest regarding the presentation of possible LFG analysis of copula sentences (section 2.4) is the fact that in many languages, the copula can be freely omitted (Pustet, 2003, Dalrymple et al., 2004). To make things even more complicated, in some languages the copula \textit{has} to be omitted in certain contexts, while in other contexts it is obligatory. See section 2.2.3 for cases of copula dropping in Irish; see Pustet (2003), pp. 34–39, for a cross-linguistic account of copula dropping.

2.2 The Irish copula

In Irish, we find two verbs which correspond to English \textit{be}. The two verbs are \textit{bí} and \textit{is}. \textit{is} is considered to be the copula verb in Irish, while \textit{bí} is generally considered to be predicative on its own, hence carries a lexical meaning by itself.
2.2.1 Two Irish verbs: *is* and *bí*

The two verbs for *be* in Irish can be distinguished syntactically and semantically. The copula verb *is* generally takes essential and inherent qualities as predicatives; hence the copula predicates are most commonly noun phrases, such as occupation, nationality, group membership and the like. The substantive verb *bí*, on the other hand, takes as predicatives less inherent qualities, such as temporal specifications, location, (temporal) possession etc. (Stenson, 1981).

Below, I give examples of the usage of the substantive verb *bí* and the copula *is*. The predications which *bí* occurs in express information about time, location, possession, personal situation and so on; the copula verb *is* occurs with predicates which refer to inherent properties of the subject. The substantive verb *bí* can also be used to express that some entity *exists*, while this is not possible with the copula verb (see (15a) vs. (15b)).

(15) a. Tá-im.
    be.Pres-1P.Sg
    ‘I am.’ / ‘I exist.’

b. * Is mé.
    COP.Pres I

The difference between (16a) and (16b) in interpretation of possession is that (16a) expresses that at the time of the utterance, the speaker claims that he has a pen with him, while (16b) states that some specific car has the inherent property of belonging to the individual named *Pól*.

(16) a. Tá peann ag-am.
    be.Pres pen at-me
    ‘I have a pen.’ [lit. ‘A pen is at me.’]
b. Is le Póil an carr.
COP.Pres with Paul ART.Def.Sg car
‘The car belongs to Paul.’ [lit. ‘The car is with Paul.’]

Note the word order properties of the two types of predication. While the subject comes first in substantive verb constructions (i.e. sentences containing *bí*, e.g. (16a)), the predicate normally precedes the subject in copular sentences (Ó Siadhail, 1989, Stenson, 1981). In this context, (17a) vs. (17b) are of particular interest.

(17) a. Is é Máirtín an múinteoir.
COP.Pres AGR.Masc.Sg Máirtín ART.Def.Sg teacher
‘Máirtín is the teacher.’

b. * Is é an múinteoir Máirtín.
COP.Pres AGR.Masc.Sg Art.Def.Sg teacher Máirtín

Here, the normal word order of copular sentences is permuted in that the predicate (*an múinteoir*) *follows* the subject (*Máirtín*), and not vice versa. This seems to be due to a rule dictating that in sentences of the form COP – NP – NP, the most referentially exclusive noun phrase directly follows the copula; this rule seems to override the general rule that the predicate NP should follow the copula and precede the subject NP (Ó Siadhail, 1989).

Consider the four examples in (18) for further illustration. (18a) and (18b) describe present and past situations, while (18c) and (18d) refer to qualities which are permanent. The examples further illustrate the difference between the two verbs.

(18) a. Tá na daoine ag damhsa.
be.Pres ART.Def.Pl people.Com.Pl at dance.VN
‘The people are dancing.’
b. Bhí-omar sa choláiste inné.
   be.Past-1P.Pl in.Def college yesterday
   ‘We were in the college yesterday.’

c. Ba dhuine deas é.
   COP.Past man nice he
   ‘He was a nice man.’

d. Is as an Ghearmáin mé.
   COP.Pres from ART.Def.Sg Germany I
   ‘I am from Germany.’

We therefore have to make a clean-cut distinction between the substantive verb and the copula verb. The substantive verb is used for dynamic predications and in contexts that refer to particular situations, while the copula predicates over semantic relations such as equality, class membership, and the like, hence inherent qualities of entities. The next section gives a more systematic overview of the single functions of the Irish copula.

2.2.2 A systematic overview

Since the uses of the Irish copula are often presented in the literature in a confusing manner, I provide a systematic overview. Basically, the copula system in Irish can be distributed across five different uses (Stenson, 1981, The Christian Brothers, 1960).

Classification

Under this use, the so-called classificatory copula sentence is subsumed. It is used to express that some individual belongs to a specific group or class. The structure of sentences of this class is copula – (indefinite) predicate classificatory noun – subject pronoun or noun phrase (see also Ó Siadhail, 1989), as in (19) and (20).
(19) Is múinteoir mé.
   COP.Pres teacher I
   ‘I am a teacher.’

(20) Is múinteoirí iad na fir.
   ‘The men are teachers.’

There is a rule which inserts é ‘he’, í ‘she’ and iad ‘they’ before a definite noun (i.e. proper noun or noun preceded by the article), but apparently not before a pronoun. The rule applies in all of the copula system of Irish (Ó Siadhail, 1989). Stenson (1981) calls this element the “subpredicate”, but does not offer an explanation for its occurrence; nor does Ó Siadhail (1989). I gloss the element as AGR, as it agrees in number and gender with the following NP. See also section 2.4 and the subsections below for further discussion of the subpredicate.

Identification

Subsumed under this category is the identificatory copula sentence. This construction is used to identify an individual for being in a certain position. The structure of this class of copula sentences is copula – definite noun phrase – definite noun phrase (Ó Siadhail, 1989, Stenson, 1981).

(21) Is iad na daoine sin mo thuismitheoirí.
    COP.Pres AGR.Pl ART.Def.Pl people that my parents
    ‘These people are my parents.’ (Stenson, 1981, p. 96)

(22) Is é Pól an dochtúir.
    COP.Pres AGR.Masc.Sg Paul ART.Def.Sg doctor
    ‘Paul is the doctor.’

The subpredicate introduced above is also present here. Since it is introduced with definite NPs, and since in identification sentence, two definite
NPs are equated, most of these sentences contain the subpredicate. Also note the variability in word order in these sentences. Since the status of the two NPs in both cases is definite, it is not always clear which NP is the predicate and which NP is the subject (Stenson, 1981). Note the ambiguity of the sentences in (23).

(23) a. Is é an sagart an dochtúir.
   COP.Pres AGR.Masc.Sg ART.Sg.Def priest ART.Sg.Def doctor
   ‘The priest is the doctor. / The doctor is the priest.’

   b. Is é an dochtúir an sagart.
   COP.Pres AGR.Masc.Sg ART.Sg.Def doctor ART.Sg.Def priest
   ‘The doctor is the priest. / The doctor is the priest.’

In both sentences, both meaning configurations can be expressed (i.e. both NPs can either be subject or predicate). The word order variability is, however, overridden if one of the NPs is realized as a proper noun or a pronoun. In this case, the rule of semantic exclusivity applies, which was already mentioned above in section 2.2.1 and dictates that the semantically more exclusive NP will precede the other NP (Ó Siadhail, 1989). See the two examples below and sections 2.4.1 for further discussion.

(24) a. Is é Pól mo dhochtúir.
    COP.Pres AGR.Masc.Sg Paul my doctor
    ‘Paul is my doctor.’

    b. * Is é mo dhochtúir Pól.
    COP.Pres AGR.Masc.Sg my doctor Paul

The problem of the choice of predicate vs. subject remains, however. It is not clear which one of the NPs in (24a) is the predicate and which one is the subject. The issue of predication in identification sentences is further discussed in section 2.4.2.
Ownership

One can express ownership using a copula statement and the structure COP – PP (le + owner NP) – NP (the owned entity) (Stenson, 1981, Ó Siadhail, 1989). See the example in (25).

(25) Is le Pól an rothar dearg.
COP.Pres with Paul ART.Def.Sg bicycle red
‘The red bicycle belongs to Paul.’

If the NP expressing the owned entity is indefinite, the use of the copula construction is not allowed, and the substantive verb tá has to be used (Ó Siadhail, 1989, Stenson, 1981):

(26) a. * Is le Pól rothar.
COP.Pres with Paul bicycle

b. Tá rothar ag Pól.
be.Pres bicycle at Paul
‘Paul owns a bicycle.’

In this context, the distinction between the copula verb and the substantive verb referring to different kinds of predication, permanent and situational respectively, is somewhat blurred. Ownership is a rather permanent relation. Therefore, it should be expressed by a copula, but is not in examples like (26b) (see also Stenson, 1981, p. 98).

The prepositional phrase in this structure is the only kind of PP that can occur as a predicate in simple copula sentences (Stenson, 1981) (but note the use of PPs in cleft sentences, see the subsection on emphasis below). The use described by Ó Siadhail (1989) is interesting in this context; he shows that the copula can be used in what he calls prepositional phrase construction in examples like (27).
(27) Is maith li-om tae.
    COP.Pres good with-me tea
    ‘I like tea.’

In cases like these, Stenson (1981) argues that the sequence ADJ – P (e.g. maith liom in (27)) constitutes a single predicate. See the subsection on adjective predicates below.

**Emphasis**

The use of copula verbs in constructions of emphasis is the copula use which I focus on in this thesis; it is the use of copula verbs in fronting. Constructions such as these are called *clefting* in the literature. In Irish, these constructions display more freedom regarding the category of the fronted constituent (i.e. the constituent put in the predicate of the copula) (Ó Siadhail, 1989, Stenson, 1981). NPs, adjectives, PPs, adverbs, verbal nouns and other constituents may be fronted. The use of adjectives as the clefted constituent seems to be more marginal; not all adjectives can be clefted freely, and with most of the ones that can be clefted, speakers often prefer an exclamatory or question use (Stenson, 1981). See examples (28d) and (28e) and the next subsection on adjective predicates.

Since a comprehensive account of Irish clefting is given in chapter 3 of the thesis, I just give some simple examples here. Note, however, the general structure of these sentences: the copula standing at the beginning of the sentence, its predicate immediately following the copula, and a relative clause introduced by *a* containing the remaining parts of the original proposition.

(28) a. Is airgead a-tá ag teastáil ua-im.
    COP.Pres money COMP.Rel-be.Pres at lack.VN from-me
    ‘It is money that I need.’ [lit. ‘It is money that is lacking from me.’]
    (Stenson, 1981, p. 99)
b. Is ar an mbóthar a bhuaílfidh mé le-at.
   COP.Pres on ART.Def.Sg road COMP.Rel meet.Fut I with-you
   ‘It is on the road that I will meet you.’ (Stenson, 1981, p. 99)

c. Is aréir a léigh mé an leabhar.
   COP.Pres last night COMP.Rel read.Past I ART.Def.Sg book
   ‘It was last night that I read the book.’

d. Is tinn a-tá sé!
   COP.Pres sick COMP.Rel-be.Pres he
   ‘He is sick!’ [lit. ‘It is sick that he is!’] (Ó Siadhail, 1989, p. 237)

e. Nach dóna a-tá tú!
   COP.Pres.NegQ bold COMP.Rel-be.Pres you
   ‘Aren’t you bold!’

Adjective predicates

The possibility of having adjectives in the predicate of the copula is confined to a small variety of clearly defined cases, most of them being somehow marked and non-productive (Ó Siadhail, 1989, Stenson, 1981); see the example in (29a). Although this is valid for all kinds of copula sentences, the restriction is even stricter with identification type copula sentences than with clefts. Stenson (1981) maintains that the adjective in the copula predicate position is governed by specialized usage; she mentions modals and psychological predicates (Stenson, 1981, p. 101); an example is (29b). Stenson (1981) and Ó Siadhail (1989) also note that the use of the copula in comparison statements such as (29c) is not exclusive to the copula: the substantive verb is used far more often for comparison constructions, as in (29d).
2.2.3 Copula dropping in Irish

In Irish the copula may be dropped in certain environments, and when certain requirements are met. A general rule seems to be that the present tense copula form *is* may be dropped at the beginning of an utterance (Ó Siadhail, 1989). Note that other copula forms, i.e. when the copula is inflected for mood, past tense, negation, interrogation, or when it is embedded in a sentence, may not be deleted from the surface. Also note that the agreement markers *é, í, iad* following *is* are simultaneously deleted. See examples (30) and (31). In a sentence like (31), the copula cannot be dropped, since it is inflected for past tense. (30b) is the deleted form of (30a).

(30) a. Is dochtúir í an bhean.
    COP.Pres doctor AGR.Fem.Sg ART.Def.Sg woman
    ‘The woman is a doctor.’
b. Dochtúir í an bhean.
   doctor AGR.Fem.Sg ART.Def.Sg woman
   ‘The woman is a doctor.’

(31) Ba dochtúir í an bhean.
    COP.Past doctor AGR.Fem.Sg Art.Def.Sg woman
    ‘The woman was a doctor.’

Ó Siadhail (1989) states that copula deletion is most frequent in identification sentences and cleft sentences, the latter of which is at the focus of this thesis. The obvious question raised by copula deletion is: When there is no copula, what in the sentence provides the subcategorization frame for the subject, i.e. what links the predicative element to its subject? I discuss the consequences of copula dropping for the syntactic analysis of copula constructions in section 2.3.

I have presented some data from copula constructions from different languages as well as from Irish, and described syntactic and semantic properties of copula in three languages. I have given a systematic overview of the use of copula in Irish. The next section describes possible analyses of copula in the theory of LFG.

2.3 Analyses of copula in LFG

Within the theory of LFG, there have been numerous attempts at analyzing copula constructions. Since linguists working in LFG constantly try to come up with analyses that can be applied to related phenomena cross-linguistically, and since there is a great deal of variation in the realization of copula across languages, no definitive LFG approach for copula can be made out as yet. The general picture of copula analyses in LFG is depicted in Figure 2.1 (adapted from Attia, 2008).
The question I raise in this section of the thesis therefore is: Is there striking evidence in Irish for choosing one particular analysis for Irish copula constructions? If so, which properties of the Irish copula prohibit the assumption of some other analysis?

2.3.1 The single-tier analysis

In the so-called “single-tier analysis”, the predicate\(^1\) of the copula construction functions as the sentential head; it selects for a subject. Dalrymple et al. (2004) claim that this is the appropriate analysis for copula constructions where the copula itself is optional. In cases where the copula is optional, it is argued that the complement itself can subcategorize over the subject. In examples (32a) and (32b) from Japanese, taken from Dalrymple et al. (2004), the adjective \textit{akai} is assumed to be able to function as a subcategorizing head of the sentence. Note that Dalrymple et al. (2004) propose that a complement should be treated as subcategorizer for a subject \textit{if it can ever be used without the copula}. Hence, both in (32a) and (32b), the adjective is treated as a subcategorizing head. The corresponding f-structure is identical for both sentences and is depicted in Figure 2.2.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure2.1.png}
\caption{Copula construction analyses in LFG}
\end{figure}

(32) a. hon wa akai.
\hspace{3cm} book TOP red
\hspace{3cm} ‘The book is red.’

(Dalrymple et al., 2004, p. 190)

\(^1\)I use the terms “(copula) predicate” and “(copula) complement” for referring to the same thing: the constituent in the copula sentence that says something about the subject.
b. sono hon wa akai desu.

this book TOP red is

‘This book is red.’ (Dalrymple et al., 2004, p. 191)

\[
\begin{align*}
\text{PRED} & \quad \text{‘red} \langle \uparrow \text{SUBJ} \rangle \text{’} \\
\text{SUBJ} & \quad \text{[PRED ‘book’]}
\end{align*}
\]

Figure 2.2: single-tier analysis in Japanese

The problem that arises in Japanese is that, contrary to adjectival complements, the copula may not be omitted in sentences with a nominal complement. Therefore, it seems that nouns in Japanese are not able to subcategorize for a subject, but adjectives are, and the copula may therefore be omitted. For Dalrymple et al. (2004), the conclusion is that within Japanese, one has to assume different analyses for copula constructions that look very similar on the surface, but behave differently as far as copula occurrence is concerned, depending on the category of the predicative.

Copula occurrence also seems to be governed by tense (Dalrymple et al., 2004). Cross-linguistically, it seems that the copula has the tendency to be null in the present tense, but overt in past and future tenses. This is the case in Russian and Arabic, for example; see Dalrymple et al. (2004) and references therein. For such languages, Dalrymple et al. (2004) claim that a unified analysis is desirable, as there does not seem to be any difference in the syntax and semantics of such constructions. They discuss both a single-tier analysis and a closed complement double-tier analysis (see 2.3.3) for these languages. They claim, however, that a unified analysis might not be possible for all languages where copula occurrence is partially controlled by tense.

Attia (2008) argues against the assumption of divergent analyses in cases such as Japanese, where Dalrymple et al. (2004) motivate the differentiation in analysis with the presence vs. absence of the copula. Attia (2008) claims
that one should only use different syntactic analyses for the same function (i.e. copula predication) when there is a compelling motivation. The optionality vs. necessity for the copula to appear on the surface is not enough motivation to assume diverging analyses, whatever the deciding factor may be: category of the predicate constituent or tense. Attia (2008) also provides evidence from Arabic, where it is clear that the predicate cannot be the head because it does not assign case to the subject. However, it remains questionable in how far the statement about Arabic case-assigning predicates holds valid as a cross-linguistic evidence that copula predicates can never be the head of a sentence. In essence, however, the criticism by Attia (2008) is valid and points very strongly into one direction: whether the copula is in the sentence or not does not matter; the syntactic function of the copula predicate is the same in both contexts. Posing different analyses for copula constructions does not capture the generalization of predication across copula sentences, both within a language and cross-linguistically.

Nordlinger and Sadler (2006) argue for the single-tier analysis in a more compelling way. They present examples from the Abkhaz language, a Caucasian language spoken in Georgia and Turkey. Here, the predicate carries morphological features which are normally present on verbs, such as tense, aspect and mood. In (33), a whole sentence is formed containing just one word, namely the inflected predicate.

(33) Da-psó-w-p’.

3P.Sg.Sbj-dead-Pres-Decl
‘He is dead.’

Attia (2008) acknowledges this as a compelling motivation for a single-tier analysis as depicted in Figure 2.2. In cases such as (33), i.e. in a specific language and under specific conditions, the predicate is realized as a subcategorizing head, argues Attia (2008). Supposedly, by the specific conditions, cases are meant where we have a predicate carrying verbal inflection.
2.3.2 The open complement double-tier analysis

In this analysis, the predicate of the copula clause does not function as a sentential head. The head and main predicator of the clause, therefore, is either the copula itself, when it is present, or a null element, when it is not overtly expressed. In this analysis, the copula clause predicate selects for a subject, which is functionally controlled by the main subject of the sentence. In LFG, functional control is realized with an XCOMP (“predicate complement”) attribute; the subject of XCOMP is unified with the main subject of the sentence. The category of XCOMP is an “open” category, meaning that control equations are defined between the subject of the sentence and the XCOMP’s subject. The partial XCOMP f-structure alone does not contain a subject value, i.e. is not complete; it receives the subject’s value by functional control through the linking verb (Bresnan, 2001, Butt et al., 1999b, Lodrup, 2008).

Dalrymple et al. (2004) state that this is the preferable analysis for cases in which the predicate shows agreement with its subject. Dalrymple et al. (2004) give the example of French adjectives in predicate position, agreeing with the subject of the main clause in the same way as verbs do. See (34a) and (34b) and the resulting f-structure in Figure 2.3; the examples are taken from Dalrymple et al. (2004). The index 1 indicates that the XCOMP’s subject is functionally controlled by the subject of the main clause.

(34) a. Il est petit.
   he.Masc.Sg COP.3P.Sg.Pres small.Masc.Sg
   ‘He is small.’ (Dalrymple et al., 2004, p. 195)

b. Elle est petite.
   she.Fem.Sg COP.3P.Sg.Pres small.Fem.Sg
   ‘She is small.’ (Dalrymple et al., 2004, p. 195)

Dalrymple et al. (2004) argue that the agreement is a strong indication for a control relation between the subject and the predicate. They also state
that if one assumes an XCOMP type of analysis, one can write basic lexical
entries for the predicate as in (35).

(35) petite (↑ PRED)='small<↑ (↑ SUBJ)>'
       (↑ SUBJ NUM)=c sg
       (↑ SUBJ GEND)=c fem

The alternative to this solution would be a closed complement analysis;
“closed” because the PREDLINK category is a closed category and there-
fore does not allow functional control (see section 2.3.3). Figure 2.4 shows the
same sentence under the assumption that the adjective does not select for a
subject. The PREDLINK type of analysis shown in Figure 2.4 models the
fact that a particular property is predicated of the subject in a syntactically
reasonable way.
Figure 2.4: Closed double-tier analysis of French copula

The main PRED of the f-structure expresses that a specific property is predicated of the subject *he*, namely the property *small*. This is captured by the PREDLINK function (Butt et al., 1999b).

A lexical entry for *small*, accounting for a structure like Figure 2.4, would have to look like in (36), since agreement in this case occurs outside the adjective’s partial f-structure (i.e. the agreement is not governed locally by the adjective, since the subject is not in the subcategorization frame of the adjective).

(36) petite (↑ PRED)=’small’
    ((PREDLINK ↑) SUBJ NUM)=c sg
    ((PREDLINK ↑) SUBJ GEND)=c fem

This, however, in the view of Dalrymple et al. (2004), is quite cumbersome, and the type of inside-out control equations shown in (36) are not standard in LFG. Since the XCOMP analysis shown in Figure 2.3 and (35) allows a much simpler annotation, and since it is more similar to the analysis of other cases of agreement, such as subject-verb-agreement, the XCOMP analysis should be preferred over the closed complement analysis.

Attia (2008) again counters this analysis. He makes the valid point that in French, the agreement of the copula predicate with the subject is not the same type of agreement verbs show with their subjects. He maintains that
the agreement is not enough evidence to assume that the copula predicate subcategorizes for a subject. He also argues that the non-standard equations shown in (36) can be rewritten using standard notation; in his opinion, one should not prefer one syntactic analysis over the other because of notational preferences. Moreover, Attia (2008) points out that the control equations can be annotated in the syntax of the grammar (i.e. in the phrase-structure rules), and not in the lexical entries of the adjectives, since it seems implausible to generally assume that all adjectives subcategorize over a subject and agree with it.

Another important argument contra the XCOMP open complement analysis comes from a predicational perspective. The analysis in Figure 2.3 is not suitable for copula constructions according to Attia (2008) because this is exactly the way normal subject raising verbs (such as seem, appear) are analyzed in LFG. Assuming an f-structure like Figure 2.3 would therefore mean that there is no difference between copula constructions and subject raising verbs, such as (37).

(37) Elle semble contente.
    she.Fem.Sg seem.3P.Sg.Pres happy.Fem.Sg
    ‘He seems happy.’
For Attia (2008), the problem is that if the structures in Figures 2.3 and 2.5 are assumed, then there is no difference in analysis between copula verbs and simple raising verbs. There is some tradition in generative grammar in treating copula constructions as raising constructions (Moro, 1997, Carnie, 1997, Adger and Ramchand, 2003, Bresnan, 2001). Attia (2008) rejects this view. He claims that one has to differentiate between raising constructions with verbal complements, and raising constructions without verbal complements. Hence, sentences like (38) and (39) have different underlying structures, because the predication is different.

(38) He seems to go. (Attia, 2008)

(39) He seems happy. (Attia, 2008)

The verbal complement in (38) selects for a subject; hence, it is plausible to assume a raising construction. Attia (2008) claims, however, that it does not seem plausible, and also hard to prove, that the adjective phrase in the complement position in (39) (and also adverbial phrases, noun phrases, prepositional phrases) subcategorizes for a subject. The analysis proposed by Attia (2008) treats the two verbs *seem* in (38) and (39) differently; *seem* in (38) is analyzed as an actual raising verb with subject control, while *seem* in (39) is treated as a “quasi-copula”, linking subject and predicate:
(40) The difference between *he seems to go* and *he seems happy* is the same as the difference between *he goes* and *he is happy* which are completely different syntactic structures. The first is a verbal construction while the second is a predicational construction. (Attia, 2008)

Dalrymple et al. (2004) provide the most compelling evidence against an XCOMP analysis, further emphasizing their argumentation for a pluralist approach towards copula constructions. In cases where the post-copular complement already *has* a subject which is different from the subject of the main clause, the closed complement PREDLINK analysis is the preferred analysis. See the examples in (41a) and (41b).

(41) a. The good thing is that he did not throw the snowball.

b. The main goal is (for the student) to succeed in the exam.

If we assume an XCOMP analysis for sentences like these, the result is a clash of PRED values, i.e., because of the control equations, the XCOMP f-structure would contain two subjects which are not unifiable (Attia, 2008, Dalrymple et al., 2004, Butt et al., 1999b). See the illformed f-structure in Figure 2.6 for sentence (41a).

```
[ PRED 'be (↑ XCOMP) (↑ SUBJ)'
  [ SUBJ [ PRED 'thing'
        [ PRED 'throw (↑ SUBJ), (↑ OBJ)']
          [ XCOMP [ SUBJ [ PRED *he/things'
                        [ OBJ [ PRED 'snowball']
]
```

Figure 2.6: Clashing open double-tier analysis with divergent subjects
The problem can be avoided by using a closed complement analysis, i.e. if the copula takes a closed complement, as in the f-structure in Figure 2.7. There are no control equations, and hence the closed PREDLINK f-structure can have its own subject, distinct from the subject of the main clause.

\[
\begin{align*}
\text{PRED} & \quad \text{‘be } \langle (\uparrow \text{PREDLINK}) \rangle \langle \uparrow \text{SUBJ}\rangle' \\
\text{SUBJ} & \quad \begin{cases} 
\text{PRED} \quad \text{‘thing’} \\
\text{PRED} \quad \text{‘throw } \langle (\uparrow \text{SUBJ}), (\uparrow \text{OBJ})\rangle' \\
\text{PREDLINK} \quad \begin{cases} 
\text{SUBJ} & \quad \begin{cases} 
\text{PRED} \quad \text{‘he’} \\
\text{OBJ} & \quad \begin{cases} 
\text{PRED} \quad \text{‘snowball’} \\
\end{cases}
\end{cases}
\end{cases}
\end{cases}
\end{align*}
\]

Figure 2.7: Well-formed closed double-tier analysis with divergent subjects

2.3.3 The closed complement double-tier analysis

The alternative is a so-called closed complement analysis; “closed” because the PREDLINK category is a closed category and therefore does not allow functional control. The PREDLINK type of analysis models the fact that a particular property is predicated of the subject in a syntactically reasonable way. The main PRED of the f-structure expresses that a specific property is predicated of the subject. Exactly this is captured by the PREDLINK function (Butt et al., 1999b).

The closed complement double-tier analysis is a universal LFG analysis for copula constructions according to both Attia (2008) and Butt et al. (1999b). The main advantage of this approach is that since it does not necessarily rely on control equations, it does not have any issues with sentences such as (41a) and (41b). On the other hand, when control equations become
mandatory, i.e.

| mandatory, i.e. when there is obvious agreement (see (34) and (37)), these can also be formulated in a closed complement PREDLINK analysis (Attia, 2008, Butt et al., 1999b), although maybe in a slightly more troublesome way. Therefore, the main arguments for an open complement XCOMP type of analysis presented by Dalrymple et al. (2004) are not enough to motivate a pluralist approach to copula constructions in terms of analyses, since all of the advantageous properties of the XCOMP analysis can be reproduced within a PREDLINK closed complement analysis. Attia (2008) maintains that syntactic features such as the presence and absence of the copula form and the presence and absence of agreement on the predicate do not affect the syntactic function of the predication. As stated above, only because languages like French show agreement on the predicate, copula predication does not necessarily require diverging syntactic analyses.

There are several advantages to the double-tier closed complement PREDLINK analysis. First, it does not matter what kind of constituent the copula complement is; this analysis seems to be the only one that succeeds in providing valid representations for all constituent types, which can take different semantic roles; see Attia (2008) for an overview. Other approaches seem to have problems with this unified approach. Bresnan (2001), assuming that adjectives can subcategorize for subjects, also assumes that nouns and prepositional phrases can do so. To account for this, she proposes to manipulate the PRED of the noun or preposition by means of lexical rules; see the sentences in (42) and the corresponding rules in (43a) and (43b), cited by Lodrup (2008).

(42) a. The pills made him a monster. (Lodrup, 2008, p. 22)
    b. She seems in a bad mood. (Lodrup, 2008, p. 22)

(43) a. ‘monster’ => ‘be-a-monster<↑SUBJ>’ (Lodrup, 2008, p. 22)
    b. ‘in<↑OBJ>’ => ‘be-in-a-state-of<↑SUBJ>, (<↑OBJ>)’
       (Lodrup, 2008, p. 22)
Both Attia (2008) and Lodrup (2008) find this approach problematic, since it not only results in artificial and complex annotation, but also presumes that any PP or NP in a given language can in principle subcategorize for a subject. Dalrymple et al. (2004) and also Rosén (1996) in an earlier paper maintain that this type of analysis is certainly not desirable. Within the closed complement analysis, these problems vanish, since there is no XCOMP f-structure, hence we do not need any subject that is functionally controlled. I give the f-structure for (42b) in Figure 2.8, assuming a closed complement double-tier analysis.

\[
\begin{align*}
\text{PRED} & : \text{‘seem } \langle \uparrow \text{PREDLINK} \rangle \langle \uparrow \text{SUBJ} \rangle \text{’} \\
\text{SUBJ} & : [\text{PRED ‘she’}] \\
\text{PREDLINK} & : [\text{PRED ‘in } \langle \uparrow \text{OBJ} \rangle \text{’}] \\
\text{OBJ} & : [\text{PRED ‘mood’}]
\end{align*}
\]

Figure 2.8: Well-formed closed double-tier analysis of \textit{She seems in a bad mood}.

In sentences where the copula is deleted from the surface, the PREDLINK analysis has great intuitive appeal, since, at f-structure level, it mirrors the juxtaposition of constituents when the copula is missing (Attia, 2008). Since many languages contain be-less sentences (Carnie, 1995), the analysis further gains cross-linguistic appeal. Attia (2008) further argues that the presence vs. absence of the copula itself is a parameter of variation. Since the copula is generally considered as semantically empty, there is no functional distinction to be made between sentences containing the copula and sentences without the copula. We can model the predication in the absence of the copula using a \texttt{null-be} predicator in the LFG rule notation. See the sentence in (44) and the rule in (45); the resulting f-structure is shown in Figure 2.9.
(44) hwa țālībun
he student
‘He is a student.’ (Attia, 2008)

(45) S --> NP: (↑ SUBJ)=↓;
ε: (↑ PRED)='null-be<(↑ SUBJ), (↑ PREDLINK)>'
(↑ TENSE)=pres;
{NP | AP}: (↑ PREDLINK)=↓
(↓ GEND)=(↑ SUBJ GEND)
(↓ NUM)=(↑ SUBJ NUM)

[\[
\begin{array}{c}
PRED \\
\text{‘null-be} \left( \left( \uparrow \text{SUBJ}, \left( \uparrow \text{PREDLINK} \right) \right) \right) \\
\text{SUBJ} \\
\text{PRED} \quad \text{‘he’} \\
\text{PREDLINK} \\
\text{PRED} \quad \text{‘student’} \\
\text{TENSE} \\
\text{pres}
\end{array}
\]
]

Figure 2.9: Closed double-tier analysis without surface copula in Arabic

The rule in (45) looks complicated, and some clarification might be necessary. The rule states that a sentence can be formed out of an NP, annotated for subject function; an empty element with the predicate null-be, selecting for a subject and a PREDLINK and providing the sentence with present tense; and a noun phrase or an adjective phrase as the head of the PREDLINK f-structure, agreeing in gender and number with the subject of the main clause. Note that the tense feature is provided by the empty element ε, which captures the insight that the copula in Arabic can only be omitted in present tense. If a copula was present in the sentence, then the tense feature would be provided by the copula itself.

To sum up the discussion about the different approaches towards copula constructions in LFG, I stress that each one of the possible analyses has its
advantages and disadvantages. While we have to assume a subcategorization frame for predicate elements in the single-tier and open-complement types of analysis which might be less appropriate for some languages than for others, the PREDLINK approach is more neutral in this respect. However, the annotation methods required in connection with the PREDLINK might be a little more complicated than the intuitive functional control annotations of the XCOMP analysis. I think in this discussion it is important to see LFG in the context of parallel (i.e. cross-linguistic) grammar designing and engineering. As long as there are no serious reasons to object the PREDLINK analysis (e.g. the case of Abkhaz discussed above), I maintain that a universally applicable analysis should be favored to increase the cross-linguistic appeal of LFG.

2.4 Towards an analysis of Irish copula in LFG

In this section, I present my own analysis of Irish copula predication in LFG. This analysis provides the basis for the approach to Irish cleft constructions, presented in chapter 4, since simple cleft constructions in Irish rely on copula predication. The section first repeats some of the examples for copula sentences from section 2.2 and takes a closer look at them, then sums up the main properties of copula constructions in Irish, which have been presented in section 2.2. Then, I choose one of the three LFG approaches to copula constructions presented in section 2.3.

Consider the examples in (46), reproduced from section 2.2.2. The examples all have one thing in common. They all contain a copula, a predicate and a subject — nothing more and nothing less.
(46) a. Is dochtúir í an bhean.
   COP.Pres doctor AGR.Fem.Sg ART.Def.Sg woman
   ‘The woman is a doctor.’

b. Is le Pól an carr.
   COP.Pres with Paul ART.Def.Sg car
   ‘The car belongs to Paul.’ [lit. ‘The car is with Paul.’]

c. Ba dhuine deas é.
   COP.Past man nice he
   ‘He was a nice man.’

d. Is maith li-om tae.
   COP.Pres good with-me tea
   ‘I like tea.’

e. Is tinn a-tá sé!
   COP.Pres sick COMP.Rel-be.Pres he
   ‘He is sick!’ [lit. ‘It is sick that he is!’]

Consider the arguments given by Dalrymple et al. (2004) in favor of a divergent analysis for copula constructions across languages and within a certain language. First, agreement is given as an argument for an XCOMP (open complement double-tier) analysis. In Irish, however, the predicate does not show any agreement whatsoever with the subject. Therefore, I maintain that there is no reason concerning agreement to choose XCOMP as a possible copula analysis for Irish.

Second, in cases where the copula is absent, Dalrymple et al. (2004) argue for a special analysis: the single-tier analysis. They present the case of Japanese, where the occurrence of the copula is governed by the category of the predicate. They propose that different analyses have to be assumed depending on the presence or the absence of the copula. Attia (2008) however, as we have seen, maintains that this is merely a case of stylistic variation. In
principle, the predication is the same, no matter if the copula is present or not.

In Irish, like in Russian and Arabic, the factor governing the occurrence of the copula is tense. The copula may be *is* or *null* in the present tense, but its occurrence is mandatory when tensed for future or past. Still, the predication in the sentences does not change in principle, whether the copula is present or not. Therefore, I maintain that the presence of the copula is a means of stylistic variation in the present tense, but must be present in clauses with future or past tense. Additionally, the present tense copula may not be deleted when the negative form is used (Ó Siadhail, 1989). This can also be modeled via c-structure rule annotations.

I follow Butt et al. (1999b) and Attia (2008) in claiming the closed complement double-tier analysis as a universally applicable analysis for copula constructions. The possible variation in the choice of the predicate constituent is immediately reflected by the variation in the PREDLINK f-structure; as described in section 2.2.2, different categories can be head of the predicate, hence head of the PREDLINK f-structure. I present some sample analyses, produced by the Irish LFG grammar developed along this thesis, in the next subsection.

### 2.4.1 Sample analyses

This subsection presents some sample analyses of different copula constructions, using the closed-complement double-tier PREDLINK analysis. For each sentence, I provide explanation where necessary, and show c- and f-structures.

(47) *Is dochtúir é.*

```
COP.Pres doctor he
```

‘He is a doctor.’
In (47), a simple classification sentence, the indefinite NP *dochtúir* is the predicate of the sentence, and the pronoun *é* acts as the subject of the sentence. The copula *is* is present in the sentence and provides the subcategorization frame for the sentence. The PREDLINK analysis of the sentence is shown in the c-structure in Figures 2.10 (c-structure) and 2.11 (f-structure).

![C-Structure](image)

Figure 2.10: PREDLINK analysis of *Is dochtúir é*: c-structure

"Is dochtúir é."

```
PRED 'is<[229:é], [103:dochtúir]'
PRED 'é'
SUBJ 229 CASE com, GEND masc, NUM sg, PERS 3, PRON-TYPEpers
PRED 'dochtúir'
PREDLINK 103 CASE com, GEND masc, NUM sg
1 CLAUSE-TYPE decl, TENSE pres, VTYPE copular
```

Figure 2.11: PREDLINK analysis of *Is dochtúir é*: f-structure

The present copula form *is*, however, may be null in any copula sentence (although the null copula seems to be used more often in identification sentences and clefts, see Ó Siadhail (1989)). So, in the following example, the subcategorization frame of the sentence can not be provided by the lexical entry of the copula, as it is not present, but has to be introduced by an empty category; see the rule in (49).
(48) Dochtúir é.
   doctor he
   ‘He is a doctor.’

Figure 2.12: null-copula analysis of Dochtúir é: c-structure

"Dochtúir é."

PRED  'null-cop<191:é, [1:dochtúir]'
SUBJ  191 CASE com, GEND masc, NUM sg, PERS 3, PRON-TYPEpers
PREDLINK 1 CASE com, GEND masc, NUM sg
256 CLAUSE-TYPEdecl, TENSE pres

Figure 2.13: null-copula analysis of Dochtúir é: f-structure

(49) COPP --> {COP: ↑=↓
   | ε: (↑ PRED)=null-cop<(↑ SUBJ)(↑ PREDLINK)>'
   (↑ TENSE)=pres}
   NP: @PREDLINK;
   PRON: (↑ SUBJ)=↓
   (↓ CASE)=com

44
The rule in (49) basically has the same result as the rule in (45), agreement left aside, which is not necessary to encode in the Irish case. In cases where a copula is present, the lexical entry is used for introducing the subcategorization frame (as indicated by the ↑=↓ annotation on the copula. If the copula is not present, an empty category is used for introducing the frame; in that case, the tense of the clause is set to present, as only the present tense copula may be dropped. Then, the predicative NP follows; the @PREDLINK annotation is a pointer to a macro which introduces the PREDLINK f-structure; the macro is shown in (50).

\[(50)\text{ PREDLINK} = (↑ \text{ PREDLINK})=↓.\]

The result of the annotation is that the NP is assigned the head function of PREDLINK. The COPP rule in (49) is a simplified version of the actual rule in the grammar, as more constituents are allowed in the predicate position.

The next example shows how the grammar deals with the so-called ownership copula expressions. Here, I assume the PP immediately following the copula to be the predicate of the sentence, and the NP after the PP to be the subject.

\[(51)\text{ Is le Pól an carr dearg. }\]
\[
\text{COP.Pres with Paul ART.Sg.Def car red}
\]

‘The red car belongs to Paul.’ [lit. ‘The red car is with Paul.’]
"Is le Pól an carr dearg."

In the next analysis, we have an identification sentence. The copula here is used to equate to definite NPs. The examples from (23) above are repeated here for convenience.

(52) a. Is é an sagart an dochtúir.
   COP.Pres AGR.Masc.Sg ART.Sg.Def priest ART.Sg.Def doctor
   ‘The priest is the doctor. / The doctor is the priest.’

   b. Is é an dochtúir an sagart.
   COP.Pres AGR.Masc.Sg ART.Sg.Def doctor ART.Sg.Def priest
   ‘The doctor is the priest. / The priest is the doctor.’
Two obvious problems arise here. First, we have to deal with the sub-predicate (i.e. the element glossed as AGR in the examples above). Neither Stenson (1981) nor Ó Siadhail (1989) provide an explanation; Stenson (1981) suggests inserting the marker by a low-level syntactic rule, which does not offer any explanation of function. Carnie (1997), in a transformational grammar account of Irish non-verbal predication, mentions the agreement marker and compares it to the agreement found on full verbs in Irish. For the purposes of this thesis, however, I limit myself to mentioning the problem of this agreement marker; I do not have an explanation yet. For the time being, the LFG grammar inserts the agreement marker in a flat way as a sister constituent under COPP. There, it checks if the NP in the predicate position is definite and, in addition, if the whole sentence is a copula sentence. If the AGR marker is not introduced, then a definite NP in PREDLINK is not allowed. See the excerpt from the COPP rule in (53).

\[(53)\ COPP \rightarrow \ldots \{ \text{AGR: } (↑\ PREDLINK\ DEF)=c \ + \]
\[\ (↑\ VTYPE)=c\ \text{copular} \]
\[| \varepsilon: \sim(↑\ PREDLINK\ DEF)\ldots \]

The second problem arises with the choice of the predicate and the subject. Carnie (1997) suggests that neither of them is a predicate, and that both are arguments of the copula. He makes a clear distinction between predicative and equative sentences, the former consisting of a predicate and a subject, and the latter consisting of a null predicate, a subject and 'other' constituents, meaning the second NP of the equation. The status of the second NP is left unclear by Carnie’s (1997) treatment.

In my analysis, I take the view of Stenson (1981) in claiming that the status of both NPs in sentences such as (52a) and (52b) is not clearly defined. These sentences seem to be inherently ambiguous: both of the NPs can either be the predicate of the subject of the sentence. A purely syntactic analysis can not disambiguate between the two readings. Therefore, both readings are
produced by the computational LFG grammar; see the f-structures for both of the readings of (52b) below (Figure 2.16 and 2.17).

"Is é an dochtúir an sagart."

**Figure 2.16: ambiguous analysis of identification sentence (I)**

"Is é an dochtúir an sagart."

**Figure 2.17: ambiguous analysis of identification sentence (II)**

Despite the fact that in principle both readings are acceptable, the predicate-first reading is preferred in the majority of cases (as depicted in Figure 2.17); both Stenson (1981) and Ó Siadhail (1989) note the general rule that the predicate tends to come first. Instances where the subject comes first (Figure 2.16) are understood to be clearly marked. See Stenson (1981) for additional discussion on the word order in identification sentences.

There are, however, other cases in which the word order of the two NPs seems to be much stricter. These are cases in which one of the NPs is realized as a proper noun or as a pronoun (i.e. a ‘narrow’ NP in terms of semantics, meaning a noun which has few possible referents). Proper nouns, being the semantically most exclusive NPs, tend to appear first in the sentence, i.e.
immediately after the copula (Stenson, 1981, Ó Siadhail, 1989). Hence, (54a) is strongly preferred over (54b). The same is valid for pronouns: these precede both proper nouns and definite NPs in identification sentences. Hence, (55a) is strongly preferred over (55b). Note that the agreement marker is also inserted before proper nouns, as these are inherently definite (Stenson, 1981, Ó Siadhail, 1989).

(54) a. Is é Pól an dochtúir.
   COP.Pres AGR.Masc.Sg Paul ART.Def.Sg doctor
   ‘Paul is the doctor.’ / ? ‘The doctor is Paul.’

   b. ??* Is é an dochtúir Pól.
   COP.Pres AGR.Masc.Sg ART.Sg.Def doctor  Paul

(55) a. Is mise an sagart.
   COP.Pres PRON.Emph.1P.Sg ART.Def.Sg priest
   ‘I am the priest.’

   b. ??* Is é an sagart mise.
   COP.Pres AGR.Masc.Sg ART.Def.Sg priest  PRON.Emph.1P.Sg

This, of course, raises several interesting questions, as noted by Stenson (1981), about the basic word order of NPs in identification sentences. What comes first: subject or predicate? How can the concepts of subject and predicate be even defined in copula sentences? Since, as will become clear in chapter 3, the structure of identification sentences is crucial to the understanding of simple Irish clefts, I will shortly sum up the discussion involving subject and predication of the copula given in Stenson (1981).
2.4.2 Some remarks on the semantics of Irish copula constructions

Since there is some variation in the word order of copula sentence (as noted above), we are not able to determine the copula predicate of the copula taking only the linear order of constituents into account. Stenson (1981) mentions interesting examples from The Christian Brothers (1960). They are repeated in (56).

(56) a. Cé hé an bainisteoir? Is é
    who he ART.Sg.Def manager COP.Pres AGR.Masc.Sg
    an Búrcach an bainisteoir.
    ART.Sg.Def Burke ART.Sg.Def manager
    ‘Who is the manager? Burke is the manager.’
    (Stenson, 1981, p. 105)

b. Cé hé an Búrcach? Is é
    who he ART.Def.Sg Burke COP.Pres AGR.Masc.Sg
    an Búrcach an bainisteoir.
    ART.Def.Sg Burke ART.Def.Sg manager
    ‘Who is Burke? Burke is the manager.’
    (Stenson, 1981, p. 105)

The identification sentences which constitute the answers in (56a) and (56b) have the same structure, but they have different stress patterns. The stressed elements are given in italics in each sentence. The Christian Brothers (1960) identify an Búrcach as the predicate of (56a) and an bainisteoir as the predicate of (56b). Why is this, one might ask?

The sentences have different subjects and predicates, depending on the question preceding the answer. The consequence for Stenson (1981) is that, when looking for the copula predicate in a particular sentence, one has to take into account the notion of “new information”. The copula then would always have as its predicate information that is new to the reader or listener.
In most of the copula constructions, it is clear which words convey new information, i.e. what is the predicate, since in most cases, the predicate follows the copula. The specialty of identification sentences is just that either of the NPs may contain the new information, or that the new information may be the identification relation itself. Ó Siadhail (1989, p. 249) also points out the general function of copula sentences to separate new from old information and show their relationship.

Stenson (1981) concludes that, from the viewpoint of semantics, simple copular constructions can be compared to clefts (which I will discuss in the next chapter), in that they are used to separate new from old information.

Summing up, this chapter has dealt with the Irish copula and presented a possible analysis within the LFG framework. I first gave a systematic overview of the Irish copula system, then discussed different approaches towards analyzing copula that have been taken in the past in LFG literature. I have shown why the PREDLINK closed-complement analysis is an appropriate analysis for Irish copula, and presented some analyses using the computational LFG grammar for Irish.
Chapter 3

Irish Clefting

This chapter finally presents my analysis of simple Irish clefts. I first provide a possible definition of the term clefting, then give a collection of sample data from Modern Irish. If we keep in mind the treatment of the Irish copula verb discussed in chapter 2, especially the case of identification sentences, the syntactic analysis of simple Irish clefts becomes straightforward and fits perfectly in the closed complement double-tier PREDLINK analysis introduced in the relevant LFG literature (Attia, 2008, Butt et al., 1999b).

3.1 Clefting: a possible definition

The description of the syntactic process of clefting has given rise to a huge amount of literature within different fields of linguistics - syntax, semantics, pragmatics. Clefting has consequences not only for the syntax of a sentence, but also for the interpretation of a sentence in a particular context.

In most of the literature, a cleft sentence is defined as a focusing construction, composed of a copula verb, the focused word or phrase, and a relative clause which refers to the focused material (Bußmann, 2002, Sornicola, 1996, Declerck, 1988, Collins, 1991, Halvorsen, 1977). The clefted material, immediately following the copula, normally consists of or contains a stressed item,
which is usually called the focus of the sentence (Sornicola, 1996, Collins, 1991, Halvorsen, 1977). The remainder of the sentence is usually called the presupposition or the topic of the sentence. It is important to note that cleft sentences consist of two clauses: one embedding, with the copula as the linking head, and one embedded, with the verb of the original sentence as the head (Bußmann, 2002, Collins, 1991). Also note that it is a language-specific feature whether a semantically empty dummy subject for the copula is inserted or not (Sornicola, 1996).

3.2 The Irish data

In this section, I present a collection of data containing simple Irish cleft sentences. The data shows the general form of clefts and the variability in the choice of the clefted constituent. All of these sentences, however, have the same underlying structure. A constituent is fronted and inserted as the predicate of the copula, and a relative clause is formed containing the rest of the original sentence. These are the basic syntactic facts acknowledged throughout the literature (Stenson, 1981, Ó Siadhail, 1989, The Christian Brothers, 1960). In the following examples, I first give the original (i.e. non-clefted) sentence and then variations of that sentence. That means, for example, (57b) and (57c) are permutations of (57a), and so on.

In (57b), the subject NP is fronted. The example shows that, as with normal copula sentences, the subpredicate has to be inserted between copula and a definite NP in predicate position. It is possible to front both subject NPs and object NPs. (57c) is another permutation of (57a) whereby the adverbial inné ‘yesterday’ is fronted.

(57) a. Léigh an muinteoir leabhar inné.
   read.Past ART.Def.Sg teacher book yesterday
   ‘The teacher read a book yesterday.’
In (58b), the prepositional phrase *ag péinteáil* in (58a) is fronted. The prepositional phrase here expresses a progressive action; in this case, it takes over the grammatical function of a predicate of the substantive verb, i.e. it is not used as an adjunct. In (58c), the object NP *cathaoir* ‘a chair’ is fronted.

(58) a. Bhí an fear ag péinteáil cathaoir inné.
be.Past ART.Sg.Def man at paint.VN chair yesterday
‘The man was painting a chair yesterday.’

(Ó Siadhail, 1989, p. 236)

b. Is ag péinteáil cathaoir a bhí an
COP.Pres at paint.VN chair COMP.Rel be.Past ART.Sg.Def
fear inné.
man yesterday.
? ‘It is painting a chair that the man was (doing) yesterday.’

(Ó Siadhail, 1989, p. 236)
c. Is cathaoir a bhí an fear ag péinteáil inné.

‘It is a chair that the man was painting yesterday.’

(Ó Siadhail, 1989, p. 236)

(59) and (60) show the possibilities of adjective fronting. In (59b) the predicative adjective *te* ‘hot’ is focused; it is not possible to front an attributive adjective out of an NP, as is shown by (60b).

(59) a. Tá sé fuar.

‘It’s cold.’

b. (Níl sé fuar.) Is te a-tá sé.

‘(It’s not cold.) It is hot.’ [lit. *‘It is hot that it is.’]

(60) a. Léigh sí an leabhar dearg.

‘She read a red book.’

b. * Is dearg a léigh sí an leabhar.

As is obvious from examples like (58b) and (59b), English is more rigid concerning the choice of the clefted material. While the English equivalent of (58b) is at least questionable, even with the insertion of the verbal residue *doing*, the literal translation presented in (59b) is completely unacceptable. (60) shows that it is, as in English and other languages, not possible to front an attributive adjective, while (59) shows that it is possible, on the other
hand, to focus adjectives functioning as predicates in the original sentence. The clefting of adjectives often has a marked connotation as a result, as it is the case in (59b). Both Ó Siadhail (1989) and Stenson (1981) note that an exclamatory or question form in these cases is often preferred, as in (61).

(61) Nach te a-tá sé!
    COP.Pres.NegQ hot COMP.Rel-be.Pres it
    ‘Isn’t it hot!’

It is also worth mentioning that only adjectives occurring as predicates of the substantive verb may be clefted; they cannot be clefted when they occur with the copula.

As noted above, clefting is a means of syntactic focusing — the material in the predicate position of the copula is marked as the focus of the sentence. The structure of clefts seems, at first glance, not to have much in common with normal copula constructions. In the next section, however, I will argue along the lines of Stenson (1981) that cleft sentences in fact have much in common with copula constructions — especially with identification sentences.

3.3 Copula constructions, questions, and simple clefts: some similarities

As mentioned above in section 2.4.2, the predication in identificatory copula sentences is similar to the predication found in cleft sentences. If the basic assumption is that the copula in fact always predicates over the focus of the sentence, then some interesting syntactic consequences follow.

Consider the case of predicative adjectives; these can only be clefted if they occur together with the substantive verb tá. This follows naturally from the above assumptions: if the copula already predicates over a focus element, then an additional syntactic focus using a cleft sentences is not possible. But since Irish adjectives are not focus-marked when occurring in the predicate
of the substantive verb, they can be clefted in these cases. This was also noticed by Stenson (1981), who distinguishes Irish adjectives from English adjectives. In English, predicative adjectives are inherently marked for focus, which is why they cannot be fronted in any case.

Next, consider question sentences. A simple constituent question is clearly a device for distinguishing presupposed from new (i.e. focused) information. If I ask the question in (62), then the anticipated referent of the WH-phrase is in the focus, and the remainder of the question is presupposed.

(62) What cleared the road? (The snowplow cleared the road.)

Interestingly, in Irish, the structure of constituent questions is very similar, if not identical to the structure of simple clefts, especially if the fact that the copula form is may be deleted from the surface is taken into consideration, as in (63) and (64).

(63) a. Cé a léigh leabhar?
   who COMP.Rel read.Past book
   ‘Who read the book?’

   b. Is é an múinteoir a léigh leabhar.
      COP.Pres teacher COMP.Rel read.Past book
      ‘It is the teacher who read a book.’

(64) a. Cé-n t-am a chuaigh sí abhaile?
   what-ART.Def.Sg time COMP.Rel go.Past she home
   ‘When did she go home?’

   b. Is a trí a chlog a chuaigh sí abhaile.
      COP.Pres PART.Num three DET.Poss clock COMP.Rel go.Past she home
      ? ‘It is at three o’clock that she went home.’
The syntactic similarity of (63a) to (63b) and (64a) to (64b) is obvious. While Ó Siadhail (1989) notes that the copula is mostly dropped in the context of clefts, Stenson (1981) mentions that

(65) Indeed, one of the commonest situations for copula deletion is the clefted answer to a Wh-question, suggesting that the opposition between old and new information that the copula expresses has already been established by the question itself. (Stenson, 1981, p. 108)

So, a more appropriate answer to the question in (63a) would be (66), where the copula form is deleted from the surface. The similarity between the question phrasing and the cleft becomes even more obvious then.

(66) An m´ uinteoir a l´ eigh leabhar.
   ART.Def.Sg teacher COMP.Rel read.Past book
   ‘It is the teacher who read a book.’

Finally, and most importantly, consider the similarities between identification sentences and clefts with an NP as the clefted constituent. If we try to distinguish these structures from one another by simply looking at the surface, we might conclude that simple cleft sentences differ from identification sentences only in that they seem to lack one of the NPs which are equated.

(67a) is an identification sentence, equating the two NPs as indicated by the bracketing. (67b) is a cleft sentence.¹

(67) a. Is é [mo dhearth´ air] [an fear a bh´ ı tinn inn´ e].
   COP.Pres AGR.Masc.Sg my brother ART.Sg.Def man be.Past sick yesterday
   ‘My brother is the man who was sick yesterday.’

¹Since the status of the subpredicate (the element glossed as AGR, see section 2.2.2) is not clear, I assume for the time being that it attaches in a flat way as a sister to the other NPs.
b. Is é [an fear] [a bhí 
COP.Pres AGR.Masc.Sg ART.Sg.Def man COMP.Rel be.Past 
tinn inné.)
sick yesterday
'It is the man who was sick yesterday.'

Stenson (1981) claims that there is no surface head to the relative clause in sentences such as (67b). Now consider the sentences in (68), which are in fact pseudo-cleft sentences, using the same type of paraphrase as English pseudo-clefts.

(68) a. Is é [mo dheartháir] [an té 
COP.Pres AGR.Masc.Sg my brother ART.Def.Sg one 
a bhí tinn inné.)
COMP.Rel be.Past sick yesterday
'My brother is the one who was sick yesterday.'

b. Is é [an té a bhí tinn 
COP.Pres AGR.Masc.Sg ART.Def.Sg one COMP.Rel be.Past sick 
inné] ná [mo dheartháir.] 
yesterday CONJ my brother
'The one who was sick yesterday is my brother.'

In these pseudo-clefts, the relative clauses do have a nominal head; therefore, they resemble run-of-the-mill identification sentences such as (67a) even more. In fact, (68a) and (67a) are identical in structure. The nominal head inserted in the pseudo-clefts does not add any semantics, but serves as a syntactic head for the relative clause. As can be seen by comparing (68a) with (68b), the sentences also show the same variability in word order as identification sentences. Stenson (1981) notes that when the nominal head does not surface, then the head of the relative clause in a sentence like (67b) is automatically understood to be human. In other sentences, the nominal
head might not be human; then, another provisory head noun surfaces in pseudo-clefts. (69a) and (69b) are examples.²

(69) a. Is é [an carr dearg] [an COP.Pres AGR.Masc.Sg ART.Sg.Def car red ART.Sg.Def rud a cheannaigh sé.] thing COMP.Rel buy.Past he ‘The red car is the thing that he bought.’

b. [S-éa-rd a cheannaigh sé] ná COP.Pres-AGR.Masc.Sg-thing COMP.Rel buy.Past he CONJ [an carr dearg.] ART.Sg.Def car red ‘The thing that he bought is the red car.’

The nominal heads in these sentences are understood, even when they do not surface. The conclusion would be that that cleft sentences such as the one in (67b) are derived by deletion from pseudo-clefts such as the ones in (68) – (69). Since (68) – (69) in turn resemble identification sentences like (67a), the following (underlying) structure for nominal cleft sentences and identification sentences is proposed.

(70) COP [NP]focus [NP – relative clause]subject

In normal cleft sentences, the subject nominal would be deleted from the surface, but other than that, the structures are in fact identical. I conclude that there is a clear functional similarity between copula sentences and simple Irish cleft sentences. I focus on the aspects of focus (or predicate) and topic (or subject) in chapter 4 and discuss the various consequences of clefting for the Information Structure of a sentence.

²Séard in (69b) is obviously a contraction of is é (an) rud; see Stenson (1981, p. 113) and also Ó Siadhail (1989, p. 239).
3.4 Towards an analysis of simple Irish clefts in LFG

The preceding discussion in section 3.2 points strongly into one direction. Irish copula take as their predicate the element which is in the focus of the sentence. Cleft sentences are permutations of simple, copula-less sentences, in which the copula is introduced to be able to focus certain material in the sentence. Without the copula, this does not seem possible. Therefore, an appropriate analysis seems to be one that models these similarities between simple copula constructions and cleft sentences. The analysis I propose for clefts is the same closed-complement PREDLINK analysis presented for copula constructions in chapter 2.

3.4.1 A tentative analysis

In this section, I present some cleft analyses from the computational LFG grammar for Irish implemented along with this thesis and discuss the structures in detail. I discuss the consequences for Irish clefts from assuming the same PREDLINK analysis used for copula constructions. The consequences predict the considerations from the theory of Information Structure concerning clefting, as I illustrate in chapter 4.

Consider the first example, where an adverbial phrase is clefted. The sentence in (71a) is the original sentence, and (71b) the clefted version, focusing the adverbial adjunct *inné*.

(71) a. Léigh múniteoir leabhar inné.
   read.Past teacher book yesterday
   ‘A teacher read a book yesterday.’

b. Is inné a léigh múniteoir leabhar
   COP.Pres yesterday COMP.Rel read.Past teacher book
   ‘It is yesterday that a teacher read a book.’

61
Now, recall that the cleft is used to separate an element which is in focus, and in the copula predicate position, from the remaining elements, which refer to presupposed information and are in the copula subject position; see the structure in (72). The structure is repeated from (70) and slightly modified (i.e. it applies to all clefts, not only those with nominal phrases in focus position).

(72) COP [XP]_{focus/predicate} [relative clause]_{subject}

Since this overall structure is identical to run-of-the-mill copula predication in Irish (see chapter 2), I analyze cleft predication in exactly the same way. Hence, I assume the following structure for (71b):

(73) Is [{inné}]_{focus/predicate} [{a léigh an múinteoir leabhar}]_{subject}

That is, the whole relative clause is analyzed as being the subject to the main clause (i.e. to the copula). The head of the relative clause in this case is the main verb léigh; hence, together with its subcategorization frame, it takes on the subject role in the resulting f-structure. The c-structure for (71b) is given in Figure 3.1, the f-structure in 3.2.
Is inné a léigh múinteoir leabhar.

Figure 3.1: cleft sentence with adverbial in focus position: c-structure

Apart from providing the subject function to the copula, the relative clause is also assigned the function of TOPIC-CLEFT, within the PREDLINK f-structure. The partial TOPIC-CLEFT f-structure contains any elements from the relative clause and is annotated directly in the CP_cleft rule of the grammar.

(75) CP_cleft --> PART: (↑ PREDLINK TOPIC-CLEFT PART-TYPE)=cp
                (↑ PREDLINK TOPIC-CLEFT ... { V: (↑ PREDLINK TOPIC-CLEFT)=↓ ...

63
That is, any of the elements included in the relative clause is inserted in the TOPIC-CLEFT f-structure. Note that the f-structure in Figure 3.2 does not display anything but the verbal head léigh in the TOPIC-CLEFT f-structure, since TOPIC-CLEFT is mapped in its entirety to the subject function of the main clause. The numbering at the left edge of the brackets indicates that the main clause’s subject is unified with the contents of TOPIC-CLEFT. What the notion of topic exactly means and how it relates to the concepts of predicate and focus are among the main topics of chapter 4.

As mentioned above, in the present tense the copula is may often be deleted. This is particularly common in cleft sentences and identification sentences. As with copula predication, a rule similar to the one in (47) providing a copula subcategorization frame has been added to the grammar, so that sentences like (76) can be parsed correctly; see the corresponding structures in Figure 3.3 and 3.4.

(76) Aréir a léigh sí sa leabharlann.
  last_night COMP.Rel read.Past she in.Def library
  ? ‘It was last night that she read in the library.’

Figure 3.3: cleft sentence with dropped copula: c-structure
Next, consider the example in (77b). This is an example of a clefted predicative adjective. The original sentence is given in (77a).

(77) a. Bhí sé tinn.
   be.Past he sick
   ‘He was sick.’

   b. Is tinn a bhí sé.
   COP.Pres sick COMP.Rel be.Past he
   ‘He was sick’ [lit. ‘It is sick that he was!’]

Since this is a predicative adjective, the original sentence in (77a) already uses a PREDLINK structure to link the subject sé to the predicate. Here, the substantive verb bhí is used as a predicate to give information about some particular situation in the past (see also the copula discussion in 2.2.1). Otherwise, the resulting structures in 3.5 and 3.6 are not different from the preceding ones in that they take as the copula subject the entire relative clause.
(78) Is tinn a bhí sé.

![Diagram of cleft sentence with adjective in focus position: c-structure]

Figure 3.5: cleft sentence with adjective in focus position: c-structure

"Is tinn a bhí sé."

![Diagram of cleft sentence with adjective in focus position: f-structure]

Figure 3.6: cleft sentence with adjective in focus position: f-structure

The example discussed next contains an NP as the clefted constituent. Suppose the sentence in (79a) is the original sentence, and the sentence in (79b) is a clefted version of the sentence, focusing the original subject NP *an múinteoir*.

(79) a. Léigh an múinteoir leabhar inné.

`read.Past ART.Def.Sg teacher book yesterday`

‘‘The teacher read a book yesterday.’’
b. Is é an múinteoir a léigh leabhar inné.

'It is the teacher who read a book yesterday.'

The subject NP is missing from the relative clause in (79b) as it has been moved to the copula clause (i.e. the embedding one). The resulting f-structure would be incomplete since the subcategorization frame of léigh is not provided with a subject. By linking the NP in the focus position to the function of the subject of the embedded clause, the problem can be avoided:

$\text{(80)}  \text{XPcleft } \rightarrow \{ \text{NP: } \{ (\uparrow \text{PREDLINK})=\downarrow \\
(\uparrow \text{PREDLINK TOPIC-CLEFT SUBJ})=\downarrow \\
(\downarrow \text{CASE})=\text{com} \ldots \}
$

The resulting c- and f-structure for (79b) are given in Figure 3.7 and 3.8.

$\text{(81)}  \text{Is é an múinteoir a léigh leabhar inné.}$

![Cleft sentence with NP in focus position: c-structure](image)
Note that both subject NPs and object NPs may be focused. The consequence is some ambiguity in the choice of the grammatical function of the focus NP; in principle, the sentence in (79b) could have both of the syntactic readings in (82).

(82) a. It was [the teacher]SUBJ that read [a book]OBJ.

b. It was [the teacher]OBJ that [a book]SUBJ read.

Since, obviously, the reading in (82b) is semantically at least questionable, it must be ruled out; this cannot and should not be done by the syntax, this has to be dealt with by semantic processing. Although cases like (82) might be possible for the syntax to disambiguate\(^3\), there are sentences where the ambiguity is much more difficult to resolve, i.e. where we have genuine ambiguity. (83) is an example.

(83) a. Is é an múinteoir a see.Past ART.Def.Sg child yesterday
chonaic an leabhar inné.
‘It is the teacher who saw the child yesterday.’ or
‘It is the child who saw the teacher yesterday.’

\(^3\)Syntactic disambiguation in this case might be possible by annotating the feature [+/- ANIMATE] and only allowing NPs that have [+ ANIMATE] to receive the subject function of verbs like read.
This chapter dealt with the structure of simple Irish clefts and presents an implementation in LFG using a PREDLINK analysis, which reflects the structural similarities between clefts and copula constructions. In the next chapter, I will present the concept of Information Structure and how it can be integrated in the LFG architecture. Since, for analyzing clefting structures, an annotation for focus and topic is crucial, I maintain that a full analysis of these sentences can only be attained if Information Structure features are integrated.
Chapter 4

Information Structure and LFG

In this chapter I discuss several issues. First, I introduce the concept of “Information Structure” (IS). I briefly discuss the theoretical assumptions behind this concept and the interfaces to other levels of grammar, like syntax and phonology; I use the work of Knud Lambrecht as the main point of reference (Lambrecht, 1994, 1986). Then, I show how the idea of IS can be integrated within the framework of Lexical-Functional Grammar. I discuss work in LFG that has been done in the past on interfacing IS with LFG.

4.1 Information Structure

The linguistic subfield of “Information Structure” is subject to huge variation, both in terminology and theoretical assumptions (Lambrecht, 1994, Bosch and van der Sandt, 1999, Bußmann, 2002). This has two main reasons. First, perhaps more than any other linguistic objective, Information Structure lies at the interface of different other, more established fields of linguistics. Phonology, syntax, semantics, pragmatics — all of these have a significant influence on IS, which is in between all of these.

Second, analyses within IS have to be concerned with the relationship between linguistic form (e.g. word order) and the mental states of speaker
and hearer (e.g. what is known and what is new information). Therefore, any linguist dealing with IS has to take both formal and communicative aspects of language into consideration. In the study of IS, one can neither focus exclusively on the formal aspects, leaving aside pragmatic function and hence approach IS in a way syntacticians approach syntax; nor can one study the function of language in communication and social interaction without taking into account issues of formal nature (Lambrecht, 1994).

4.1.1 Information Structure: a possible definition

Since there is so much variation in the literature about what IS is and what its basic elements are, I provide a definition of the term and describe what its aim is in this chapter. For the purposes of this thesis, I find the following definition to be intuitive enough.

(84) Information Structure: That component of sentence grammar in which propositions as conceptual representations of states of affairs are paired with lexicogrammatical structures in accordance with the mental states of interlocutors who use and interpret these structures as units of information in given discourse contexts. (Lambrecht, 1994, p. 5)

We can already see from this definition that it is in the nature of IS to touch several interfaces within language. It is part of the sentence grammar of a language, hence connected to syntax; it deals with propositions and how they are paired with lexicogrammatical structures and is therefore at the syntax-semantics interface; and this pairing happens according to the mental states of the participants of the discourse, who use the structures as units of information - a clue that IS is connected to pragmatics as well (Lambrecht, 1994, Sornicola, 1996).

Lambrecht (1994) points out an important point in defining what Information Structure is not about. One has to make a clear distinction between
what Lambrecht calls *conversational pragmatics* on the one hand and Information Structure, or *discourse pragmatics* on the other hand. He emphasizes that conversational pragmatics (the “general pragmatics” as studied, for example, by Grice (1975)) is

(85) ...not so much concerned with grammatical structure as with the INTERPRETATION of sentences in relation to conversational settings.

(Lambrecht, 1994, p. 4, emphasis in original)

The remark also reveals the close connection from pragmatics to semantics. Lambrecht (1994) mentions Grice’s *conversational implicature* (Grice, 1975), which essentially deals with those aspects of meaning which cannot be captured by the use of simple truth conditions, but go beyond, taking into account conversational context.

Information Structure as defined by Lambrecht (1994), on the other hand, deals with those discourse functions that are determined by one rather than another possible morphosyntactic or prosodic sentence form. Linguists working on IS therefore ask: Why can one and the same (truth-conditional) meaning be expressed by two or more grammatical structures, e.g. sentence forms or prosodic phrasing?

### 4.1.2 Main concepts of Information Structure

What is Information Structure about? How do linguists working on IS apply the theory to language? In IS, the main goal is to assign *pragmatic relations* to any meaningful unit in a given sentence. Linguists working in the field can choose from a set of such functions. Since defining this set is mandatory for any work done within or in connection with IS, and since there is a lot of variation in the terminology, I define the set of possible relations required for the work in this thesis below. I use linguistic examples to explain the terms.
Topic

In Lambrecht (1994), the notion of *topic* is defined as the thing which the proposition of a sentence (i.e. the compositional semantic content of the sentence) is about. Of course, this seems to be a very loose definition, since the concept of “being about something” is not neatly defined. Lambrecht (1994) further characterizes the concept of topic as follows. In a certain discourse, one particular referent within a sentence is interpreted as being the topic of the sentence, if the proposition expressed by the sentence is construed as being about this referent. *Aboutness* is defined as a pragmatic relation from proposition to referents; a proposition says something about a certain referent if the proposition is relevant to the referent and if it increases the knowledge of this referent. The formal definition given by Lambrecht (1994) is the following.

\[(86)\] TOPIC: A referent is interpreted as the topic of a proposition if in a given situation the proposition is construed as being about this referent, i.e. as expressing information which is relevant to and which increases the addressee’s knowledge of this referent. (Lambrecht, 1994, p. 131)

Consequently, a topic expression is defined as a constituent that is contained within a clause which, in its proposition, is about the referent of the constituent.

The notion of topic is closely tied to the notion of the subject, particularly the notion of the logical subject, which goes back to traditional grammar. This, in turn, is connected to the aboutness defined above: the subject, in most discourse instances, is, after all, the thing which the proposition expressed by the sentence is about, hence the topic. In spite of this close connection, one must not conflate the two terms subject and topic, since in many cases they do not fall on the same word (Lambrecht, 1994, Sornicola, 1996).

Another important feature of the concept of topic is its connection to *givenness*. In the pragmatic literature, topic is often said to refer to some
information that is present in situational context, meaning that the referent of the topic is an already established referent in the discourse. In other words, topic is information that is presupposed by the speaker to be existent in the mind of the hearer (Sornicola, 1996). I clarify the notion of topic using the examples below.

In a sentence such as (87), which is an unmarked English sentence with respect to intonation and emphasis, the grammatical subject is naturally understood as the topic expression of the sentence (Lambrecht, 1994, Sornicola, 1996, Büring, 1999). Suppose the sentence was uttered in a situation where no ongoing discourse is given, e.g. if the sentence was uttered at the beginning of a discourse between two individuals. Then the grammatical subject the snowplow would be the topic expression, and the proposition expressed by the sentence, namely that X cleared the road, would be about the snowplow.

(87) The snowplow cleared the road.

In (88), however, where a sentence-initial accent is given, the grammatical subject would normally be assigned a focus role instead of a topic role, since the snowplow is, as indicated by the question, not an established referent in the discourse, and hence must be recognized as new information. The sentence therefore has no structural element which can be identified as topic. Lambrecht (1994) calls the focus role that the subject receives in this case argument focus.

(88) What cleared the road? — The snowplow cleared the road.

Finally, in a discourse such as (89), the grammatical subject again cannot be assigned the topic role. Lambrecht (1994) calls the answer in (89) an event-reporting utterance; the whole sentence is in the focus of the discourse, hence the subject cannot be the topic. It is not presupposed that it is the snowplow that cleared the road; in fact, nothing is presupposed, apart from the fact that something happened.
(89) What happened? — The snowplow cleared the road.

It is therefore obvious that the assumption subject equals focus would oversimplify things in a dramatic way. What also becomes obvious by these examples alone is that it will be impossible to unambiguously assign discourse functions in Information Structure to single constituents in a single sentence, without taking into account additional contextual or intonational (i.e. phonological) information. The consequences of this restriction for the theory of LFG, and particularly for the parsing of sentences in XLE, are discussed in section 4.2.

Focus

In the context of this thesis, the concept of focus is the most important concept within Information Structure, since the syntactic process of clefting is inherently a focusing device, meaning that a certain portion of a sentence is “highlighted” (Collins, 1991, Declerck, 1988, see also section 3.1). I briefly define the notion of focus in this subsection.

Focus is defined in Lambrecht (1994) as the element of information whereby the presupposition and the assertion of an utterance in a given context differ from each other:

(90) **FOCUS**: The semantic component of a pragmatically structured proposition whereby the assertion differs from the presupposition.

(Lambrecht, 1994, p. 213)

What the expression *pragmatically structured proposition* in the definition means is made obvious by the following example.

(91) Where did the fisherman go last night? — He went to the SEASIDE.

The accentuated element would be considered as the focus constituent by many researchers, but Lambrecht (1994) introduces a distinction concerning
the scope of the focus. The answer in (91) conveys the proposition that *the place the fisherman went to last night was the seaside*. Hence, we can only say that the referent of the expression *the seaside* is new information *in relation* to the predicate of the above proposition. Lambrecht (1994) calls this relation a focus relation. At the core of the definition, however, is the difference between assertion and presupposition; the assertion in (91) (the proposition denoted by the answer) differs from the presupposition (*the fisherman went somewhere last night*) by the accentuated element.

**The relation between topic and focus**

The relation between the two notions of topic and focus is the most important relation within Information Structure; it is also subject to a lot of terminological variation throughout pragmatics. *Given* and *New, Theme and Rheme* (Sornicola, 1996), *Topic and Comment* (Bußmann, 2002, Gundel, 1974), *Topic and Predication* (Bußmann, 2002), *Topic and Focus* (Lambrecht, 1994, Bosch and van der Sandt, 1999) — all of these are found expressing more or less the same polar opposites. One tendency that is consistent throughout literature is the idea of *segmentation* of information into old and new — topic and focus. By this assumption, a focused constituent (i.e. a constituent to which an accent is assigned) contributes new information to the discourse, while a non-focused constituent conveys old information.

Lambrecht (1994) maintains that such a segmentation can only be assumed in the context of pragmatal relations. It is not the constituents or lexical units themselves which convey information; information is conveyed by establishing relations between referents and propositions. While it is important to keep in mind this restriction in the differentiation between topic and focus, and especially the restriction pointed out by Lambrecht (1994) in the way these roles are assigned, the intuitive appeal of the concepts of old and new information is out of the question.
4.2 Information Structure and Irish Clefting

Now that the main concepts of Information Structure are defined, their application to Irish clefting is modeled in this section. First consider the general structure of clefts; the structure is repeated from (72).

(92) COP [XP]_focus [relative clause]_subject

Putting this general layout in the perspective of Information Structure, I maintain that the relative clause constitutes the topic of cleft sentences, i.e. it takes the role of what is being talked about in the sentence. The proposition expressed by the topic material is understood to be presupposed, i.e. already established in the discourse. The clefted constituent receives the focus role. This means that it is not part of the presupposition; it is in fact the contrary concept to presupposition. The focus is understood as the assertion - material which has not yet been established in the discourse.

The whole picture is more complicated, however. Note that in most of the examples above, and in all the examples showing the implementation in 3.4.1, the focus constituent was a simple one, meaning that it contained a single PRED. Relevant examples are repeated here for convenience in (93).

(93) a. Is é an muinteoir a léigh leabhar inné.

b. Is tinn a bhí sé.

c. Aréir a léigh sí sa leabharlann.

In these cases, it is easy to assign focus, as there is only one candidate for focus in each sentence: the NP _an muinteoir_ in (93a), the adjective phrase _tinn_ in (93b) and the adverbial phrase containing just the word _aréir_ in (93c).

There are, however, cases of sub-constituent focus where the predicate constituent is complex (i.e. has multiple PREDs) and contains elements with contrastive stress. This is exemplified in the examples in (94).
(94) a. Is é [an carr DEARG] a cheannaigh sé.
   buy.Past he
   ‘It was the RED car that he bought.’

   ART.Def.Sg cat
   ‘It is UNDER the table where the cat was.’

c. Is é [an MÚINTEOIR sean] a léigh leabhar inné.
   ART.Def.Sg teacher read.Past book yesterday
   ‘It was the old TEACHER that read a book yesterday.’

Since it is not possible to focus just parts of the relevant constituent (e.g. just the adjunct adjective of the NP in (94a)), the only possibility is to front the entire NP. Only partial elements of the focus constituent, however, bear contrastive stress. One can easily imagine to answer (94a) to a question such as “Did he buy the green car?” In such cases, the view across literature is that only the parts of the predicate bearing stress should be treated as the focus of the sentence (Halvorsen, 1977, Declerck, 1988).

The problem now is that the concepts of focus and predication can not be unified anymore. I stress that predication and focus are concepts belonging to different levels of grammar: predication (in the meaning in which it is used here) is at the level of syntax, while focus is at the level of pragmatics, or, more precisely, Information Structure. It is therefore not surprising that there are structural mismatches between the two. The question is how we can describe these mismatches in the context of LFG.
4.3 Projecting Information Structure into LFG

LFG is a modular grammar theory (Bresnan, 2001, Falk, 2001, Dalrymple, 2001). What this means is that, while its main focus is on the syntax of a language (c- and f-structure), a linguist working on LFG can annotate the phrase-structure rules to account for processes on extra-syntactic levels within a language, such as prosody (Butt and King, 1998, Bögel et al., 2008), optimality theory (for disambiguating between different analyses) (Broadwell, 1999) and Information Structure. In defining such annotations, the very same mathematical mechanisms are used which map c-structure nodes onto f-structure attributes and values (Kaplan, 1995).

4.3.1 Previous work on interfacing IS and LFG

Some work has already been done by researchers working on LFG on interfacing Information Structure with LFG. In the classical textbooks, such as Falk (2001) and Bresnan (2001), the domain of f-structures is expanded in that discourse functions such as topic and focus are inserted directly into f-structures. In particular Bresnan and Mchombo (1987) have argued that discourse functions like topic and focus can be integrated in the f-structure. In Chichewa, the language Bresnan and Mchombo (1987) is concerned with, these discourse functions have been syntacticized in that they show agreement like other grammatical functions. That is, the language displays syntactic counterparts on the surface string for the encoded discourse function.

Other researchers have assumed a separate projection for discourse functions (Butt and King, 1997, 1996). King (1997) in particular has argued that an annotation of discourse functions directly in the f-structure is not able to capture all IS processes and that a separate LFG projection called i-structure has to be assumed for the encoding of IS. The paper presents data from Russian and discusses three different phenomena, which are all related to the focusing of f-structure heads: yes-no-questions, contrastive focus and
sub-constituent focus. With these focus structures, a direct treatment in the
f-structure is not possible. An example for contrastive focus on the verb from
Russian is given in (95).

(95) Ona PROČITALA knigu.
    she read book
    ‘She READ the book.’ (King, 1997, p. 5)

In this sentence, the contrastive focus is on the verb; imagine the sentence
as an answer to a question such as “Did she burn the book?” or “What did
she do with the book?” If we were to annotate this in the f-structure, we
have two options. One is by using the annotation in (96a), the other is by
using (96b).

(96) a. ↓ ε (↑ FOC)

b. (↓ PRED) ε (↑ FOC)

Both annotations do not yield the correct result. (96a) will insert anything
on the top level of the f-structure into the focus feature FOC. That is, it
will put the entire f-structure in FOC, including the subject, the object and
possibly any adjunct that might occur in the sentence. (96b) chooses the
PRED value as the feature to be inserted in FOC, hence it takes the f-
structure’s head. This also results in too wide a scope: selecting the verbal
PRED will not only select the core (semantic) meaning of the verb, but also
its subcategorization frame, together with the subject and the object (King,
1997). Figure 4.1 shows the structure created by (96), and Figure 4.2 shows
the intended structure.
Since simple Irish clefting cannot be used to focus verbal elements, this is not a problem that we are concerned with. The example simply shows a general issue with the annotation of discourse functions in the f-structure.

Another problem King (1997) mentions is the issue of sub-constituent focus. In Russian, declaratives with neutral intonation have right-edge focus, with the size of the focus depending on the context of the utterance. An example is given in (97).

(97) Ona pročitala knigu.

\[
\begin{array}{c}
\text{SUBJ} & \text{PRED} & \text{OBJ} \\
\text{FOC}
\end{array}
\]

\[
[\text{read} \langle (\uparrow \text{SUBJ}), (\uparrow \text{OBJ}) \rangle]
\]

‘She read the book.’ (King, 1997, p. 6)

The sentence can have its focus either on just the object or on the verb and the object. Annotating the whole I’ (i.e. the verb together with the object)
with $\downarrow \varepsilon (\uparrow \text{FOC})$ as shown in Figure 4.3 will include the main PRED of the sentence, hence the entire f-structure (including the subject) will be in FOC, as depicted in the f-structure in Figure 4.4.

Another possibility would be to annotate the phrasal nodes with $(\downarrow \text{PRED})\varepsilon (\uparrow \text{FOC})$, coming from the right edge of the clause. The problem remains; as soon as the verbal PRED is annotated, the whole subcategorization frame is put into FOC along with the verb — including the subject, which should not be in there.

King (1997) proposes a two-fold solution:

1. projection of a separate level for the annotation of discourse functions: i-structure;

Figure 4.3: Russian c-structure: sub-constituent focus

Figure 4.4: f-structure focus annotation: scope too wide (II)
2. removal of the argument structure of the predicate.

Two separate types of annotation are used within the c-structure rules: one projects the LFG-usual f-structure, the other projects the i-structure. In Figure 4.5, the general layout of such a grammar is given.

phonetic string
    \|--
c-structure
    \|--
f-structure i-structure
    \|-- semantics

Figure 4.5: i-structure projection in LFG architecture

There are two reasons the i-structure maps into the semantics; first, i-structure can serve as input to semantic interpretation, and second, the semantics module can check if all strings in a sentence containing a PRED value are assigned a discourse function role. In an overall architecture like this, Information Structure is treated as a separate module making its own contribution to the interpretation of a sentence.

The second part of the solution involves cutting off the subcategorization frame of any PREDs. In the Russian examples above, the inclusion of the verbal PRED resulted in too wide a scope for the focus function. The core predicate values can be accessed using the function PRED FN (Kaplan and Maxwell III, 1996). Below is an example of the application of PRED FN, taken from King (1997).

(98) PRED 'read<(↑ SUBJ), (↑ OBJ)>'
PRED FN read (King, 1997, p. 9)

If the mapping between c-structure and i-structure is annotated with PRED FN and not with simple PRED notation, then the problems with the
scope of the focus discourse function disappear, since the subcategorization frame will not be included in PRED FN.

I have presented approaches to the integration of discourse functions in LFG; I have given an account of the solution presented by King (1997). I will not discuss the eventual solution given for Russian by King (1997), but rather elaborate on the ideas and use them to derive an analysis of Irish clefting which is more complete in that it contains an interface to Information Structure.

4.4 Towards a more complete analysis of Irish clefting

What are the consequences of IS for Irish clefting? As discussed in 4.2, we also get sub-constituent focus in Irish clefts. I show in this section that the ideas of King (1997) can be applied in a straightforward way to Irish clefts to account for the IS-relevant consequences of these constructions.

First, I again present examples with simple PREDs as focus. The examples from (93) are repeated here for convenience.

(99) a. Is é an múinteoir a léigh leabhar inné.
   COP.Pres AGR.Masc.Sg ART.Def.Sg teacher COMP.Rel
   read.Past book yesterday
   ‘It is the teacher who read a book yesterday.’

   b. Is tinn a bhí sé.
   COP.Pres sick COMP.Rel be.Past he
   ‘He was sick!’ [lit. ? ‘It is sick that he was!’]
c. Aréir a léigh sí sa leabharlann.

last_night COMP.Rel read.Past she in.Def library

? ‘It was last night that she read in the library.’

In (99a), the entire focus NP can be assigned the focus function in i-structure, since it only contains a single PRED. The PREDs coming from the relative clause are put in a set and assigned the topic function in i-structure, because they represent the established ground of the discourse. The f-structure for (99a) is given in Figure 4.6, the i-structure in Figure 4.7. I gloss the i-structure in English.

(100) Is é an múinteoir a léigh leabhar inné.

Figure 4.6: cleft sentence with NP in focus position: f-structure

\[
\text{FOC} \{\text{teacher}\} \\
\text{TOP} \{\text{read, book, yesterday}\}
\]

Figure 4.7: cleft sentence with focus NP: i-structure

The topic of Figure 4.7 does not include the subject of the main verb, since the subject is not part of the established ground of the discourse; it is new information, hence in FOC. The presupposition of the sentence is that
someone read a book (leaving aside the subject-object ambiguity discussed above), and the assertion is that this someone is the teacher.

Next, see the f-structure for (99b) in Figure 4.8 and the resulting i-structure in Figure 4.9. Here, we have a predicative adjective which is clefted, hence receives the focus function.

(101) Is tinn a bhí sé.

"Is tinn a bhí sé."

The verb bí and the subject are annotated for the discourse function topic in this example: the referents of these words are presupposed, given in the discourse.

The last of the simple examples, (99c), includes an adverbial phrase, aréir, in the focus position. In the i-structure, anything belonging to the relative clause is analyzed as part of the presupposition, i.e. the topic, while the adverbial in the predicate position is analyzed as the focus of the sentence.

(102) Aréir a léigh sí sa leabharlann.
"Aréir a léigh sí sa leabharlann."

Figure 4.10: cleft sentence with adverbial in focus position: f-structure

```
FOC {{last,night}}
TOP {{read,she,in,library}}
```

Figure 4.11: cleft sentence with adverbial focus: i-structure

But what about the sentences with complex phrases in the focus position? In these, the focused phrases are complex in that they contain multiple PREDs, and different smaller parts of the clefted constituent may be contrastively stressed.

(103) a. Is é an carr DEARG a cheannaigh sé.

‘It was the RED car that he bought.’

b. Is an bhí an cat.

‘It is UNDER the table where the cat was.’
c. Is é [an MÚINTEOIR sean] COP.Pres AGR.Masc.Sg ART.Def.Sg teacher old a léigh leabhar inné.
COMP.Rel read.Past book yesterday
'It was the old TEACHER that read a book yesterday.'

In these cases, only partial elements of the copula predicate have to be associated with the focus function. Since we have highlighting of the element bearing contrastive stress in these cases, we can unambiguously assign the focus function to DEARG, FAOIN and MÚINTEOIR in (103a), (103b) and (103c), respectively. This is shown below for sentence (103b); the f-structure is given in Figure 4.12 and the i-structure in Figure 4.13.

(104) Is [FAOIN mbord] a bhí an cat.

"Is faoin mbord a bhí an cat."

Figure 4.12: cleft sentence with partial focus: f-structure

```
FOC {under}
BCK {table}
TOP {be, cat}
```

Figure 4.13: cleft sentence with partial focus: i-structure

88
What discourse function does the remainder of the phrase in predicate position, *mbord* ‘table’, receive? In Figure 4.13, it is analyzed as BCK, background. Backgrounded material is similar to topic information in that it is already established in the discourse, ‘known’ material. It is, however, different from the topic in that provides information as to how exactly the new information fits in with the old information But \(t\) and King (1997), King (1997), i.e., how the focus fits in with the topic. This is exactly what we find in this sentence: the focus element *under* fits in with the topic *be, cat* in that it was under the *table* where the cat was; *table* receives background function. Lambrecht (1994) does not distinguish background from the topic function role, that is, in his approach, backgrounded information would be included in the topic function. However, I agree with King (1997) and Butt and King (1997) in assuming a separate discourse function for backgrounding material, since it is in fact cross-linguistically operating as a linking element between new and old information and therefore different from topic; see also Hoffman (1995) for motivation of the background role.

Note also that in the f-structure, the prepositional phrase in the predicate position takes an object. If we want to map only the preposition *faoin* ‘under’ onto the focus function, we have to use the PRED FN function in the c-structure annotation; otherwise, the whole PRED of the preposition, including its object, will be projected to FOC, which is not what we want. A possible prepositional phrase rule accounting for this type of partial focus is depicted in (105).

---

1The rule is simplified. \(\dagger_i\) is used to indicate projection onto i-structure. The actual annotation scheme is more complex; see Crouch et al. (2008).
(105) PPfoc --> \{ P: (↑ PREDLINK)=↓
     (↓ PRED FN) ∈ (↑i FOC) ...\n     NP: (↑ OBJ)=↓
     | P: (↑ PREDLINK)=↓
     NP: (↑ OBJ)=↓
     (↓ PRED) ∈ (↑i FOC) ...\n
The rule allows for either projecting the PRED FN of the preposition or the PRED of the preposition’s object to the FOC function in i-structure. There is no reason for using the PRED FN function if the object is projected to FOC, since the nominal phrase does not contain a subcategorization frame in its PRED.

4.4.1 Consequences for a computational grammar

A computational grammar designed for the processing of text, like an LFG grammar, mostly will encounter text which is not formatted in a way such as the grammar can make out the focused word or phrase. If the focus contains just a single PRED, then the grammar can unambiguously map the predicate of the copula to the FOC function within i-structure. The problem is that in normal text, the focus is rarely highlighted.\(^2\) Consider a sentence such as (106).

(106) Is é an carr dearg a
     COP.Pres AGR.Masc.Sg ART.Def.Sg car red COMP.Rel
     cheannaigh sé.
     buy.Past he
     ‘It was the red car that he bought.’

\(^2\)Also note that, even if the focus element in a complex cleft predicate is highlighted, the linguist would have to teach the grammar how to interpret the highlighting, which might not be totally straightforward.
In this sentence, either the whole NP, just the nominal head, or just the adjective adjunct may be the focused element. The problem is that the grammar will not be able to make out the focused element, because there is no highlighting as in the previous examples. The c- and f-structure for the sentence will be unambiguous, but the i-structure will look different, depending on the allocation of the discourse functions.

Therefore, I propose a one-to-many mapping between c-structure and i-structure. For the disambiguation between the different possibilities in choosing the focus role, I assume that LFG grammar-external processing has to be used. Discourse analysis techniques (e.g. Centering Theory (Grosz et al., 1986, 1995)), can take on further processing tasks and disambiguate between structures generated by the syntax grammar.

\[
\begin{bmatrix}
\text{FOC}\{\text{red}, \text{car}\} \\
\text{TOP}\{\text{bought, he}\}
\end{bmatrix}
\]

Figure 4.14: cleft sentence with complex focus NP: i-structure (I)

\[
\begin{bmatrix}
\text{FOC}\{\text{red}\} \\
\text{BCK}\{\text{car}\} \\
\text{TOP}\{\text{bought, he}\}
\end{bmatrix}
\]

Figure 4.15: cleft sentence with complex focus NP: i-structure (II)
Figure 4.16: cleft sentence with complex focus NP: i-structure (III)

The LFG grammar is able to produce all these structures from the c-structure, but will not be able to choose the correct one depending on the context. Further processing is needed to unambiguously determine the focus, topic and background discourse functions. This, however, is not within the scope of this thesis.
Chapter 5

Conclusion

This thesis has presented an analysis of Irish clefting. I first discussed some preliminaries in chapter 1, then presented LFG approaches to copula predication in chapter 2. Chapter 2 also included my own analysis of Irish copula. In chapter 3, I first defined the concept of clefting, then presented the Irish data, then explained the similarities between simple copula constructions, question formation, and clefting in Irish. In the analysis of Irish clefting, the similarities are reflected in that the f-structures are designed using the same basic structures. Chapter 4 finally discussed not only the basic assumptions about Information Structure, but also showed how IS can be integrated into the LFG architecture. The chapter also described the Information Structure of Irish cleft sentences and showed how the LFG analysis of clefting can be enhanced by using i-structure, a separate projection within the modular LFG framework. Chapter 4 concludes with remarks about the consequences of assuming an i-structure for a computational grammar used for parsing standard, non-highlighted text; a one-to-many mapping between c- and i-structure is suggested.
5.1 Future work

Future work on the aspects discussed in this thesis point into various directions. First, because of the obvious relation between clefts and other, more simple copula constructions in Irish, an extension of the ideas of Information Structure to the copula constructions discussed in chapter 2 could provide valuable insights, especially since the relationship between predicate and subject is not always clear in some copula constructions.

Another possible area of further research are the other variants of Irish clefting. Ó Siadhail (1989) presents four different cleft constructions: simple clefts, as discussed in this thesis, pseudo-clefts, verb focusing, and inverted clefting. Both a syntactic analysis from the viewpoint of LFG and an analysis within IS of these constructions could provide additional insights into the structure of clefting.

Third, the projection to i-structure has not yet been implemented in the XLE program. The i-structures depicted in chapter 4 are not taken from XLE output. The implementation seems straightforward, however; the XLE documentation gives detailed instructions on how to annotate the c-structure rules to project additional levels of representation.
Bibliography


Butt, Miriam, Stefanie Dipper, Anette Frank, and Tracy Holloway King. 1999a. Writing Large-Scale Parallel Grammars for English, French and German. In M. Butt and T. H. King, eds., *Proceedings of the LFG99 Conference*.


