Doktori (PhD) értekezés

ON CLAUSE-INITIAL DISCOURSE-RELATED CONSTRUCTIONS IN ENGLISH AND HUNGARIAN

Szűcs Péter

Témavezető: Dr. Laczkó Tibor

Debreceni Egyetem

BDT

2017

ON CLAUSE-INITIAL DISCOURSE-RELATED CONSTRUCTIONS IN ENGLISH AND HUNGARIAN

Értekezés a doktori (Ph.D.) fokozat megszerzése érdekében

a ..... tudományágban

Írta: ..... okleveles .....

Készült a Debreceni Egyetem ..... doktori iskolája

(..... programja) keretében.

Témavezető: Dr. . . . . .

A doktori szigorlati bizottság:

elnök: Dr. ……

tagok: Dr. ……

Dr. ……

A doktori szigorlat időpontja: 200… . …… … .

Az értekezés bírálói:

Dr. .....

Dr. ……

Dr. .....

A bírálóbizottság:

elnök: Dr. .....

tagok: Dr. ……..

Dr. ……..

Dr. ……..

Dr. ……..

A nyilvános vita időpontja: 200… . …… … .

Én Szűcs Péter teljes felelősségem tudatában kijelentem, hogy a benyújtott értekezés önálló munka, a szerzői jog nemzetközi normáinak tiszteletben tartásával készült, a benne található irodalmi hivatkozások egyértelműek és teljesek. Nem állok doktori fokozat visszavonására irányuló eljárás alatt, illetve 5 éven belül nem vontak vissza tőlem odaítélt doktori fokozatot. Jelen értekezést korábban más intézményben nem nyújtottam be és azt nem utasították el.

TABLE OF CONTENTS

ABSTRACT

The dissertation investigates discourse-related clause-initial constructions within the framework of Lexical-Functional Grammar (LFG). The following structures are to be scrutinized: English Topicalization (TOP-Eng), Clause-initial adjuncts (CIADJ-Engs) and Left-dislocation (LD-Eng); Hungarian Left-dislocation (LD-Hun) and Operator fronting (OF). The overall goal is to provide an adequate account for these structures, especially with regards to their syntactic and information-structural properties.

The dissertation builds on several earlier approaches, expands and improves them in various ways. As for information-structure, it is argued that no satisfactory account for the understanding and representation has been proposed and thus a new framework is put forward. The proposal is an amalgamation and enhancement of earlier feature-based approaches. I argue that the features NEW, DISCOURSE-STRUCTURING and CONTRASTIVE are well-definable and able to account for the existing information-structural categories. In syntax, I rely on Laczkó's works (2014a, 2014b, 2015) on Hungarian and supplement them in two ways. One, I incorporate my proposals about information-structure into them, and two, I expand them to cover subordinate clauses as well. For this latter part, I investigate the theoretical and empirical landscape and argue extensively that contra the standard view (Kenesei 1992/1994) the pronouns associated with Hungarian subordinate clauses (*az* 'that') are contentful and not expletives.

Once the theoretical background is laid, I discuss the target-constructions in detail. After examining their syntactic and information-structural properties, argue that these structures necessitate analyses where the left-peripheral constituents are integrated into the sentence-structure in various ways and to different degrees. I support my analyses with various evidence and provide a LFG-theoretic formal account for each of the constructions.

In English, TOP-Eng is argued to be a "fronting proper" construction, where the topicalized element is functionally identified with a clause-internal function. At information-structure, TOP is always contrastive. Some CIADJs also follow this pattern, but there also exist CIADJs that are "base-generated" neutral topics, without a link to the inner part of the sentence. Left-dislocated entities in English are even less integrated: they are argued to be "syntactic orphans" which are not related to the host-sentence at the level of syntax.

In Hungarian, three types of Left-dislocation are distinguished: Contrastive topic left-dislocation (CTLD), Noncontrastive left-dislocation (NCLD) and Contrastive focus left-dislocation (CFLD). While the first two are syntactically integrated structures (involving an anaphoric dependency between the discourse-prominent element and the associate pronoun),

CFLD is not and is argued to involve a sentence-external left-peripheral entity, akin to LD-Eng. Operator fronting is also divided into two types. The first one is a proper fronting construction, sharing many properties with TOP-Eng, while the second type is a representative of "prolepsis", where the fronted element becomes a thematic argument of the main verb and is anaphorically linked to a clause-internal function.

The last part of the dissertation is concerned with the theoretical and cross-linguistic ramifications of the findings and proposals put forward in the dissertation. I show that "fronting", "left-dislocation" and "proleptic" structures exist in various languages and substantial parallelisms may be observed with regards their properties. I compare and contrast these properties and outline a taxonomy. As prolepsis is a relatively little-studied phenomenon, I put special emphasis on its place in syntactic typology. I argue that it may be regarded as a type of finite control and substantiate this claim with providing an overview of such constructions. I also show the place of prolepsis in the typology of control.

Overall, the dissertation reaches two goals: it offers comprehensive analyses for various clause-initial discourse-related linguistic structures and it also provides wider perspective by investigating the theoretical and cross-linguistic place for the ideas put forward.

ÖSSZEFOGLALÁS

Az értekezés az angol és magyar nyelv olyan mondatszerkezeteivel foglalkozik a Lexikai-Funkcionális Grammatika (LFG) elméleti keretében, melyekben valamely diskurzusperiférikus összetevő balperiférikus pozíciót foglal el. A vizsgált mondattípusok a következők: topikalizáció (TOP-Eng), tagmondateleji adjuntumok (*Clause-initial adjuncts*, CIADJ-Eng) és balra kihelyezés (*Left-dislocation*, LD-Eng) az angolban, valamint balra kihelyezés (LD-Hun) és operátor-előrevivés (*Operator fronting*, OF) a magyarban. A cél az, hogy ezen szerkezetek megfelelő elemzést kapjanak, különös tekintettel a szintaxisra és az információs szerkezetre.

Az értekezés több korábbi elmélet is felhasznál, módosítva és bővítve azokat. Az információs szerkezetet illetően amellett érvelek, hogy a korábbi megközelítések nem kielégítőek, és ezért egy új taxonómiára van szükség. Az általam felállított taxonómia a korábbi jegyalapú rendszerek továbbgondolása. Állításom szerint az ÚJ (*NEW*), DISKURZUSSZERKEZET-ÉPÍTŐ (*DISCOURSE-STRUCTURING*) és a KONTRASZTÍV (*CONTRASTIVE*) jegyek jól definiálhatók és lefedik a létező információs-szerkezeti kategóriákat. A szintaxist illetően nagyban építek Laczkó (2014a, 2014b, 2015) magyar nyelvet taglaló műveire, de két szempontból is kiegészítem azokat: az információs-szerkezetet illető elgondolásaim beépítésre kerülnek, valamint az alárendelő mondatok elemzésére is sor kerül. Ez utóbbival kapcsolatban részletesen érvelek amellett, hogy a magyar alárendelő mondatok névmási asszociáltja az általánosnak tekinthető nézettel szemben (Kenesei 1992/1994) nem expletívum, hanem tartalmas névmás.

Az elméleti háttér felállítása után részletesen tárgyalom a célszerkezeteket. A mondattípusok tulajdonságainak a vizsgálata után arra a következtetésre jutok, hogy a balperiférikus összetevők különböző mértékben és módokon vannak a mondatba integrálva. Ezt részletesen alátámasztom bizonyítékokkal és a szerkezetekre formális elemzési javaslatot teszek LFG-keretben.

Az angol topikalizáció meglátásom szerint egy klasszikus „előremozgatás”-szerű szerkezet, ahol a mondateleji összetevő funkcionális azonosításra kerül egy mondatbeli grammatikai funkcióval. A topikalizált összetevők mindig kontrasztívak az információs-szerkezetben. Bizonyos tagmondat-eleji adjunktumok is így viselkednek, azonban vannak olyanok is, amelyek semleges topikként teljesen a mondat elejéhez kötődnek, a mondat belsejéhez való kapcsolódás nélkül. Az angol balra kihelyezésben részt vevő összetevők még ennél is kevésbé integráltak: meglátásom szerint ezek olyan „szintaktikai árvák”, vagy „függő topikok”, melyek nem tartoznak az anyamondathoz szintaktikai szempontból.

A magyarban a balra kihelyezés három típusát különböztetem meg: kontrasztív topikos balra kihelyezés (*Contrastive topic left-dislocation*, CTLD), nemkontrasztív balra kihelyezés (*Noncontrastive left-dislocation*, NCLD) és kontrasztív fókuszos balra kihelyezés (*Contrastive focus left-dislocation*, CFLD). Az első kettőben a balra kihelyezett elem a mondat szerves részét képezi (és anaforikus függőségben áll az asszociált névmással), a CFLD az angol balra kihelyezéshez hasonlóan egy az anyamondathoz szintaktikailag nem kapcsolódó szerkezet. Az operátor-előrevivésben szintén megkülönböztetek két altípust: az egyik az angol topikalizációhoz nagyban hasonló „előrevivés"-szerű szerkezet, míg a másikat „prolepszis"-ként elemzem: a mondateleji összetevő a főpredikátum thematikus argumentumává válik és anaforikusan azonosítódik egy alárendelt mondat-beli grammatikai funkcióval.

Az utolsó fejezet az értekezésben megfogalmazott állítások elméleti következményeivel és nyelvközi összehasonlításokkal foglalkozik. Megállapítom, hogy hasonló „előrevivő" és „balra kihelyező" szerkezetek sok egyéb nyelvben is megtalálhatóak, és a tulajdonságaik összehasonlítása után fölvázolok egy taxonómiát. Mivel a prolepszis egy viszonylag keveset kutatott szerkezettípus, nagy terjedelemben foglalkozok az ő helyével a szintaktikai elméletekben. Amellett érvelek, hogy egyfajta finit kontroll-szerkezetről van szó. Ezen állításom mentén áttekintést nyújtok az ilyen szerkezekről és bemutatom, hogyan helyeződik a prolepszis egy lehetséges kontroll-tipológiába.

Összességében az értekezés két célt is szolgál: átfogó elemzést nyújt a szóban forgó diskurzushoz kapcsolódó balperiférikus szerkezeket illetően, valamint tágabb perspektívát is nyújt azáltal, hogy megvizsgálja az elemzési javaslatok elméleti hátterét és következményeit.

ACKNOWLEDGEMENTS

During the past years, there were times when the completion of this dissertation seemed so distant that I doubted that I would ever get there. That these doubts have remained unfounded is to a great extent a result of me being supported by a number of wonderful people, to whom I cannot be grateful enough.

First and foremost, I would like to express my gratitude towards my supervisor, Tibor Laczkó. He introduced me to the world of linguistics and LFG in particular, so but for him, the idea of me entering PhD-studies would have never occurred to me. During my research, he was always there for me to discuss my half-baked ideas, gently push me in the right directions, away from the wrong ones and he also helped me out in various formal aspects of the dissertation. Without his support and guidance, the previously mentioned doubts would surely have been proven.

Rákosi György is another vital member of my close linguist community. His sharp observations and insightful suggestions have formed and improved my ideas tremendously.

I must also individually mention Réka Jurth. I forever indebted to her ,because without her the carefull eyes of Réka, THE entire disssertation w$_{ou}$ld look like *th*is sentence “looks′ .

I would also like to thank the entire Institute of English and American Studies at the University of Debrecen for providing a superb background for me. This community of people, instructors, students and administrative personnel included, has been truly stimulating in both professional and personal sense.

I also owe a lot to the reviewers, Gábor Alberti and Balázs Surányi. Their comments, criticisms and corrections about the first version of this dissertation have vastly improved the content.

The annual LingDok conferences in Szeged were the highlights of the academic year for me. The outstanding community of linguists I met there has really been formative for me. I hope that their knowledge, skills and attitudes have made a permanent mark on my scientific progression. I am forever indebted to the organizers for putting these conferences together and to all the participants for their insightful remarks and also for the sense of linguist-community that I have experienced there.

A similar thanks should go to the organizers and participants of all the other conferences and other linguistic events that I attended throughout the years: the LFG conferences, CECIL’S 2013, ConSOLE XXIII, DELITE PhD Symposium, Budapest Linguistics Conference,

SinFonIJA 6, EGG 2011 and 2017. I am thankful to life for making it possible for me to participate in these. These have greatly contributed to my professional and personal growth.

Last but by no means least, I am immensely grateful to my family who supported me with a solid background throughout the years. A very special thanks goes to my father, Sándor Szűcs, whose scientific mind, academic rigor and superb work-ethic always inspired me to bring the best out of myself.

LIST OF ABBREVIATIONS

Throughout the dissertation, if not indicated otherwise, nominative case, present tense and indicative mood are assumed.

ABL	ablative
ACC	accusative
AUX	auxiliary
BG	background information
CF	contrastive focus
CI	completive information
CL	clitic
COMP	complementizer
COND	conditional
CT	contrastive topic
DAT	dative
DEF	definite
DEL	delative
DIST	distal
FOC	focus
FUT	future
IF	information focus
INDEF	indefinite
INF	infinitive
MASC	masculine
NOM	nominative
PASS	passive
PL	plural
POSS	possessive
PROX	proximal
PRT	particle
PV	preverb
Q-PRT	question particle
SBJV	subjunctive
SG	singular
SM	subject marker
that(C)	complementizer *that*
TOP	Topic

CHAPTER 1

INTRODUCTION

The left periphery of the clause is commonly viewed as one of the most prominent area for the interaction of syntax and information structure. This is true not only in the so-called discourse-configurational languages (languages where the primary structuring force behind the order of constituents is their discourse/information-structural status, see É. Kiss 1995), but also in conventionally configurational ones like English, where the phrase structure rules impose a more rigid syntactic hierarchy and constituent-order. The aim of this dissertation is to investigate some of these constructions, where the order of constituents deviate from what is considered as "normal" word order. More precisely, in the structures to be scrutinized, some elements occur in a left-peripheral position, not necessarily in the one that they thematically belong to. The non-canonical word order in each case has information/discourse-structural effects. That is, the dislocated constituents receive a specific discourse-function, which would not necessarily be the case if they were not involved in the configurations to be discussed.

The dissertation investigates such discourse-related left-peripheral constructions in two typologically distant languages, English and Hungarian. The structures under investigation are the following: (1) illustrates Topicalization in English (TOP-Eng), (2) shows a Clause-initial adjunct in English (CIADJ-Eng) and (3) demonstrates Left-dislocation in English (LD-Eng). (4) displays versions of Hungarian Left-dislocation (LD-Hun) and (5) showcases what I label Hungarian Operator fronting (OF). In every case, the label of the constructions should simply be viewed as convenient names, not as theoretical/analytical suggestions:

(1) *John, I like.*

(2) *In New York, there is always something to do.*

(3) *John, I like him.*

(4) a *Jánost, azt Kati szereti.*
John.ACC that.ACC Kate likes
'As for John, Kate likes him.'

b *Erre János az fogta magát, és elszaladt.*
then John that took.3SG himself and away.ran.3SG
'Then John, he went and ran away.'

c *A könyvet, AZT olvastam a szobában (és nem az*
the book.ACC that.ACC read.1SG the room.in and not the
újságot).
newspaper.ACC
'The book, I read THAT in the room (and not the newspaper).'

(5) *János(t) mondtam, hogy jön a partira.*
John(ACC) said.1SG that(C) comes the party.to
'(Of) John I said that he will come to the party.'

Throughout the dissertation, expressions like "fronting", "dislocation", "extraction" will be heavily used. As with the names of the constructions above, in each case, such usage should be viewed as descriptive labelling for the configurations/processes, not as theoretical/analytical commitments on my part. These terms are deeply embedded in the linguistic canon and my concern here is neither terminological innovation nor the use of cumbersome paraphrases, so I stick with the tradition in this respect.

The ultimate question of the dissertation may be put simply:

What is the relationship between the fronted expression and the rest of the sentence?

Elaborating on this, the answers that I seek are related to two levels of linguistic analysis: syntax and information-structure.

The first level is concerned with the thematic/argument structural, formal and functional links between the left-peripheral expression and the host-sentence. As we will see, there are different ways to incorporate the fronted constituent into the sentence and these different methods are detectable in how the fronted element interacts with the composition of the sentence. Part of this are "connectivity-effects": formally observable markings that appear as if the fronted element were occupying its canonical position. These vary in the constructions shown in (1). Such variability will be shown to be demonstrable in various syntactic phenomena including formal features (agreement), binding patterns and interaction with syntactic islands. We will see that the nature of the link between the fronted elements and their intra-sentential correlates may range from a strict syntactic dependency through a looser semantic/anaphoric one to a mere pragmatic inference.

As for information-structure, I will seek to establish what kind of communicative purpose is served by fronting the element from its "normal" position. I will aim to do so in a systematic and principled framework. The construction of such a framework is essential as despite a lot of work in the area, no consensus has been reached about the proper view of

information-structure. While it is clear that some sort of feature-based proposal is favorable, there are substantially different understandings of what the relevant features are, what they really mean, how they interact and what are the categories that these features should ultimately build. Once an adequate picture of information-structure has been constructed, it will become available for us to understand how these constructions interact with the discourse.

It is clear that the proper analysis of these constructions involves various levels of linguistic theory. As such, the framework of Lexical-Functional Grammar (LFG), with its parallel levels of representation is a suitable choice for the investigation. A related goal of the thesis is to contribute to the LFG-theoretic analysis of Hungarian, as not much work has been done in these respects on the area of long-distance dependencies, including subordinate clauses.

The structure of the thesis is the following. Chapter 2 introduces LFG in detail. The various levels of representation will be showcased in subchapters. Every relevant concept of the theoretical framework will be presented to the reader. Of the representational levels, information-structure will get the greatest emphasis. After providing a brief overview of the field itself and its previous implementations in LFG, I will put forward a new information-structural taxonomy. The proposed system is an amalgamation of previous approaches from several theoretical backgrounds. It follows the tradition of analyzing information-structural categories in terms of features and it will be shown that the features +/–NEW, +/–D-STRUCTURING and +/– CONTRASTIVE are necessary and sufficient for the proper classification of the information-structural categories.

Chapter 3 gives an overview of Hungarian sentence structure. This is necessary for the subsequent analyses of the Hungarian structures. I follow the LFG-theoretic approach to Hungarian of Laczkó (2014a, 2014b, 2015), which in turn is influenced by É. Kiss (1992). The preverbal field is assumed to be hierarchical, while the postverbal field is structurally flat. As Operator fronting involves subordinate-clauses, this chapter also provides a summary of the main issues concerning them. I will confront the influential idea presented in Kenesei (1992/1994) that the pronoun associates of Hungarian subordinate clauses are expletives and I will subscribe to the alternative view in which they are contentful pronouns (Tóth 2000, Rákosi & Laczkó 2005). I will support my choice with theoretical, empirical and typological data.

Chapter 4 targets the English constructions. For each of them, I will first survey their empirical properties in a theory-neutral manner and then I will propose an LFG-theoretic analysis. TOP-Eng will be shown to be a prototypical fronting construction, with strong connectivity effects and as such, it will be analyzed as a strict functional dependency between the fronted element and its canonical position. Some CIADJ-Engs also follow this pattern but there are CIADJs that lack such clause-internal ties and are best viewed as “base-generated” at

the left-periphery. Finally, LD-Engs will be argued to be so disconnected that they call for an extra-syntactic approach. Information-structurally, TOP-Eng will be shown to be a contrast-marker, while the other two are related to the category of neutral Topics.

Chapter 5 is about the Hungarian constructions: Left-dislocation in Hungarian and Operator fronting. Following the pattern of the previous chapter, first a survey of empirical properties will be provided and then the presentation of my analyses will be presented. Three subtypes of LD-Hun will be distinguished: Contrastive topic left-dislocation, Noncontrastive left-dislocation and Contrastive focus left dislocation. It will be argued that Contrastive focus left-dislocation should be regarded as a phenomenon lying outside of syntax proper and in this respect should be treated on par with Left-dislocation in English. The other two left-dislocations in Hungarian are parts of core syntax. They come with an array of unique annotations. Operator fronting will also be shown to be able to take place in two ways. The fronted element may be part of a traditional "fronting" construction, in which case the clause-initial element maintains strong ties with its canonical position (so connectivity effects are strongly present). This results in the retention of its original case and strict agreement with the embedded predicate, to which the fronted element thematically belongs. The other possibility is what is labelled as "prolepsis": the fronted element becomes an object or an oblique argument of the main verb via an argument-structural process and is only semantically (anaphorically) linked with an embedded grammatical function. A formal analysis for both the English and the Hungarian constructions will be provided with reference to various levels of linguistic representation.

Chapter 6 puts the constructions discussed throughout the theses into cross-linguistic and theoretical perspective. It will be demonstrated that many languages possess similar constructions. We will see how the various fronting configurations relate to each other, what the main similarities and differences are. Particular attention will be paid to the exploration of prolepsis and how it can be put into the theoretical space provided by LFG. In particular, I will argue that there is a natural link between prolepsis and "control"-constructions.

The final part concludes the dissertation. I briefly summarize the questions and topics raised in the dissertation and the answers and claims that I have provided for them. Some avenues for future research will also be suggested.

The data discussed throughout the dissertation comes from various sources. I rely heavily on previous literature, but I also use examples that I have constructed. During my research, I have conducted several empirical surveys to support the claims that I make. The details of these surveys are provided at the relevant parts. Corpus examples are also invoked at some points.

Overall, the dissertation provides an investigation of a number of constructions in a considerable depth and while doing so, it fully makes use of and in some respects expands the theoretical space provided by Lexical-Functional Grammar.

CHAPTER 2

THEORETICAL FRAMEWORK

The framework of this dissertation is Lexical-Functional Grammar (LFG), a model of the generative grammar tradition. LFG, which was one of the first major alternatives of the Chomskyan line of research, grew out of two emerging research paradigms in the late 1970s and early 1980s. The first is the theoretical work of Joan Bresnan, who was skeptical of several aspects of the Chomskyan transformational theory of the time (the (Extended) Standard Theory). Her criticisms concerned the (lack of) universality and psychological realism of Chomsky's models. The other emerging research line centered on the work of computational linguist Ronald Kaplan, who aimed to develop formally coherent and computationally implementable models of natural language.

The emerging theoretical (Bresnan's research) and implementational (Kaplan's work) avenues of research converged in 1982 when the first major publication in LFG, *The Mental Representation of Grammatical Relations* was published (Bresnan ed. 1982). The book lay the foundations of a new grammatical theory, as well as showed how it can be applied to various real language phenomena (chapters by KP Mohanan, Avery Andrews, etc.) and also explored the theory's implications for psycholinguistic research (e.g. contributions by Steven Pinker and Marilyn Ford).

Since its conception, the theoretical assumptions and formal tools of LFG have been extraordinarily stable. This allowed the framework to thrive and live up to its goals to a considerable extent. Linguistic diversity, formal rigor and implementation have been nurtured in LFG from the earliest days. Prominent illustrations for this are the annual LFG Conferences and the ParGram/ParSem projects which have produced parallel computational grammars for a number of languages on a common platform (Xerox Linguistic Environment), including Hungarian. In the next sections I give a brief overview of the framework, so the reader can be sufficiently familiar with the theoretical background and motivation of this dissertation. Apart from the section on information-structure, this is a rather succinct outline of LFG, for a comprehensive look, see Falk (2001), Dalrymple (2001) and Bresnan et al. (2016).

2.1 General properties of Lexical-Functional Grammar

The first letter in the acronym "LFG" stands for "lexical". In essence, this means that LFG holds that instead of transformations, it is the (mental) lexicon that is the engine behind most grammatical phenomena. Thus LFG is a non-transformational/non-derivational generative framework.

LFG subscribes to the Lexical Integrity Hypothesis (also called the Strong Lexicalist Hypothesis). It was originally proposed in Bresnan (1978), here we use Falk's (2001:26) formulation.

(1) **Lexical Integrity Principle:** morphologically complete words are leaves of the constituent-structure tree and each leaf corresponds to one and only one constituent-structure node.

This means that unlike in Chomskyan theories, words are "atoms" for syntax. For instance, a standard tree-structure minimal representation for a sentence in a Chomskyan framework would be receive a representation along the lines of Figure 1a, where the "past tense" feature would occupy a node of its own, and then subsequently interact with the verb via some agreement-mechanism. (This is just a simplified illustration (disregarding irrelevant details such as vP or a CP-shell. Also, the bit outdated IP-notation is used in the Chomskyan tree to make it more parallel to the LFG-tree, where IP is a commonly used label). On the other hand, in the LFG-style representation (Figure 1b), each terminal node is a word.

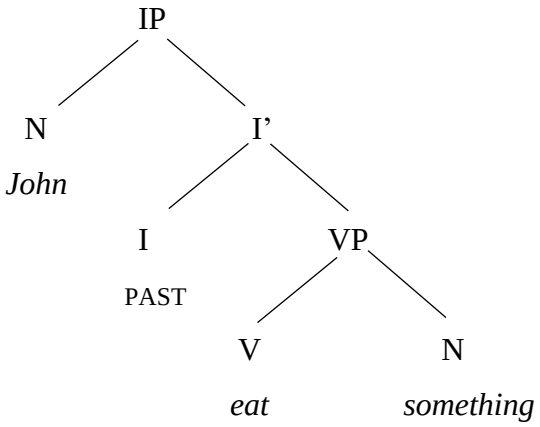

Figure 1a.
A Chomskyan-style tree-structure.

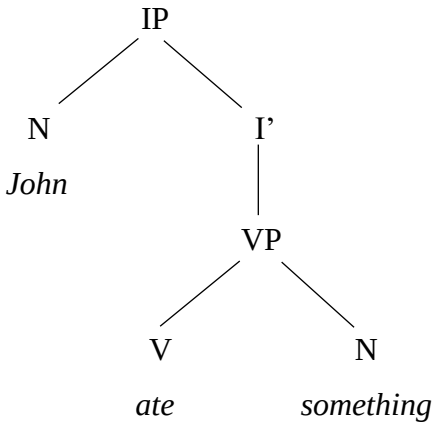

Figure 1b.
An LFG-style tree-structure.

As a result of the Lexical Integrity Principle, "empty" or "zero" categories (traces, pros, PROs, silent copies) are generally avoided in constituent-structures in LFG (although some LFG researchers have suggested that under some very restricted circumstances, they may be made use of, see Falk 2007).

The second letter in the acronym stands for "functional". What this means is that instead of assuming that traditional grammatical functions like "subject" or "object" are only configurationally derived concepts, defined as particular tree-structural positions (e.g. an element is the "subject" if it occupies the Spec/IP position), LFG assumes that these categories are primitives of grammatical theory.[1] Among others, this is motivated by the fact that while constituent-structures among languages may vary radically, grammatical functions are generally assumed to be universal. The advantages of such an approach are evident when one considers "non-configurational" languages, which may exhibit radically free word order and a lack of hierarchical constituency. Warlpiri, an Australian language is a prime example of such languages. In a sentence like (2), apart from the constraint that the auxiliary should be the second word, all possible word-orders are grammatical (the data is from Simpson 1983).

(2) *Kurdu-ngku* *kaju* *nya-nyi* *ngaju*
child-ERG AUX.1SG.OBJ see-NONPAST me.ABS
"The child sees me."

Such a sentence is plausibly analyzed as having a flat constituent structure, posing a serious challenge for configurational approaches. However, it is clear that the sentence contains a predicate (*see*), a subject (*the child*) and an object (*me*), the latter two being distinguished by case-markers. So as far as constituency is concerned, English and Warlpiri are very different, but in functional terms, the sentences receive an essentially identical analysis in both languages. Functional information of this sort thus is a major tool for LFG to capture cross-linguistic regularities.

The next general property of LFG to be mentioned here is its concept of language being analyzed via several, interconnected, parallel structures. Every representational level has its own "vocabulary" and rules and various correspondences mediate the relations between them. A sentence is grammatical if it conforms to the rules of every level of representation. Because of this, LFG is said to be a constraint-based, representational theory, where grammaticality

[1] Or more precisely, they are primitives for syntax, as later developments in LFG theorized that these notions may be decomposed into argument-structural features. The details of these developments will be discussed in section 2.4.

depends on obeying constraints on representations and not on properties of derivations. The main levels of representations are the following:[2]

- **Constituent-structure**: simple and flexible tree-structures to represent constituency and the linear order of words.
- **Functional-structure**: attribute-value matrices, they represent the grammatical and functional relations within sentences.
- **Argument-structure**: the module where various diatheses and lexical semantic information are represented.
- **Information-structure**: this is where information-packaging properties of sentences are represented.
- **Phonological-structure**: the layer where phonological information (stress, intonation) is represented.
- **Semantic-structure**: the structure containing the logical-semantic analysis of sentences using linear logic and Glue Semantics.

Out of these, the first four are the most relevant for the purposes of this dissertation, so in the following sections, I provide a more detailed description of them. The interested reader is referred to Dalrymple (2001, Chapter 9) and Dalrymple & Mycock (2011) for further information about phonological- and semantic-structure.

The final general property of LFG to be mentioned is monotonicity. This is related to the non-derivational nature of the theory: information may be added to the representation but not removed or changed. Monotonicity contributes to the computational plausibility of the theory.

2.2 Constituent-structure

Constituent-, or c-structures are essentially syntactic trees, but in LFG, they are much more compact than they are in Chomskyan frameworks. Their role is to represent constituency and linear word order. The guiding principle for c-structures in LFG is the principle of Economy of Expression (see Bresnan 2001:91, Dalrymple 2001:85, Falk 2001:33). Here, I quote Falk's formulation:

[2] Besides these, there have been proposals in the literature about a separate structure for representing (some) morpho-syntactic information (inflection, case, agreement), see Butt et al. (1999) and Falk (2006a).

(3) **Economy of Expression**: all syntactic phrase structure nodes are optional and are not used unless required to create a well-formed functional-structure or to add semantic content.

LFG's approach to phrase structure has its roots in the X'-theoretic approach of Chomskyan models, but as a result of Economy of Expression, there are crucial differences as well.

The standard lexical category-labels are readily recognized by LFG as valid c-structural categories: V, N, A, P, Adv. As for functional categories, in accordance with the Economy of Expression, the LFG approach is that they are recognized for a particular language, only if there is an overt syntactic correlate for the category. From this it follows that the presence or absence of functional categories in a given language has to be argued individually. So, only because it is generally accepted that English has the functional category "I(nflection)", it does not follow that every other language has to have it. Apart from "I", the most common functional categories in LFG are Complementizers (C) and Determiners (D).[3]

While LFG recognizes the standard X'-schema, depicted in Figure 2a, there are certain important modifications/additions. First, there is no strict binary-branching requirement. Second, as all nodes are optional, if there is no overt element in a position, the corresponding node is simply not projected. This may result in "headless" constructions. Let us take an embedded wh-question as an example, as in Figure (2b):

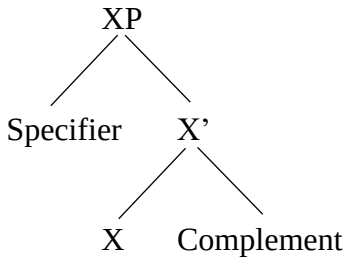

Figure 2a.
X'-schema.

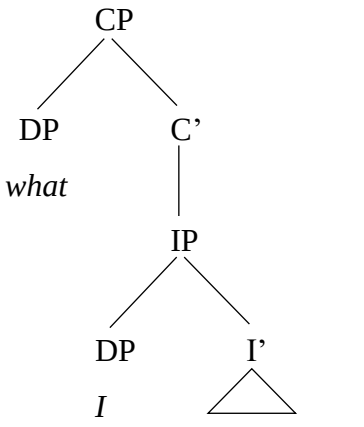

Figure 2b.
C-structure of *(I asked) what I should read.*

In Figure 2b (taken from Falk 2001:42), the question word is in Spec/CP, as standardly assumed, but as there is no overt complementizer, there is no C head, with a stipulated null complementizer.

[3] Also, Butt & King (1999) argue that Hindi has the functional category K, for certain case-markers.

What this means is that tree-structures in LFG are allowed to be exocentric. From this it follows that LFG recognizes the category S, which never has a head. It is usually assumed for non-configurational languages (like Warlpiri), although Dalrymple (2001) mentions that it has been hypothesized for a number of languages with relatively fixed word-order as well (e.g. Welsh, Icelandic).

The formal tools in LFG to specify possible c-structures for a given language are immediate dominance (ID) and linear precedence (LP) rules. ID rules are similar to traditional phrase structure rules, but they do not specify the linear order of constituents, which are defined via LP rules. So an ID rule like (4a) actually represents both phrase structure configurations in (4b). The irrelevance of ordering is notationally represented by the comma between the constituents. (4c) is an LD rule, specifying the order as verb-noun phrase.

(4) a VP → V, NP

b VP → V NP ; VP → NP V

c V < NP

Since the ID-LD distinction is not relevant for this dissertation for the sake of simplicity I will ignore this aspect of LFG. For this reason, I will use the comma-less representation in (4b), which does indicate the order of constituents. So it will be interpreted as indicating dominance and linear relations at the same time. I will also provide annotated tree-structures so ordering should be straightforward throughout the dissertation.

2.3 Functional-structure

Functional structures (f-structures) are the other side of syntactic descriptions in LFG. They are attribute-value matrices which represent information about grammatical and functional relations within sentences. As has already been noted, grammatical functions are syntactic primitives in LFG.

LFG standardly recognizes the following grammatical functions: subject (SUBJ), object (OBJ), oblique (OBL), secondary object/thematically restricted object (OBJΘ), closed complement (COMP), open complement (XCOMP), possessor (POSS), closed adjunct (ADJ), open adjunct (XADJ).

From the list above, subjects, objects and possessors are standard and well-known grammatical functions. The OBL function is assigned to elements that are associated with a unique semantic role by their predicates and usually bear idiosyncratic case, marked by an

adposition (e.g. *John appeals to Mary)* or case affixes. Because of the thematic restrictedness, obliques sometimes are referred to as OBLΘ. Secondary objects are objects that are specified for some semantic role, like dative shifted theme objects in English (*John gave Mary an apple*) or applicative objects in some Bantu languages (Kibort 2008).

Clauses that host their own subject are standardly assumed to have the closed complement function COMP (*I hope that John passes the exam*). Some researchers have expressed doubts about the necessity of the COMP function (e.g. Alsina et al. 2005, Forst 2006) and it has been proposed that it should be abandoned and clauses that had been analyzed as bearing this function are OBJs, OBJΘs or OBLs. However, arguments for the original approach have also been evoked.[4] In this dissertation, I follow the approach which could be regarded as standard today, according to which CPs may be either COMPs or OBJs. Nothing crucial hinges on this for the analyses presented in the subsequent chapters.

Clauses which lack an independent subject are analyzed as having the open complement function XCOMP. These include for instance infinitival clauses of "raising"-sentences (*John seems to be happy*) and other predicative complements (*Mary didn't sound ashamed of herself*). The open adjunct function XADJ is a direct parallel to this, on the realm of adjuncts (e.g. *Mary arrived drunk*).

Originally, LFG also posited that topic (TOPIC) and focus (FOCUS) are grammaticalized discourse functions and they are represented in f-structure. However, later research has gone in the direction where their representation is entirely relocated to the separate information-structure. Some researchers (Alsina 2008, Asudeh 2011) have suggested that dislocated/extracted elements, associated with discourse functions receive some general label at f-structure. In this dissertation I adopt Asudeh's (2011) suggestion of the Unbounded Dependency Function (UDF) for this purpose. Since discourse functions are always additional upon a normal grammatical function, they are sometimes called "overlay" function.

UDF, ADJ and XADJ are somewhat different from the other functions mentioned, as they are not uniquely specified, but usually represented as sets. This is so for the obvious reason that a sentence may contain multiple topics, foci or adjuncts.

Following Falk (2001), we may set up the following taxonomy of grammatical functions:

[4] On this topic, see Dalrymple & Lødrup (2000), Alsina et al. (2005), Forst (2006) and Lødrup (2012).

overlay	nonoverlay		overlay
argument functions		nonargument functions	
SUBJ[5]	OBJ, OBL, OBJΘ, COMP, XCOMP, POSS	ADJ, XADJ	UDF

Table 1.
An LFG-taxonomy of grammatical functions.

To put these into practice, let us demonstrate f-structures through an example. Consider sentence (5) and its f-structure in Figure 3.

(5) *Gavrilo Princip assassinated Franz Ferdinand I.*

$$\begin{bmatrix} \text{PRED} & \textit{assassinate} \langle (\text{SUBJ})(\text{OBJ}) \rangle \\ \text{TENSE} & \text{past} \\ \text{SUBJ} & \text{PRED} \left[\textit{Gavrilo Princip} \right] \\ \text{OBJ} & \text{PRED} \left[\textit{Franz Ferdinand I} \right] \end{bmatrix}$$

Figure 3.
F-structure for (5).

Figure 3 shows that the main predicate of the sentence is *assassinate*, which subcategorizes for a subject and an object. These grammatical functions are present in the f-structure (*Gavrilo Princip* and *Franz Ferdinand I*), so it is well-formed.

All well-formed f-structures conform to the three well-formedness constraints of LFG (there are several formulations of the constraints; here I quote that of Dalrymple 2001):

- **Completeness**: an f-structure is *locally complete* if and only if it contains all the governable grammatical functions that its predicate governs. An f-structure is *complete* if and only if it and all its subsidiary f-structures are locally complete.
- **Coherence**: an f-structure is *locally coherent* if and only if all the governable grammatical functions that it contains are governed by a local predicate. An f-structure is *coherent* if and only if it and all its subsidiary f-structures are locally coherent.
- **Consistency**: in a given f-structure a particular attribute may have at most one value.

[5] SUBJs are usually assumed to be default topics; this is why this grammatical function is listed as an overlay function.

Of these three, consistency is the most straightforward. It rules out sentences with value-mismatches. For example, a sentence like *A Serbians assassinated Franz Ferdinand I* is ruled out, since the indefinite article signals that the subject is singular, while the *–s* inflection marks it as plural, so the (sub-)f-structure of the subject would contain contradictory number-information.

Completeness and Coherence are basically two sides of the same coin. They ensure that all and only subcategorized grammatical functions are present in the f-structure. So (6a) would violate Completeness (because of the missing object), while (6b) would violate Coherence (by containing an "extra" element, which cannot be linked to any grammatical function subcategorized for by the predicate).

(6) a **Gavrilo Princip assassinated.*

b **Gavrilo Princip assassinated Franz Ferdinand someone.*

It is important to note that so far we have defined Completeness and Coherence on the syntactic, functional level. With this in mind, consider (7a-b) and their f-structures (Figure 4a-b).

(7) a **Gavrilo Princip seemed that someone assassinate the emperor.*

b **There assassinated the emperor.*

$$\begin{bmatrix} \text{PRED} & [\ \textit{seem}\ \langle(\text{COMP})\rangle(\text{SUBJ})\ ] \\ \text{SUBJ} & [\ \text{PRED} \quad \textit{Gavrilo Princip}\ ] \\ \text{COMP} & \begin{bmatrix} \text{PRED} & \textit{assassinate}\ \langle(\text{SUBJ})(\text{OBJ})\rangle \\ \text{SUBJ} & [\text{PRED} \quad \textit{someone}\ ] \\ \text{OBJ} & [\text{PRED} \quad \textit{the emperor}\ ] \end{bmatrix} \end{bmatrix}$$

Figure 4a.
F-structure for (7a).

$$\begin{bmatrix} \text{PRED} & \textit{assassinate} \langle (\text{SUBJ})(\text{OBJ}) \rangle \\ \text{SUBJ} & \begin{bmatrix} \text{FORM} & \textit{there} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \textit{the emperor} \end{bmatrix} \end{bmatrix}$$

Figure 4b.
F-structure for (7b).

The sentences in (7) conform to the well-formedness constraints as we have formulated them: all and only the subcategorized grammatical functions are present, and there is no value-inconsistency. What is the problem then?

The problem lies in the semantic sides of the Coherence and Completeness conditions. Semantic Coherence is violated in (7a), while (7b) violates Semantic Completeness. The explanation is the following. As commonly accepted, *seem* only has a propositional argument. So the subject of *seem* is not thematic, it does not have a theta-role. This property is represented by the COMP appearing inside and the SUBJ outside the angle brackets in the lexical entry of *seem*. Yet, in (7a), the subject is an element with semantic content, *Gavrilo Princip* (it has a PRED feature), but this SUBJECT is not associated with a semantic argument. (7b) has the mirror image of this problem. *There* in this sentence is assumed to be an expletive, not a meaningful element. This is represented by it not having a PRED attribute, only a FORM-one, as expletives are assumed to do so in LFG. But *assassinate* requires a meaningful, PRED-bearing subject (an agent, in semantic terms).

Such configurations are barred by the Semantic Coherence and Completeness Conditions, which specify that every meaningful argument function in an f-structure must be subcategorized for by a predicate (Semantic Coherence, violated by 7a) and every thematic argument function subcategorized for by a predicate must be associated with a meaningful element in the f-structure (Semantic Completeness, violated by 7b). Syntactic and semantic Completeness and Coherence coincide in most cases (like in 5, where each syntactic argument is a semantic argument as well), the difference only surfaces if nonthematic subcategorized grammatical functions or nonthematic constituents appear in the f-structure (as in 7).

A final issue that must receive a mention regarding functional structure is that LFG possesses mechanisms that make it possible to state identity relations between distinct f-structural entities. There are two such mechanisms: functional and anaphoric identification.[6]

[6] The terms “functional control” and “anaphoric control” is also used. I prefer “identification” so as to avoid any collision with other uses of the term “control” (see section 6.3.1).

Functional identification is a strict, formal identity relation: it allows one element to satisfy two grammatical functions at the same time. This happens for example in the case of wh-questions. In (8), for example, according to the LFG-analysis, the question word simultaneously serves as the OBJ of the sentence and it is also an Unbounded Dependency Function (UDF). This is graphically represented as a solid line between the two functions in the f-structure.

(8) *Whom did he assassinate?*

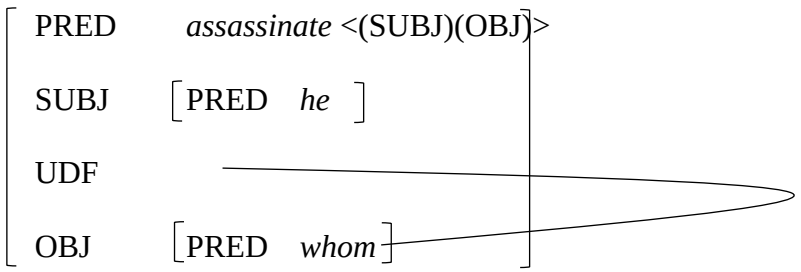

Figure 5.
F-structure for (8).

Functional identification is also utilized in "raising" sentences, like the ones in (9). In (9a) *Gavrilo Princip* is the subject of the main clause and the embedded infinitival clause at the same time. Similarly, in (9b), *Gavrilo Princip* is the object of the main verb and at the same time, it is the subject of the infinitival clause. In both cases, *Gavrilo Princip* is thematically only related to the embedded predicate.

(9) a *Gavrilo Princip seems to have assassinated the emperor.*
b *I believe Gavrilo Princip to have assassinated the emperor.*

Anaphoric identification is a looser kind of dependency. Essentially, it states the referential identity of two f-structural elements, without assuming a syntactic identity. It is primarily assumed in the analysis of "equi"-type control sentences, like (10). (a characterization of "raising" and "equi" structures will be provided in section 6.3).

(10) *Gavrilo Princip agreed to assassinate the emperor.*

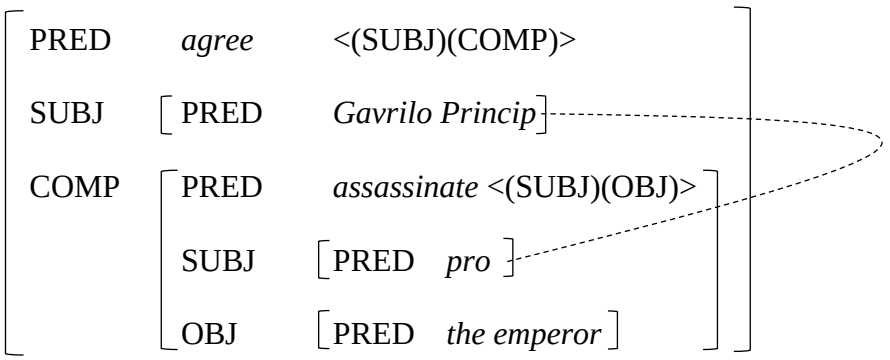

Figure 6.
F-structure for (10).

Unlike (9), here *Gavrilo Princip* is thematically related to both subject-positions. The details of the analysis are left for section 6.3, here I just give a sketch. Basically, the unexpressed embedded subject is represented in f-structure as a "pro"[7], and the main clause subject is constrained to be identical with it. This is informally represented here with the dotted line. Note that the dotted line is only used here for expository purposes; it is not a standard LFG representation. (The standard LFG representation displays INDEX attributes with identical values for the identified elements.)

2.4 Argument-structure

Argument structure is the locus of morphosyntactic and morphosemantic operations on lexical entries, so it may be regarded as the engine for syntax for LFG. Although it was not part of the original theory of LFG as it was put forward in Bresnan ed. (1982), it is now an indispensable part of the LFG architecture. The module of LFG that deals with the properties and mechanisms of argument-structure is Lexical Mapping Theory (LMT).

Bresnan and Kanerva (1989) proposed that grammatical functions can be decomposed via the features +/– o (object-like) and +/– r (semantically restricted). The feature-system builds on the intuition that SUBJ and OBJ may be associated with any semantic role (or even with semantically empty idiom chunks as in *The cat seems to be out of the bag* and *I believe the cat to be out of the bag*.), while OBL and OBJΘ are always associated with a specific semantic role. On the other hand, OBJ and OBJΘ are both object-functions while SUBJ and OBL are not. The following taxonomy is the result:

[7] Note that unlike the Chomskyan treatment this "pro" is only an f-structural entity. In accordance with Economy of Expression, no zero c-structural node is assumed.

	–o	+o
–r	SUBJ	OBJ
+r	OBL	OBJΘ

Table 2.
LMT's feature decomposition of grammatical functions.[8]

The arguments of the predicates are associated with these features and then mapping principles determine the grammatical functions of these elements. In what follows I briefly describe the most recent version of mapping in LMT, developed by Kibort (2007).

In Kibort's (2007) system, the arguments of a predicate are taken to have a set of entailments, as the participants of the event denoted by the predicate. These are associated with a fixed and universal valency-template, shown in (11).

(11) -o/-r -r +o -o … -o
arg1 arg2 arg3 arg4 arg_n

Once the semantic role entailments of the predicate are associated with the argument positions offered by the template, grammatical function assignment takes place according to the following principle (Kibort 2007):

(12) a **Mapping Principle**: The ordered arguments are mapped onto the highest (i.e. least marked) compatible function on the markedness hierarchy.
Markedness: having positive feature-specification.

b **Markedness hierarchy**: SUBJ > OBJ, OBLθ > OBJθ

When a morphosyntactic operation like passivization takes place, it is modelled as increasing the markedness (adding positive specification) of the given slot of the valency-template.

Using Kibort's (2007) version of LMT let us use (13) as an illustration. (14) contains the argument-structural representation.

[8] One may wonder about the status of COMPS and XCOMPS in LMT, as the feature decomposition does not include them. This relates to the debate mentioned earlier about the legitimacy of COMP and XCOMP as independent grammatical functions. Kibort (2007) takes them to be special kind of OBLs, following the suggestion of Zaenen & Engdahl (1994). Other researchers (Alsina et al. 2005, Forst 2006) suggest that COMP should be abandoned and its occurrences should be analyzed as OBJs, OBLs and OBJΘs. As noted, this is not the focus of this dissertation, I remain neutral on this issue.

(13) a *Jack kissed Monica.*

b *Monica was kissed by Jack.*

(14) a	*kiss*	ag	pat	("agent" and "patient")
b		–o arg1	–r arg2	← universal valency frame
c		SUBJ	OBJ	← active mapping
d		–o/+r	–r	← increasing markedness
e		OBL	SUBJ	← passive mapping

The first argument (the "agent", using a traditional label) and the second argument ("patient") of *kiss* (14a)[9] occupy the first two argument slots in the universal valency frame (14b). The least marked grammatical functions compatible with the specification are SUBJ and OBJ, so it is determined that the arguments of *kiss* will receive these functions (14c). When passivization takes place, a positive specification (+r) is added to the first argument slot (14d). This changes the mapping of the argument to OBL, that is, a *by*-phrase (Kibort assumes that the *by*-phrase is an argument, not an adjunct). Consequently, now the SUBJ function is available to the second argument (the –r specification is compatible with SUBJ) (14e).

Apart from morphosyntactic argument-structure changes, which only affect the mapping of grammatical functions, but not the semantics of predicates, there also exist morphosemantic operations, which do have effects on the meaning of the predicates. One example is the dative shift. Although the prepositional dative (15a,c) and the double object dative (15b,d) are obviously related lexical entries, there is a meaning difference between them. For instance, it has often been noted that the double object dative is incompatible (at least in standard dialects) with an abstract beneficiary (15c).

(15) a *John sent a letter to Mary.* b *John sent Mary a letter.*

c *John sent a letter to England.* d **John sent England a letter.*

According to Kibort (2007) such meaning differences arise because the valency slots are associated with different meaning entailments, e.g. the arg1 slot is for proto-agents, arg2 is for proto-patients and arg3 is for proto-themes. These morphosemantic variations are modelled in this version of LMT as different alignments of the arguments of the predicate on the valency

[9] For expository purposes, one may use traditional theta-role labels, but Kibort (2007) prefers to use variables like x and y, indicating that the exact semantic load may be idiosyncratically determined by the predicate.

frame. So in the prepositional version, the beneficiary is mapped to the 4th argument slot, while double object dative maps it to the 2nd position.[10]

(16)	a	*send*	ag	pat	ben	("agent", "patient" and "beneficiary")
	b		–o	–r	–o	← universal valency frame
			arg1	arg2	arg4	
	c		SUBJ	OBJ	OBL	← prepositional dative (15a,c)
	d	*send*	ag	ben	pat	
	e		arg1	arg2	arg3	← universal valency frame
			–o	–r	+o	
	f		SUBJ	OBJ	OBJΘ	← double object dative (15b,d)

Note that in the double object dative, the patient receives a +o specification, which makes it impossible for it to map to SUBJ. Hence, only the beneficiary can be a passive SUBJ (**A letter was sent Mary* vs. *Mary was sent a letter*).

Apart from cases where the dependents of the verbs are arguments (direct semantic dependents of the predicates), there are a number of cases in which the exact nature of the element in question shows an intermediate status. Let us take optional beneficiaries as examples (17a-b). Toivonen (2013) shows that they display a mixed behavior on standard argumenthood tests (17c-e). Here I survey the most straightforward cases, for the details of the tests, see Toivonen (2013) and references therein.

(17)	a	*John baked <u>Mary</u> a cake.*	
	b	*John baked a cake <u>for Mary</u>.*	
	c	*John baked (Mary) a cake.*	adjuncts are optional → adjunct
	c'	*John baked a cake (for Mary).*	
	d	**Who did you bake the parents of a cake?*	adjuncts are islands for extraction → adjunct
	d'	**Who did you bake a cake for the parents of?*	
	e	**What John did Mary was bake a cake.*	only adjuncts can form a VP-focus pseudocleft → argument
	e'	*What John did for Mary was bake a cake.*	

[10] The 3rd position for the beneficiary is not available in standard English, aside from some marginal examples like *??You can give it me back.* It is productive in other languages, e.g. Polish (Kibort 2008).

Obviously, the sharp distinction between arguments and adjuncts must be smoothened somehow. There are basically two ways one can approach the issue: from the side of arguments and from the side of adjuncts. Both directions have been pursued in LFG.

Let us begin with adjuncts. Rákosi (2006, 2012) argues that there is a wide range of predicates that license thematically marked adjuncts, ADJΘs. Such elements are not arguments, as they are not entailed by the meaning of the predicate. Still, when they are present they can be shown to bear a closer relationship to the predicate than regular adjuncts do, which are fully external to the event denoted by the predicate. An example for this is dative experiencers. (18a) shows a dative experiencer argument, while (18b) is a dative experiencer thematic adjunct. This can be seen from the fact that while the morphology of the argument is fixed (as expected), the adjunct shows variation.

(18) a *Christie appeals to/*for Bill.*
b *Christie is important to/for Bill.*

That thematic adjuncts still have closer ties to the predicate than regular adjuncts may be observed for example in Hungarian, where thematic adjuncts may appear in a predicate-internal c-structure position (19a), as opposed to non-thematic ones (19b) (Rákosi 2006:141).

(19) a *Fontos nekem, hogy itt maradsz.*
important DAT.1SG that(C)[11] here stay.2SG
'It is important to me that you stay here.'

b **Butaság nekem, hogy itt maradsz.*
stupidity DAT.1SG that(C) here stay.2SG
'It is a stupidity to me that you stay here.'

It is an idiosyncratic property of each predicate what kind of thematic adjunct it can combine with, e.g. *important* may license an experiencer but not an instrument, but *bake* may take a beneficiary.

As for the argument-side, Needham & Toivonen (2011) and Toivonen (2013) have put forward the idea that some arguments may be best characterized as derived arguments, arguments that are not part of the initial argument list, but added through argument structure operations. In fact, Needham & Toivonen (2011) analyze all the PPs in (17) as derived arguments. Rákosi (2012) shows that this may not be a very favorable move for several reasons.

[11] I use the gloss "that(C)" to indicate that *hogy* is a complementizer, not a distal demonstrative. (These are homophonous in Hungarian.)

His argumentation includes syntactic issues (they often have alternative expressions), semantic issues (true optionality) and implementational issues (the burden of creating OBL-ADJ ambiguities, since then in principle every PP can be analyzed as an adjunct or as a derived oblique argument) as well.

But even if one eliminates such PPs as potential derived arguments, there are a number of cases where such an analysis may be justified. Toivonen (2013) argues that English beneficiary NPs are such elements (17a: *John baked Mary a cake*.). Here the beneficiary clearly occupies an argumental c-structural position, yet shows adjunct-properties as well. As we have seen, in this case, the beneficiary alternates with a thematic adjunct (the PPs in 17). A similar situation may be observed in the Swedish Copy Raising construction.

(20) a *Det verkar pa Tom som om han har vunnit.*
it seems on Tom as if he has won
'Tom gives the impression that he has won.'

b *Tom verkar som om han har vunnit.*
Tom seems as if he has won
'Tom seems as if he has won.'

Although the issue is not completely settled, a good case may be made for analyzing *Tom* as an argument in (20b) (see Landau 2011), which state of affairs would mean that the adjunct *pa Tom* in (20a) is "argumentalized". More details about the Copy raising construction will be provided in section 6.3.2.1.1.

What this line of thought aims to show is that there seems to be some mechanism in the argument structure which can mediate between (thematic) adjuncts and derived arguments. This procedure takes some adjunct from the thematic domain as input, and creates a core argument out of it (an OBJ in the case of English beneficiary NPs and a SUBJ in the case of Swedish Copy Raising).

2.5 Information-structure

Information-structure (i-structure, IS) here is conceptualized as an autonomous level of linguistic representation, which, following Krifka and Musan (2012:1) may be defined as those aspects of linguistic representation which concern not the information-content of a sentence itself, but rather the way it is presented, "packaged", and also "helps speakers to take into consideration the addressee's current information state, and hence to facilitate the flow of communication". The main purpose of IS is the structuring of the discourse: to mediate and

assist the integration of the sentence into the conversation. Thus, it will be argued that the nature of the relationship between elements of an utterance and the wider discourse is a crucial notion in the structure of i-structure.

Particular IS categories may also have ties to the narrowly defined semantic structure. Examples for this are the often noted referentiality requirement on topics or the exhaustivity effect of certain focus-structures. Occasional references will be made to such semantic aspects of IS, but in the spirit of strong modularity, my primary perspective is that these should be detached from information-structure proper. What this amounts to is that i-structural categories like "topic" and "focus" may have properties which are not derived from i-structure itself, but from the way IS interacts with other levels of linguistic representation.

Information-structure was not part of LFG at its origins. As we noted in 2.3, information-structural labels like TOPIC and FOCUS were used only as f-structure elements. The addition of i-structure as an independent label was motivated by two factors. On the one hand, it seemed that a number of phenomena pertaining word-order variations may be traced back to different information-structural status of the elements involved. This was the primary drive behind Choi's (1996) analysis of German and Korean scrambling and Butt & King's (1996) treatment of word-order variations in Urdu and Turkish. On the other hand, King (1997) showed that there can be mismatches between f-structure and i-structure. This, given that in LFG, f-structures feed the semantic module, could lead to interpretational anomalies. An illustration from King (1997) is the case of *it*-clefts.

(21) A: *Was it the ex-convict with the red SHIRT that he was warned to look out for?*
B: *No, it was the ex-convict with the red TIE that he was warned to look out for.*

Since on the surface, the entire phrase *the ex-convict with the red SHIRT* is fronted, the constituent would have to bear the FOCUS grammaticalized discourse-function at f-structure. But this is not right, since only the piece of clothing is in focus. With no i-structure, the mismatch cannot be resolved. As a solution, a new level of representation was posited, where bare PRED values could be specified as having discourse functions.

Since then, several frameworks have been put forward about the representation of IS in LFG. A common thread running through them is the decomposition of information-structural categories with features. The advantage of this approach is that it is possible to capture the commonalities among these notions. However, as we will see, there is no consensus on the number of the needed IS-categories and on the needed features either. Before presenting my proposals about IS, here I briefly introduce some earlier LFG-approaches. As the elaboration on the IS features and categories will be provided in the next section, no in-depth discussion

about them will here be provided. The purpose is just to give the reader a bird's-eye view of the i-structural landscape.

Choi (1996), in her account of German and Korean scrambling, was one of the first researchers to incorporate information-structure in an analysis in an LFG-setting. Choi's framework involves two independent features: PROM(inent) and NEW. These features could have either positive or negative specifications, yielding the following taxonomy for IS-categories:

	+NEW	**–NEW**
+PROM	Contrastive Focus	Topic
–PROM	Completive Focus	Tail

Table 3.
IS-taxonomy of Choi (1996).

In this framework, +PROM means that the discourse function is prominent in the discourse (in some sense) and +NEW means that it introduces novel information. I illustrate i-structural categories below (these examples are from Choi 1996, the capital letters indicate phonological stress).

(22) a A: *What about the money? Who did you give it to?*
B: *I gave [the money]*$_{TOPIC}$ *to [the cashier]*$_{COMPLETIVE\ FOCUS}$.

b A: *To whom did you give the money?*
B: *I gave [the money]*$_{TAIL}$ *to the cashier.*

c A: *What did John buy?*
B1: *John bought a [Toyota]*$_{COMPLETIVE\ FOCUS}$.
B2: *John bought a [TOYOTA]*$_{CONTRASTIVE\ FOCUS}$ *(not a VOLKSWAGEN)!*

In (22a/B) *the money* is a discourse-old but salient entity that the sentence is about. The very same entity in the very same sentence is not considered to be a Topic by Choi (1996) if it is not in the center of attention in the given discourse: rather it is labelled "tail" in (22b), following Vallduví's (1992) early work on information-structure. As *the cashier* in (22a/B) and *Toyota* in (22c/B1) provides an answer for the respective questions, they are considered Completive foci. If an extra emphasis or prominence is added to the answer, as in (22c/B2), then it is categorized as Contrastive focus.

Butt & King (1996) offered another version of the four-way distinction with two features, in their account of word-order in Urdu and Turkish. They used the same features, but the set of discourse functions occupying the taxonomy is different:

	+NEW	–NEW
+PROM	Focus	Topic
–PROM	Completive information	Background information

Table 4.
IS-taxonomy of Butt & King (1996).

This framework has been used in several subsequent works, see e.g. Butt & King (1997), Dalrymple & Nikolaeva (2011) and Mycock (2013). "Background information" is basically Choi's (1996) and Vallduví's (1992) "Tail". Only one type of focus is distinguished. However, Butt and King (1997) introduce the category "Completive information" for such elements that introduce new information into the discourse but are not prominent (they are not directly relevant/highlighted in the discourse: they do not answer questions, provide emphasis, correction, etc.). An example is shown in (23) from Butt & King (1997).

(23) A: *Where is Nadya coming from?*
B: *She was just buying [toffee]* $_{\text{COMPLETIVE INFORMATION}}$ *at the market.*

Cook & Payne (2006) replace PROM with TOPICAL and they also enrich the feature-set, adding a CONTRASTIVE feature. This move doubled the available discourse functions, including an unattested one (Contrastive tail).

		+NEW	–NEW
+TOPICAL	+CONTRASTIVE	Contrastive new-information topic	Contrastive Old-information topic
	–CONTRASTIVE	New-information topic	Old-information topic
–TOPICAL	+CONTRASTIVE	Contrastive focus	Contrastive tail
	–CONTRASTIVE	Non-contrastive focus	Tail

Table 5.
IS-taxonomy of Cook & Payne (2006).

Contrastive, non-contrastive focus and Tail should sound familiar from the previous taxonomies. This is the first LFG-system where contrastiveness is explicitly represented (Choi 1996 had Contrastive focus, but the "contrastiveness" was to be construed from prominence in some way). The NEW feature distinguishes between previously and newly introduced referents.

Cook & Payne (2006) do not provide systematic minimal pair examples, so the following examples, illustrating contrastive old/new topics, are mine.

(24) a A: *What do you know about John and Mary? Are they at home?*
B: *[John]* $_{\text{CONTRASTIVE OLD TOPIC}}$ *is at home, but [Mary]* $_{\text{CONTRASTIVE OLD TOPIC}}$ *is in the office.*

b A: *Where are your friends?*
B: *[John]* $_{\text{CONTRASTIVE NEW TOPIC}}$ *is at home, but [Mary]* $_{\text{CONTRASTIVE NEW TOPIC}}$ *is in the office.*

Gazdik (2011) has the PROM feature but replaces NEW with D(iscourse)-LINKED. Discourse-linked elements "link the sentence to the discourse by introducing a subtopic of the discourse topic or by reshaping the discourse topic" (Gazdik 2011:284). As this issue is highly relevant for my view of IS, I will scrutinize this aspect of Gazdik's system in section 2.6.1.2.

	+D-LINKED	**–D-LINKED**
+PROM	Thematic shifter, Contrastive topic	Focus, Hocus
–PROM	Background information	Completive information

Table 6.
IS-taxonomy of Gazdik (2011).

Thematic shifters are topics which move the conversation forward by introducing a new subtopic. Gazdik (2011) also includes the category "Hocus", which was originally proposed by Kálmán (1985), for newsworthy preverbal NPs in neutral Hungarian sentences. (25) is an example.

(25) *János tegnap [vonattal]* $_{\text{HOCUS}}$ *utazott haza.*
John yesterday train.with travelled.3SG home
'John yesterday took the train to go home.'

It should be evident now that the landscape is not particularly clear. Three central issues obscure the view: first, it is not clear how many discourse functions we should distinguish; second, it is not clear what features we should use, and third, it is often inadequately defined what these features really mean. In what follows I present my views on these topics and outline my proposal about the organization of information-structure.

2.6 A new proposal for information-structure

In this section I aim to construct an IS-taxonomy that builds on all the previous approaches and also improves upon them. The improvement involves two aspects: firstly, my proposal will include all (and only) the well-established IS-categories, and secondly, I will attempt to provide a clearer definition of the features than the existing frameworks do. My proposal is shown in Table 7.

		+NEW	–NEW
+D-STRUCTURING	+CONTRASTIVE	Contrastive focus	Contrastive topic
	(–CONTRASTIVE)	Information focus	Topic
–D-STRUCTURING		Completive information	Background information

Table 7.
The proposed IS-taxonomy.

The –CONTRASTIVE feature is in brackets because contrastiveness is viewed as an additional, superimposed discourse-structuring relation. The absence of it automatically means negative specification, so +D-STRUCTURING is –CONTRASTIVE by default. This is significant, as I will posit in 2.6.3 that contrastiveness is not the only possible additional discourse-feature.

In the following sections, I will investigate the content of these features and the properties of the IS-categories.

2.6.1 Features of information-structure

In the proposed system, six discourse functions are distinguished with the features NEW, D-LINKED and CONTRASTIVE.

Although used in several taxonomies, "prominence" as a feature is not present here. The reasons why I discarded it are the following. At the heart of the issue is that defining prominence is very problematic. It is an inherently subjective notion that can only be stated in relation of two entities. What is more, prominence is a gradient property, so using it as a +/– attribute seems off the track. Next, it should be pointed out that i-structure is a level of linguistic representation and prominence as such is not something that seems like an exclusively linguistic property: sounds, visuals, even ideas may also be prominent in some sense. Even in the realm of

linguistics, prominence is ambiguous, as it has phonological, syntactic and pragmatic aspects and these aspects do not necessarily behave in a parallel fashion. Ultimately, I will argue that the distinctions that the PROM feature makes are either unnecessary or are statable in terms of the features that I use, so PROM itself is superfluous. It is best understood as a derivative property that piggybacks on the more fundamental features that I utilize. As we will see, the concepts of newness, discourse-structuring and contrastiveness can be defined more straightforwardly and are able to provide us with the necessary distinctions.

2.6.1.1 +/–NEW

As Gundel (1999) points out much terminological confusion stems from the failure to distinguish between two senses of newness: the referential and the relational sense.[12] The referential sense is about "a relation between a linguistic expression and a corresponding non-linguistic entity in the speaker/hearer's mind" (Gundel 1999). The cognitive status of some entity may be characterized in various ways (completely novel, identifiable, familiar, etc.) and these statuses may have consequences for the linguistic encoding of the entities but the cognitive statuses themselves are not inherently linguistic.[13]

Relational newness on the other hand makes a directly linguistic distinction. It "involves two complementary parts, X and Y, of a linguistic or conceptual representation, where X is given in relation to Y, and Y is new in relation to X" (Gundel 1999). Consider the following example, from Lambrecht (1994:48).

(26) A: *When did you move to Switzerland?*
B: *When I was sixteen.*

As Lambrecht (1994:48) notes, "what constitutes the information conveyed by this answer is not the fact that at some point in his life the speaker was sixteen (…) but the RELATION [emphasis by Lambrecht] established between an act of moving to Switzerland, the person involved in that act, and the time at which the moving occurred". In other words, the answer provides a value for X in the proposition evoked by the question: "I was X (years old) when I moved to Switzerland".

[12] A similar point is made by Lambrecht (1994:48).

[13] For approaches about the effects of referential givenness see e.g. Gundel, Hedberg & Zacharski (1993), Prince (1981a), Ariel (1988) and Lambrecht (1994:165-172).

What Gundel (1999) refers to as X and Y is termed by Lambrecht (1994:52) as the "pragmatic presupposition" (X) and "pragmatic assertion" (Y) of the sentence. The pragmatic presupposition of a sentence is "the set of propositions lexicogrammatically evoked in a sentence which the speaker assumes the hearer already knows or is ready to take for granted at the time the sentence is uttered". The pragmatic assertion on the other hand is "the proposition expressed by a sentence which the hearer is expected to know or take for granted as a result or hearing the sentence uttered".

What is part of the presupposition and what is part of the assertion may be told apart by using the "lie-test" (Lambrecht 1994:52). Consider Lamrecht's example in (27).

(27) *I've finally met the woman who moved downstairs.*

Imagine someone challenges the claim in (27) by saying "That's not true!". This would be understood as rejecting the claim that the meeting took place, not that someone did move downstairs or the speaker has a neighbor.

Let us now illustrate that the referential and the relational senses are independent with two examples. Consider the following piece of conversation, from Gundel (1999).

(28) A: *Who called?*
B: *Pat said SHE called.*

Here *SHE* in the answer has a high level of referential givennes: it refers to a proper name, referring to a specific individual, it had just been mentioned and is in the center of attention. However, it still provides a new piece of information, by providing a value for the presupposed open proposition, "X called".

In turn, some referentially new material may still be part of the pragmatic presupposition. Consider the following exchange.

(29) A: *What's that big red spot on your arm?*
B: *A mosquito bit me.*
A: *That cannot be true! Mosquito-bites aren't like that!*

In B's response, the noun phrase *a mosquito* refers to a nonspecific entity, newly introduced to the discourse. However, as A's reply shows, applying the lie-test may leave the existence of the mosquitoes intact, which means that it is part of the pragmatic presupposition.

Out of these two notions of newness, it is the second, relational sense that I take to be relevant for information-structure. A working-definition is provided in (30).

(30) +/– NEW: A linguistic entity with a +NEW feature at information-structure provides relationally new information, by being part of the pragmatic assertion of the sentence. A linguistic entity with a –NEW feature at information-structure is relationally given, is part of the pragmatic presupposition of the sentence.

2.6.1.2 +/–DISCOURSE-STRUCTURING

Gazdik (2011), building on the theories of Büring (2003) and Asher & Lascarides (2003) distinguishes between those information-structural categories which are "d-linked"[14] and those that are not: the ones of the first type "link the sentence to the discourse by introducing a subtopic of the discourse topic or reshaping the discourse topic". I find the idea attractive but Gazdik's implementation unsatisfactory, so I make a proposal in the same spirit, but with crucial modifications.

I relabel the feature as +/–DISCOURSE-STRUCTURING (abbreviated as D-STR), for two reasons. One, I believe that "discourse-structuring" describes what I want to express more precisely than "discourse-linking". Two, I would like to avoid causing terminological confusion often found in the literature, where the same word is used with a number of related but still different definitions (as in the situation of "old information" and "new information", or the term "topic"; actually, Gazdik's (2011) use of the expression is already not exactly what Pesetsky 1987 means by it, see footnote 14 below).

To understand this feature let us first review the theories of Büring (2003) and Asher & Lascarides (2003). The discussion is restricted to the fundamentals, for the details, the reader is referred to the original works.

Büring's (2003) starting point is that all discourse is structured along a hierarchical web of questions and subquestions (an earlier formulation of the idea is to be found in Roberts 1996). The top of the hierarchy may be conceptualized as (the poetic-sounding) Big Question: "What is the way things are?". Questions that are of more immediate concern are Questions Under Discussion (QUD). The structure of the discourse may be represented with a discourse-tree. Consider the conversation in (31) and its representation in Figure 7 (slightly modified from Büring 2003:4).

[14] The term originates from Pesetsky (1987), for whom some question words are D-LINKED and others are not. For example, according to Pesetsky (1987), the wh-phrase in *Which book did you read?* is D-LINKED as it may be only felicitously answered if the speaker and the hearer already have a representation of a set of books in their minds.

(31) A: *How was the concert? Was the sound good?*
B: *No, it was awful.*
A: *How was the audience?*
B: *They were enthusiastic.*
A: *How was the band? How was the drummer?*
B: *Just fantastic.*
A: *And what about the singer?*
B: *Better than ever.*
A: *Did they play old songs?*
B: *Not a single one.*
A: *So what did you do after the concert?*

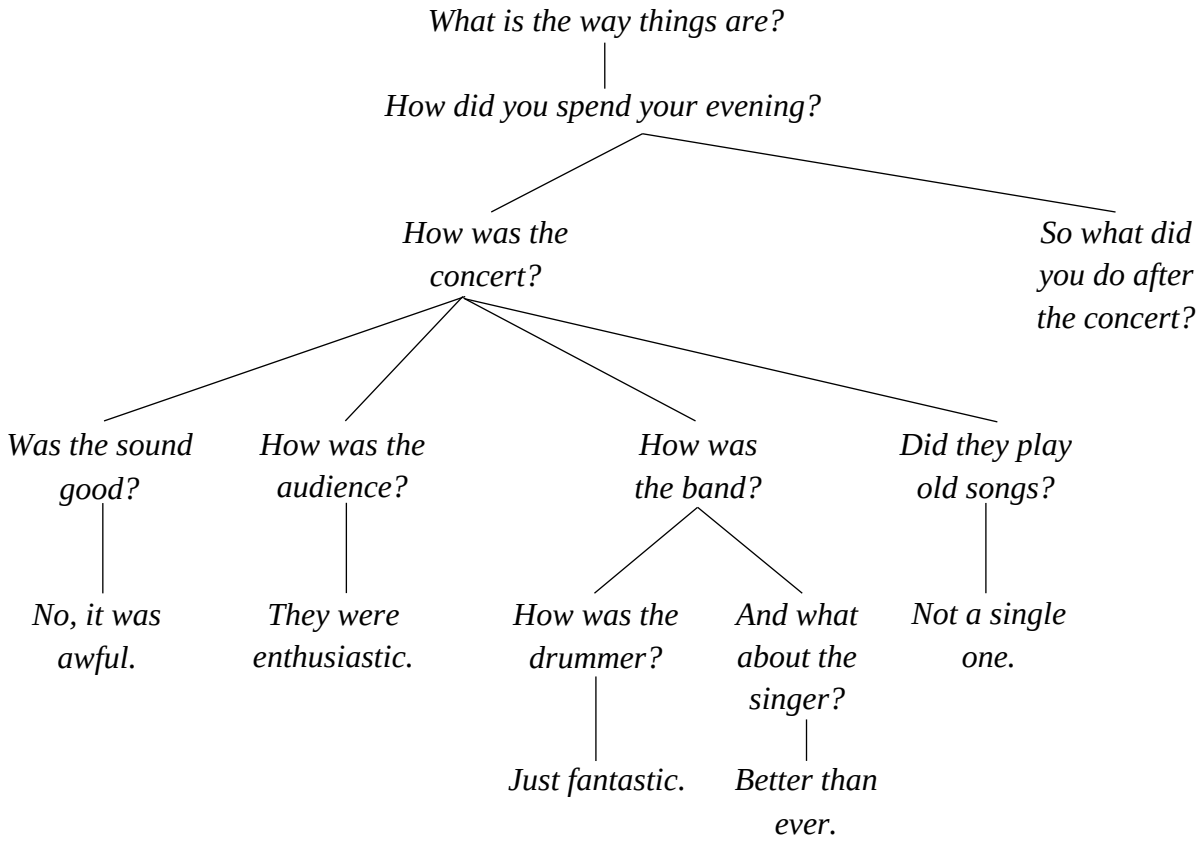

Figure 7.
Discourse-tree for (31).

As we will see in connection with the contrastive IS-categories, there are also cases where a subquestion is only implicitly present in the discourse-tree.

Gazdik (2011) also builds on the theory of Asher & Lascarides (2003) who analyze the semantics-pragmatics interface via rhetorical/discourse relations: narration, background, result, continuation, parallel, contrast, alternation, dialogue, correction elaboration, topic, explanation, consequence, and various question types. Gazdik (2011) then tries to correlate these relations with the i-structural categories, as in Table 8.

Information-structure role	Information-structure status	Discourse relation(s)
Thematic shifter	+PROM, +D-LINKED	narration, topic
Contrastive topic	+PROM, +D-LINKED	implicative answer
Focus	+PROM, –D-LINKED	answers, parallel, contrast, correction
Hocus	+PROM, –D-LINKED	narration
Question phrases	+PROM, +/–D-LINKED	question-answer pairs
Completive information	–PROM, –D-LINKED	any
Background information	–PROM, D-LINKED	any

Table 8.
Gazdik's (2011) view of i-structure roles and discourse relations.

I have already noted that I sympathize with the spirit of this approach, but I find several problematic aspects in the way it is fleshed out. My primary concern is with the featural composition of IS categories. Remember that according to Gazdik (2011:284), +D-LINKED means that the element "links the sentence to the discourse by introducing a subtopic of the discourse topic or reshaping the discourse topic". Given this, it is not clear to me why a Focus is not D-LINKED. Giving an answer for a question, offering a correction, or providing contrast does seem to me like things that reshape the discourse topic. Consider (32).

(32) A: *What concert did you attend yesterday?*
B: *Pink Floyd.*
A: *Oh, I like them too. What is your favorite album?*

Here, B's answer is a sentence including only a Focus. The rest of the conversation is about the band Pink Floyd, so the answer has introduced a new subtopic to the conversation and has reshaped the discourse topic. In fact, focussed elements prototypically do these things.

On the other hand, Background information is claimed to be +D-LINKED, which I find highly counterintuitive. Why would pieces of information, which are already known, repeated and clearly out of the center of attention participate in restructuring the discourse? Their D-LINKEDNESS only involve the fact that they are discourse-anaphoric (that is: referentially old, see the previous section), but this is not what Gazdik (2011) seems to argue D-LINKEDNESS to be. If discourse-anaphoricity were to be equaled with DISCOURSE-LINKEDNESS, then a corrective focus like the one in (33/B) would also be D-LINKED, contrary to Table 8.

(33) A: *I heard that you like [Deep Purple]$_{FOCUS}$ the most.*
B: *I like Deep Purple, but it is not [Deep Purple]$_{FOCUS}$ that I like the most. It's Pink Floyd.*

However, notice that Background information and Completive information share a property: they may be present in any discourse relation. This is because they do not participate in the structuring of the conversation in the sense that they are not directly related to Questions Under Discussion of the conversation, they are not involved in resolving these. Topics and foci directly structure the discourse: the former by being the entity that a new piece of information is predicated of, the latter by providing the new pieces of information that are needed for the resolution for a QUD. Both aspects may be stated in terms of the subquestions of the discourse, as we will see in section 2.6.2. This I claim is the core of the concept of being a +D(ISCOURSE)-STRUCTURING IS category.

(34) +/–DISCOURSE-STRUCTURING: a linguistic entity with a +DISCOURSE-STRUCTURING feature at information-structure participates in the structuring of the conversation, by being directly involved in the formation or resolution of questions under discussion of the discourse.
A linguistic entity with a –DISCOURSE-STRUCTURING feature at information-structure is not directly involved in the formation or resolution of questions under discussion of the discourse.

2.6.1.3 +/–CONTRASTIVE

There have been debates about the status of "contrast" in the linguistic literature on several levels. The central issue is whether contrast should be regarded as a concept with direct and independent linguistic relevance.

On one end of the scale is Lambrecht (1994:290) who claims that contrastiveness is a mere pragmatic effect (a conversational implicature) arising "from particular inferences which we draw on the basis of given conversational contexts" and it is not a category of grammar at all.

Another view (Roth 1985, Büring 1997, Krifka 2008) is that contrast may be regarded as a concept with direct linguistic relevance, but it is not an independent notion: it is a consequence of what the i-structural category Focus does. That is, every focus is contrastive in some sense (more on this below).

The third view, which is the closest to mine is the view of Molnár (2002, 2006) and Titov (2013). Once the necessary distinctions are made, contrast may be regarded as an

independent notion of information-structure (and thus, grammar in a looser sense)[15]. Evidence for this view is the widely-documented existence of linguistic structures which are associated with contrast, independent of focus, some of which will be discussed in this dissertation. For example, A-scrambling in Dutch is possible with elements that bear contrastive IS-categories, contrastive topics and contrastive foci, see (35), from Neeleman et al (2009).

(35) a *Ik geloof dat [alleen DIT boek]*$_{\text{CONTRASTIVE FOCUS}}$ *Jan Marie gegeven heft.*
I believe that only this book John Mary given *has*
'I believe that John has given ONLY THIS BOOK to Mary.' (and not another book)

b *Ik geloof dat [zo'n boek]*$_{\text{CONTRASTIVE TOPIC}}$ *alleen JAN Marie gegeven heeft.*
I believe that such-a book only John Mary given has
'I believe that such a book, only John has given to Mary.' (others may have given other books)

So what does it mean to be +CONTRASTIVE? It is clear that contrast is related to a set of alternatives. However, the set of alternatives may be defined on two levels. As Titov (2013) points out, referring to a set of alternatives is taken to form the basis of semantic computation for any kind of focus. Consider example (36) and a semantic interpretation for it in (37), from Kenesei (2006:144-145).

(36) A: *Where did you go in the summer?*
B: *I was [in Italy]*$_{\text{FOCUS}}$*.*

(37) a For which x, x a country, you went to x in the summer?
b ∃x, such that x ∈ {countries I went to in the summer}, and x is Italy

As it is clear from (37) answering the question does involve the semantic computation with reference to a set (a set of countries in this case). However, this should be distinguished from what Titov (2013) calls the "pragmatic set of alternatives", and what I also believe to be the relevant concept for i-structure. These are entities that are active in the minds of the hearer and the speaker once a contrastive element is uttered. That is, contrastive elements "indicate either through a link to the context or within the utterance itself that the set to which the focused constituent belongs indeed contains alternative members that are relevant for the discourse at hand" (Titov 2013:422). So *in Italy* in (36) is not contrastive because there it is not indicated

[15] However, given LFG's modular architecture, contrastiveness is not a syntactic feature, as it is in some cartographic approaches (e.g. Molnár & Winkler 2010). See Neeleman et al. (2009) for a detailed critique for the cartographic approach.

that other countries are particularly relevant for the question under discussion. In (38), they are, so *in Italy* in (38) is contrastive.

(38) A: *Did you go to Spain in the summer?*
B: *No, I was IN ITALY.*

It is sometimes claimed (e.g. É. Kiss 1998) that the open or closed nature of the referent set is a crucial factor. However, Repp (2009) calls attention to Krifka's (2008) observation that the answer in (39a) feels no more contrastive than the one in (39b), despite the fact that in (39a) an explicitly restricted set of alternatives is provided.

(39) a A: *What do you want to drink, tea or coffee?*
B: *Tea.*

b A: *What do you want to drink?*
B: *Tea.*

So when exactly these contrastive interpretations are licensed is highly dependent on the context and the communicative purpose of the speaker/hearer. That is, it is not enough that potential alternatives are present in the discourse, they have to be salient from the perspective of the interlocutors. Certain contexts may of course facilitate or hinder the obtaining of the contrastive interpretation but it is never going to be obligatory. For researching the conditions, a framework like Asher & Lascarides (2003) could prove to be a fruitful research avenue.

For present purposes, the relevant part of the discussion is that contextually salient alternatives become active in the consciousness of the interlocutors. As we will see later, this may also be represented via Büring (2003)-style discourse-trees. I propose that the discourse-structuring feature is –CONTRASTIVE by default, but it may receive a positive specification from an additional +CONTRASTIVE feature,[16] superimposed on the base concept: it manages the questions under discussion, by activating alternative nodes in the discourse-tree. Note that the +/–CONTRASTIVE is strictly subordinated to the DISCOURSE-STRUCTURING feature so "contrast" is not applicable in the realm of –DISCOURSE-STRUCTURING information-structural categories (these are –CONTRASTIVE by definition)[17].

[16] I will argue in 2.6.3 that questioning is another type of additional discourse-structuring feature.

[17] This ia comparable to Molnár & Winkler's (2010) C(oherence)-feature, which if present (positively specified, in other words) may be further specified as C-continuity and C-Contrast. However, there are two important differences. One, the C-feature is just described as being "responsible for the formally marked type of discourse linking, i.e. for cohesion in discourse" (Molnár & Winkler 2010:1396), while I strive for a characterization in terms of discourse-trees, which I believe to be a) more precise b) able to capture the distiction between Topics and Background information (I do not see how Molnár & Winkler would distinguish these categories). Also, for Molnár & Winkler such discourse-features are part of syntax, while my approach is more in

(40) +/–CONTRASTIVE: a linguistic entity with a +CONTRASTIVE feature at information-structure participates in discourse-structuring by evoking a contextually salient pragmatic set of alternatives.
A linguistic entity with a –CONTRASTIVE feature at information-structure does not evoke a contextually salient pragmatic set of alternatives.

Now that the features of my proposed information-structural taxonomy have been adequately described, I can now present the IS-categories themselves.

2.6.2 Categories of information-structure

In the following sections, I survey the i-structural categories that I posit. I repeat Table 7 here for convenience.

		+NEW	**–NEW**
+D-STRUCTURING	**+CONTRASTIVE**	Contrastive focus	Contrastive topic
	–CONTRASTIVE	Information focus	Thematic shifter
–D-STRUCTURING		Completive information	Background information

Table 7.
The proposed IS-taxonomy.

I would like to highlight at this point that the taxonomy is aimed to characterize the information-structural properties of these notions. As I will repeatedly point out, there are some properties of these categories may only be adequately described with reference to other representational levels (semantics is a prime candidate for this). So, while I will describe several of these aspects in the following sections, I do not aspire to derive each of them from the information-structural taxonomy itself. Care will be taken to delimit the scope of the system.

2.6.2.1 Topic, Contrastive topic and Background information

Topic, Contrastive topic (CT) and Background information (BI) are all –NEW , so they are part of the pragmatic presupposition of the sentence. The first two participate in discourse-

line with Horváth (2010) and Neeleman & van de Koot (2008), who argue that such notions are best captured as interface phenomena.

structuring, while Background information does not. Let us investigate the properties of these notions.

The common wisdom about the topic is that "the topic of a sentence is the thing the proposition expressed by the sentence is about" (Lambrecht 1994:118).[18] Unfortunately, it is hard to give this "aboutness"-relation substantial content, apart from intuitions. Consider (41), a sentence repeated from section 2.3.

(41) *Gavrilo Princip assassinated the emperor.*

Under default assumptions, the topic of the sentence is *Gavrilo Princip*. However, one can argue that the sentence is not just about *Gavrilo Princip*, but also about *the emperor*, or about history in general. There are some heuristics that have been proposed as empirical tests for this kind of aboutness-relation. Prince (1999), citing Gundel (1974) and Reinhart (1981), lists 3 such tests.

- **The "as for X"-test**: Can the sentence be plausibly paraphrased with an initial "as for X"-phrase, where X is the supposed topic expression?

- **The "what about X"-test**: Can the sentence plausibly answer a "what about X"-question, where X is the supposed topic expression?

- **The "say about X that…"-test**: Could the sentence be plausibly reported about by using an initial "Y said about X that…"-phrase, where X is the supposed topic expression?

If we use these tests on (41), we may get some evidence that the sentence is about Gavrilo Princip. Nevertheless, this is weak evidence as it would be difficult to argue that (42a'-c') are absolutely infelicitous paraphrases for (41).

(42) a *As for Gavrilo Princip, he assassinated the emperor.*

a′ *?#As for the emperor, Gavrilo Princip assassinated him.*

b A: *What about Gavrilo Princip?*
B: *Gavrilo Princip assassinated the emperor.*

[18] See also Reinhart (1981). Note that this definition does not include "discourse topics", as its scope is limited to the sentence-level.

b′ A: *?#What about the emperor?*
B: *Gavrilo Princip assassinated the emperor.*

c *He said about Gavrilo Princip that Gavrilo Princip assassinated the emperor.*

c′ *?#He said about the emperor that Gavrilo Princip assassinated the emperor.*

Another reflex of this aboutness-relation is the often mentioned semantic requirement that topics be referential, as only referential entities can be targets for predication. Thus, non-referential entities, e.g. quantifiers cannot be in the topic-position in Hungarian (see more about the structure of Hungarian in section 3.1), as in (43) (from É. Kiss 2002:10).

(43) **[Kevés várat]$_{TOPIC}$ meg-védtek a zsoldosok a törökök ellen.*
few forts.ACC PV-defended.3PL the mercenaries the Turks against
'Few forts were defended against the Turks by the mercenaries.'

A complicating factor is that in the case of CTs, the referentiality constraint seems to be relaxed, as certain nonreferential elements may also be contrastive topics, as in (44)[19], where the CT-elements would be ungrammatical as neutral topics (44b-c are from Gécseg 2001).

(44) a *[Kevés várat]$_{CT}$ a zsoldosok védtek meg, sokat pedig a hősök.*
few castles the mercenaries defended.3PL PV many conversely the heroes
'Few forts were defended by the mercenaries, many were defended by the heroes.'

b *[Úszni]$_{CT}$ nem tudok.*
swim not can
'Swim, I cannot.'

c *Szépnek$_{CT}$ szép a húgod.*
pretty.DAT pretty the sister.POSS.2SG
'Pretty, your sister in fact is.' (but she may not be clever)

At any rate, the concept of aboutness clearly sets Background information apart from Topics, as the former is definitely not what the sentence is about. It only provides the contextual backbone of the utterance, without being involved in the informational dynamics of the conversation.

Apart from these semantic requirements, some pragmatic constraints have also been proposed for topics. These are related to the referential newness/givenness dichotomy, as described in section 2.6.1.1.

[19] Given that semantic anomalies do not interfere, see Gyuris (2009).

It has been suggested that the entity that is denoted by the topic expression should be accessible in the discourse universe. Gundel (1985) calls this the "familiarity condition" on topics. Consider (45), from Lambrecht (1994:159), which could be the beginning of a telephone conversation, where someone had dialed the wrong number.

(45) A: *Is Alice there?*
B1: *#Alice isn't here.*
B2: *There is no Alice here.*

Even though *Alice* is a referential, definite expression, and is clearly discourse-old by the time of the answer, reply B1 in (45) is undoubtedly strange. The problem in B1 is that *Alice* is not properly established in the universe of the discourse, since the one who replies doesn't know what *Alice* the questioner could refer to. The way to circumvent this problem is to remove *Alice* from the position where she is interpreted as a topic, as in B2.

Another example for such issues is that the referential newness/givenness status of the topic entity may influence the choice of the form of the topic expression. Newly introduced topics ("thematic shifters", "shifting topics") are often distinguished from "continuing topics" (Frascarelli 2007, Gazdik 2011). Consider (46), which is based on Gazdik (2011:168)

(46) *Tamás szeret olvasni. (#Ő) intelligens, szorgalmas és sokra fogja*
Thomas likes read.INF he intelligent hard-working and much AUX
vinnni.
reach.INF
'John likes reading. He is intelligent, hard-working and he will achieve a lot.'

In (46) the pronoun is pragmatically anomalous. This shows that pronominal continuing topics are preferably dropped in Hungarian.

The semantic and discourse-pragmatic aspects of topics are certainly important in a full characterization of IS-notions, but my primary concern here is information-structure, so I will now turn to those properties of Topic, Contrastive topic and Background information that are relevant from the perspective of the features that I have introduced in section 2.6.1.

Topic as an entity of IS has the feature composition –NEW +D-STRUCTURING. In the case of Contrastive topics, it is also +CONTRASTIVE. Background information is –D-STRUCTURING (and thus –CONTRASTIVE by definition).

Being –NEW, a Topic is part of the pragmatic presupposition. As such, it is outside the scope of sentence negation. Lambrecht (1994:52) uses the following example:

(47) a *John is my friend.*

b *My friend is John.*

(47a) is assumed to be about *John*, so *John* is the topic. Imagine someone challenges the claim in (47a) by saying "That's not true!" This would be understood as claiming "John is NOT your friend," but the existence of *John* would still be taken for granted. Since it is presupposed, it is outside of the scope of sentential negation. In fact, the denial could be felicitously complemented with the presupposition-cancelling utterance "you don't have any friends," which indicates that only the existence of the topic (*John*) is presupposed, the content of the comment is not. Conversely, uttering "That's not true!" in response to (47b) where the topic is *my friend*, would still presuppose that I have a friend (just not *John*).[20] This is a property that the category Topic shares with Background information.

As for the DISCOURSE-STRUCTURING feature, from my perspective, its conceptualization should boil down to the capacity of topics for structuring of the conversation, by being directly involved in the formation and resolution of questions under discussion of the discourse. Consider the conversation in (48).

(48) A: *What car did Jack buy?*

B: *[Jack]*$_{TOPIC}$ *[bought]*$_{BACKGROUND\ INFORMATION}$ *[a Volvo]*$_{FOCUS}$.

I will consider the discourse-structuring of the focus (and other +NEW entities) in the next section. It seems straightforward that Background information does not participate in the structuring of the conversation. It is referentially old and is part of the presupposition but it is completely unaffected by the informational dynamics of the conversation: it does not provide new questions or answers for the discourse, nor does it direct the attentions of the participants.[21]

It is less straightforward exactly how Topics are discourse-structuring. It is intuitively clear and commonly accepted that topics serve as "links" to the previous discourse, they are related to the discourse topic. In Gazdik's (2011) framework their role introduce new subtopics. It is also intuitively clear that they play a central role in the information-structure of the sentence by virtue of being the targets of predication in the sentence.

How could this intuition be adapted to a discourse-tree-based conceptualization? One must realize that (48) must be part of a wider discourse. Somehow the interlocutors must have

[20] Note that the test even works if we replace *John* with a definite expression like *the king of France*. Of course one could say something like "That's not true, because the king of France doesn't even exist!", but that is an explicit modification of the presupposition.

[21] In the words of Mycock (2013:426), it does not "establish new pragmatic relations".

reached this point of the conversation from the Big Question ("What is the way things are?"). For a sentence like (48) to be uttered, speaker B has to assume that A wants to know about Jack. That is, at least implicitly, the discourse tree for B must contain a node with an instruction "Tell A about Jack!". As such, *Jack* in the answer directly corresponds to an instruction in the discourse-tree, as illustrated in Figure 8. If this is along the right track, it can be justified how topics (but not Background information) do participate in the formation and the resolution of the questions under discussion of the conversation.

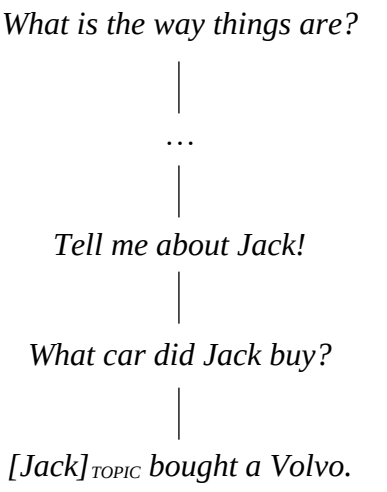

Figure 8.
Topic in a discourse tree.

This may well be connected to at least two dimensions of topics, mentioned earlier. One is the easy to see conceptual connection to the aboutness-relation of Reinhart (1981) and Lambrecht (1994). The other one is the +PROMINENT feature specifications of topics in earlier taxonomies of IS. Under the assumption that topics are implicitly but directly introduced by a process illustrated in Figure 8, it is easy to see why they are felt to be prominent. Although my approach also needs more explicitness and formalization in future research, I believe it to be a step in the right direction.

In the case of Contrastive topics, a contextually salient pragmatic set of alternatives is evoked. This could be represented, following Büring (2003), by positing that the original question is split into subquestions. This extends the discourse-tree horizontally and clearly is a discourse-structuring move.[22]

[22] Because of this, Titov (2013) argues that CTs are –PRESUPPOSED, as they introduce new material. In her framework they are eventually equated with Contrastive foci, their distinction only being the result of the different configurations they occur in. I do not agree with this position, as it may be true that CTs introduce referentially new material, but it is not necessarily the case. Consider (i):

i) A: *Did Jack buy a Volvo or an Audi, or both?* B: *A [Volvo]$_{CT}$, he did buy, I don't know about an*

CTs are always associated with a focussed element[23], which provides an answer for only one of the subquestions. Thus a sentence with a CT is felt to be an incomplete or partial answer (though later discourse may specify the other nodes too). See (49) and Figure 9.

(49) A: *What cars did Jack buy?*
B: *[A Volvo]$_{CT}$ [Jack]$_{TOPIC}$ [did]$_{FOCUS}$ [buy]$_{BACKGROUND\ INFORMATION.}$ (But I don't know about other cars.)*

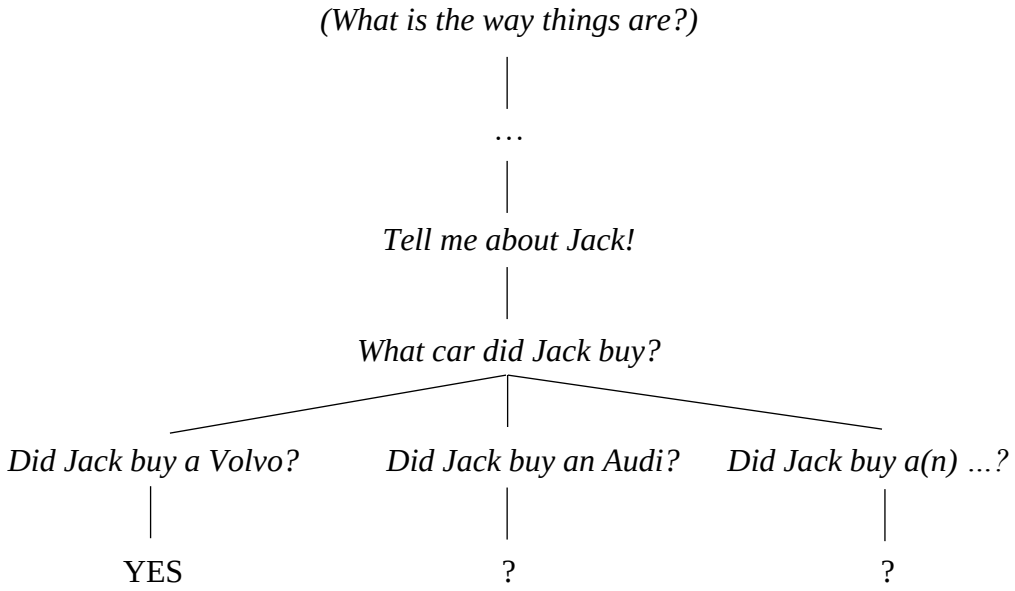

Figure 9.
Contrastive topic in a discourse-tree.

Audi.

Also, even though referentially new material may be introduced via CT-formation, the CT (simply because it is a Topic itself) is still part of the presupposition, thus they are –NEW (or +PRESUPPOSED, in reverse terminology).

Furthermore, equating CTs and CFs would make it hard to understand why certain languages have a specific morphological marker for neutral Topics and Contrastive Topics (*wa* in the Japanese example in (i) below), with the exclusion of CFs (Neeleman et al. 2009).

(i) a A: *Tell me about that dog.*
B: *Sono [inu-wa]$_{TOPIC}$ kinoo John-o kande-simatta.*
that dog-wa yesterday John-ACC bite-closed
'The dog bit John yesterday.'
b A: *What did John eat at the party yesterday?*
B: Hmm, John-wa doo-ka sira-nai-kedo,
('Well, I don't know about John, but…)
[Bill-wa]$_{CT}$ 8-zi-goro mame-o tabeteita (yo).
Bill-wa 8 o'clock-around beans-ACC eating (PRT)
'As for Bill, he was eating beans around 8 o'clock.'

This suggests that CTs form a natural class with Topics and not with CFs, so I subscribe to the standard view that upholds the distinction between Contrastive foci and Contrastive topics.

[23] I would like to remain neutral on the exact nature of the focussed element. It could be a noncontrastive or a contrastive focus or verum focus, as in the case of (49). The characterization of verum focus is less than straightforward though. For Titov (2013) it also comes in two versions (contrastive/noncontrastive), and there are proposals according to whom verum should be entirely detached from focus (Gutzmann, Hartmann & Matthewson 2017).

2.6.2.2 Information focus, Contrastive focus and Completive information

Let us now turn to the +NEW i-structural categories, Information focus (IF), Contrastive focus (CF) and Completive information (CI). Having this feature, they are all part of the pragmatic assertion of the sentence. The difference is that while the two foci are +DISCOURSE STRUCTURING, thus directly involved in the formation/resolution of questions under discussion, Completive information is not.

The interpretation of focus is traditionally tied to answers to questions (see e.g. Krifka 2008). It has also become a general consensus especially since É. Kiss (1998) that at least two kinds of foci should be distinguished. She makes the distinction between Identificational focus and Information focus.[24] A classic example is (50), note that the translations are from É. Kiss (1998).

(50) A: *Hol jártál a nyáron?*
where went.2SG the summer.on
'Where did you go in the summer?'
B1: *[Olaszországban]*$_{\text{IDENTIFICATIONAL FOCUS}}$ *jártam.*
Italy.in went.1SG
'It was Italy where I went.'
B2: *Jártam [Olaszországban]*$_{\text{INFORMATION FOCUS}}$
'I went to Italy (among other places).

According to É. Kiss (1998), while the Identificational focus in (50/B1) exhaustively[25] identifies the places where the speaker had gone, the Information focus in (50/B2) merely mentions one such place.

While the recognition of this distinction is very influential, there are certain aspects of É. Kiss's (1998) approach that have come under criticism. Kenesei (2006) criticizes É. Kiss (1998) on the grounds that both types of foci are "identificational" in the sense that they identify a member from a set of countries that B may have visited. The difference, according to Kenesei, is that while "Information focus" operates on a subset relation, "Identificational focus" involves a *proper* subset relation, that is at least one more member of the set is involved in the semantic computation. Thus Kenesei (2006) suggests that the label "Contrastive focus" should replace Identificational focus.

[24] Already at this point, I would like to emphasize that the way that I use the term "Information focus" does not correspond to É. Kiss's terminology.

[25] Also, it must be pointed out that the "exhaustive" nature of Identificational focus is not an inherent property of the IS-category itself. Rather, it seems to be a property of certain constructions, in which foci is manifested in various languages, the Hungarian preverbal "focus-position" being the prime candidate for it. For É. Kiss (1998) this focus is +EXHAUSTIVE and +CONTRASTIVE, there is an ongoing debate about this issue, see e.g. Onea (2009), Wedgewood (2009), Gerőcs, Babarczy & Surányi (2014), Pintér (2016).

Returning to the answers in (50), one must observe that while in (50/B1), *Olaszországban* ('in Italy') strictly corresponds to the wh-phrase in (50/A) (so it is +D-STRUCTURING), an answer like the one in (50/B2), may also answer a question like "What did you do in the summer?", where the answer is the entire VP, not the preverbal constituent (Gazdik 2011:209). *Olaszországban* is still NEW information, part of the pragmatic assertion, but it is not directly linked to the question under discussion. In such a scenario, it should be regarded as –DISCOURSE-STRUCTURING, thus having the status of Completive information. Another example for this IS-notion is (23) earlier, repeated here for convenience.

(51) A: *Where is Nadya coming from?*
B: *She was just buying [toffee]$_{CI}$ at the market.*

As for the label "Information focus", I would like to reserve it for cases of new information that directly answer questions under discussion, so in the context of (50/A), both instances of *Olaszországban* count as IF. (That is, "identificational focus" in Kenesei's (2006) sense.[26]). Let me repeat example (48) and its discourse-tree from the previous section, in light of this discussion.

(52) A: *What car did Jack buy?*
B: *Jack bought [a Volvo]$_{IF}$.*

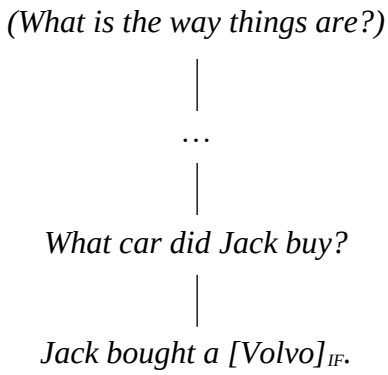

Figure 9.
Information focus in a discourse tree.

Additionally, the focus may also be +CONTRASTIVE, in the sense defined in 2.6.1.3. Similarly to Contrastive topics, I assume that in this case, new subquestions are added to the discourse-tree. The difference is that in the case of CF, at least one of the new subquestions are answered

[26] Titov (2013) uses the term New information focus (NIF), with the same justification.

in an opposite manner compared to the original question.[27] See (53) and its discourse-tree in Figure 10.

(53) A: *What care did Jack buy? An Audi?*
B: *No, [A Volvo]$_{CF}$ [Jack]$_{TOPIC}$ [bought]$_{BACKGROUND\ INFORMATION}$ (and not an Audi).*

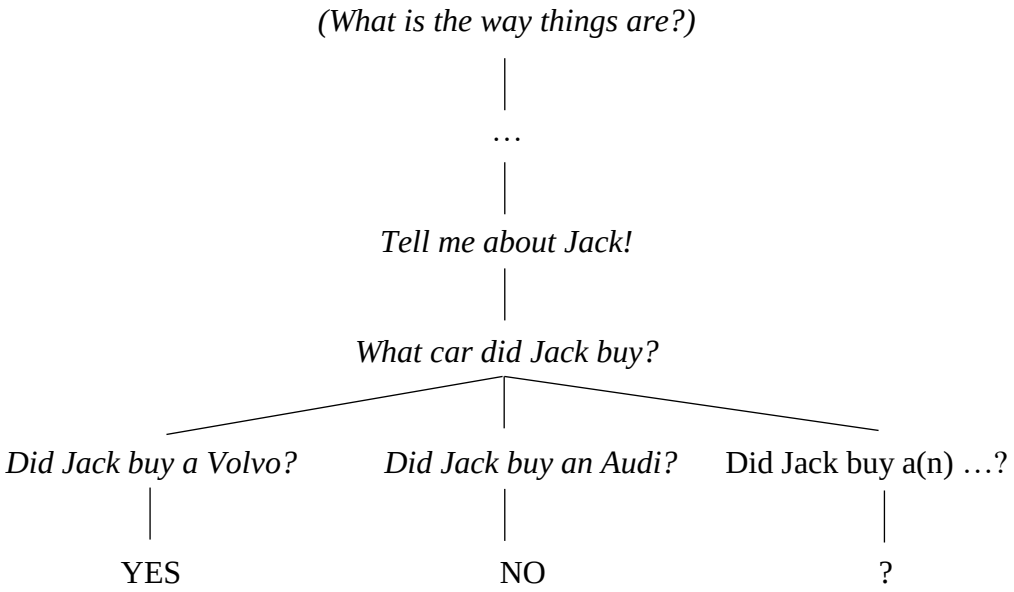

Figure 10.
Contrastive focus in a discourse-tree.

2.6.3 Miscellaneous issues regarding i-structure 1: question-words

Before concluding the presentation of my view of information-structure, I would like to tie some loose ends, in order to clarify the relation of the present taxonomy to some issues that are prominently present in the literature about IS.

One such issue is the issue of the i-structural status of question words. The most recent LFG-account is that of Mycock (2013). She takes the taxonomy of Butt & King (1996) and enriches it with the feature Q, yielding the following outcome (Table 9).

[27] How this opposing truth value is added to the semantic computation is an important question for further research. In Titov's (2013) system it is a consequence of an additional contrastive verum focus.

	+NEW	–NEW
+PROM	Focus	Topic
	Q: Questioning Focus	**Q: Sorting Key**
–PROM	Completive Information	Background Information
	Q: Non-sorting Key	**Q: Echo-Question**

Table 9.
The taxonomy of question words in Mycock (2013).

I readily agree that utilizing an additional feature like Q is a good move. It actually seamlessly integrates into my taxonomy: the Q-feature clearly belongs to the discourse-structuring domain, as nothing is more directly involved in the formation of questions under discussion than questions themselves. What I propose is that Q is an additional discourse-structuring feature like contrastiveness: it further specifies the discourse-structuring relation. While +CONTRASTIVE extends the discourse tree horizontally, a question may be said to extend it vertically, by invoking an answer. Since Q and CONTRASTIVE are at the same level of specification, I also propose that they are in complementary distribution. In other words, a discourse-structuring relation may be +CONTRASTIVE or +Q, but not both. The result is this:

		+NEW	–NEW
+D-STRUCTURING	+Q	Questioning focus	Sorting key
	+CONTRASTIVE	Contrastive focus	Contrastive topic
	(–CONTRASTIVE) (–Q)	Information focus	Topic
–D-STRUCTURING		Completive information	Background information

Table 10.
Question words in IS taxonomy

There are a number of crucial differences between Mycock's (2013) classification and the one that I propose. The most obvious one is the number of question-types, 4 vs. 2. This is essentially because I think that Non-sorting keys and Echo-questions (at least from the perspective of information-structure) are just different subtypes of Questioning Focus.

First let us clarify what the various question-types are. Questioning focus is the simplest case, it is a regular clause-initial wh-word, asking for a piece of information. Such information is provided by Information focus, so their correspondence is straightforward.

Sorting key and Non-sorting key are notions that are relevant from the perspective of questions containing multiple question words.[28] Consider (54):

(54) A: *[Ki]*$_{\textit{SORTING KEY}}$ *[kit]*$_{\textit{NON-SORTING KEY}}$ *kedvel?*
who whom likes
'Who likes whom?'
B: *János Katit szereti, Tamás Annát szereti…*
John Kate.ACC likes Thomas Anne.ACC likes
John likes Kate, Thomas likes Anne…

When an answer is provided for a question like (54/A), it is the first question word is the one along which the answer is organized. The answer is a list of pairs, and the pairs are arranged/sorted according to the people. Hence the term "Sorting key", from Kuno & Takami (1993). Furthermore, É. Kiss (2002:101) observes that the sorting key is not even a "real" question word: it is interpreted more like a distributive quantifier: "tell me about each person whom they like". Also, according to Surányi (2006), in Hungarian "high wh-expressions [the sentence initial one in [54/A, SZ.P.] invariably quantify over presupposed sets". That is, they are –NEW, in contrast to the "real" question word of the question (which occupies the immediately preverbal position). This makes such question words natural parallels for Topics. In turn, the other question word (the one I have been referring to as the "real one") is the Non-sorting key.

As can be seen from Table 10, I propose that Non-sorting key as an independent category should be dispensed with, since its role is exactly what the role of Questioning focus is, asking for information. Mycock (2013:430) claims that the –PROM Non-sorting key is distinct from the +PROM Questioning focus as "it represents information that is relatively less important than other information – most notably the Sorting Key – in the given context". Apart from the problems with the PROM feature mentioned in 2.6.1, it is not at all clear in what sense is the Non-sorting key less important. If anything, it is more important, as it represents an informational gap in the mind of the interlocutor, while the Sorting key being part of the presupposition, only provides the schemata for the answer.

The other category that I do not include in my taxonomy is Echo question. These occur in sentences where the listener for some reason (e.g. noise, lack of attention) did not understand an element and immediately asks back:[29]

[28] For more on this topic see Mycock (2006) and Gazdik (2011).

[29] Note that this is distinct from the surprise/shock-type Echo questions, as in (i). These are distinct, since they do not represent an information gap, as in (i). I leave their analysis for further research.

(i) A: *I've bought a Trabant.*

(55) A: *I've bought a Volvo.*
B: *You've bought a [what]*$_{ECHO\ QUESTION}$*?*

For Mycock (2013), these are –NEW, –PROM. I challenge both specifications and argue that Echo questions should also be analyzed as Questioning foci. Mycock (2013:426) argues that the relational status for an echo question is not at issue, since the element under discussion has been mentioned in the preceding utterance, hence the –NEW feature. However, as has been established in 2.6.1.1, the issue of being mentioned in the discourse is a matter of referential and not relational newness/givenness.

With regards to prominence, the negative specification is justified by Mycock (2013) on the grounds that Echo questions should be distinct from Sorting keys. Also, echo questions are of "secondary nature". I have already criticized the +PROM feature of Sorting keys, so the first aspect is not particularly important here. As for the second one we once again encounter the problematic issue of defining prominence. While from a referential point of view, Echo questions are "secondary", this does not necessarily mean that they are not "prominent", whatever that means. The answer for such a question is clearly relevant for the asker, otherwise they would not ask the question. Moreover, echo questions exhibit a high degree of prosodic prominence.

Consequently, while the parallelism between Information focus and Questioning focus on the one hand, and Topic and Sorting key on the other hand is a natural one, the connection between Background information and Echo questions is unmotivated and stipulative. Based on the above considerations, Echo question could be more naturally related to Questioning foci. However, to equate them would be premature at this point, as there are important differences as well.[30]

B: *You've bought a WHAT?!*

[30] An intriguing aspect of Mycock (2013) is her Principle of Relative Prominence Encoding (PRPE): A −PROM question word will only be syntactically "highlighted" in a language (i.e. appear ex situ, as the filler element in a long-distance dependency) if its +PROM question word counterpart is also by default syntactically highlighted.

There seems to be empirical merit behind this generalization, but my approach lacks the PROM feature and lacks two of her categories, so some work has to be done for a fuller picture about the relation of my approach and hers.

There are languages where in a question with a Sorting key and a Questioning focus, only the former is fronted, but the reverse situation (in situ Sorting key, ex situ Questioning focus) is never the case. This follows from Mycock's (2013) taxonomy (Table 9) and PRPE. However, the distinction is also statable in terms of NEW: in these languages, the –NEW question word should be ex situ. (If the assertion that a negative specification licenses syntactic highlighting seems unintuitive note that –NEW could be also labelled +GIVEN.)

More problematic for my proposal is the other generalization: there are languages where Questioning foci are fronted but Echo questions are not, but there is no example for the opposite situation (in situ Questioning focus, fronted Echo questions). So some distinction must be made, but the exact nature of this is not clear to me at this point.

2.6.4 Miscellaneous issues regarding i-structure 2: hocus

The Hocus i-structural category was proposed in Kálmán (1985) and is explicitly present in Gazdik's (2011) taxonomy. According to Gazdik (2011:191), a Hocus is an immediately preverbal element in Hungarian "expressing some participant or circumstance in the event denoted by the predicate. Such noun phrases can bear main stress and appear in the immediately preverbal position, when the event denoted by the verb is not particularly newsworthy". Prosodically, a Hocus is much less prominent than a Focus. Example (25) is repeated here.

(56) *János tegnap [vonattal]$_{HOCUS}$ utazott haza.*
John yesterday train.with travelled.3SG home
'John yesterday took the train to go home.'

Gazdik (2011:192) notes that in (56) travelling home counts as a usual event, but implies that John usually does not take the train, so the Hocus-element is the newsworthy piece of information. From this, it may seem that the Hocus is always contrastive. However, Gécseg (2013:34) claims that the Hocus of a sentence, in contrast with the Focus, is not contrastive, at least not explicitly. Gécseg (2013) supports this with the observation that the role of Hocus is often the introduction of a new discourse-referent, like in (57), which could be part of a narration and where the contrastive interpretation is completely absent.

(57) [*Egy ősz hajú férfi]$_{HOCUS}$ lépett be a szobába.*
one grey haired man stepped.3SG in the room.in
'A grey haired man stepped into the room.'

A third type of Hocus is where the element in question is not a new piece of information at all (either relationally or referentially), but acts rather like a topic (Gécseg 2013).

(58) A: *Ki az a Szentgyörgyi Albert?*
who that the Szentgyörgyi Albert
"Who is Szentgyörgyi Albert?"
B: *[Ő]$_{HOCUS}$ fedezte fel a C-vitamint.*
he discovered.3SG up the vitamin.C
'He (is the one who) discovered vitamin C.'

For languages where wh-phrases appear either uniformly in situ or ex situ, a feature-based explanation is basically superfluous: here the picture boils down to elements with a +Q feature being never or always fronted.

As these examples show, the Hocus is a heterogeneous category. What I suggest is that these various uses may be integrated into the IS-taxonomy as cases of existing categories, rather than having Hocus as an independent concept. This would be a desirable outcome, since the category Hocus is discussed exclusively with regard to Hungarian and a language-specific IS-notion is not particularly desirable in any taxonomy.

The proposal is this. If (56) really implies that John usually does not take the train, then it could be a case of Contrastive focus, perhaps with a bit weaker sense of contrast. As for (57), I consider it to be a case of Completive information. The preverbal element is some additional piece of information, not directly requested by a question under discussion, which is simply "What happened?" and not "Tell me about a grey haired man!". Finally (58), in accordance with Gécseg (2013) should be regarded as a Topic. So Hocus should be thought of as an umbrella-term for the manifestations of various other i-structural categories.

2.7 Conclusion of chapter 2

In this chapter I gave a general outline of the theoretical framework of the dissertation, Lexical-Functional Grammar. LFG is a lexicalist, constraint-based theory with a parallel architecture. An overview of the main levels of representational levels was provided (constituent-, functional-, argument- and information-structure). The most attention was given to information-structure as the currently available frameworks are argued to be deficient in various respects. I have proposed a new taxonomy for information-structure, which builds on the insights of previous approaches but is more advanced in comparison to them because i) it includes all but the only IS-categories that are well established ii) it utilizes the i-structural features in a principled and theoretically sound way.

CHAPTER 3

RELEVANT ASPECTS OF HUNGARIAN SYNTAX

In this section, I lay the theoretical and empirical groundwork for the analysis of the Hungarian fronting constructions to be discussed in Chapter 5. First, I will outline the basic structure of Hungarian simple sentences, and then I will offer some insights into subordinate clauses as well. Following the general framework of this dissertation, I will take the perspective of Lexical-Functional Grammar.

3.1 The simple clause in Hungarian

Hungarian is standardly assumed to be a discourse-configurational language, thus, its phrase structure is primarily determined by discourse functions like topic and focus (É. Kiss 1995, chapter 1). It is generally recognized that a Hungarian sentence shows a duality in terms of its phrase structure. While the preverbal field is hierarchically structured, the primary organizing forces being scope and discourse-functions, the postverbal area is usually assumed to be flat.[31]

There have been a number of proposals that have been put forward about the LFG-theoretic analysis of the Hungarian phrase structure: Payne & Chisarik (2000), Mycock (2006), Gazdik (2011) and Laczkó (2014a, 2014b, 2015). In this dissertation I chose Laczkó's works as a baseline, as these constitute the most articulated framework and they have also been designed to be compatible with LFG's computational platform, XLE. This approach is heavily inspired by É. Kiss's (1992) GB account. As Laczkó's works mostly concentrate on syntax, I will supplement them with my proposals regarding information-structure.

Figure 1 shows the basic structure of Hungarian, as assumed in this dissertation. (I will simplify the diagram, leaving out those annotations which are not directly relevant for the purposes of the dissertation. The interested reader is referred to the cited works of Laczkó). For ease of interpreting the information-structural notations, let me also repeat the proposed taxonomy from section 2.6.

[31] While this picture might still be called the standard view, there have been proposals that argue for a hierarchical postverbal field, see Surányi (2006b).

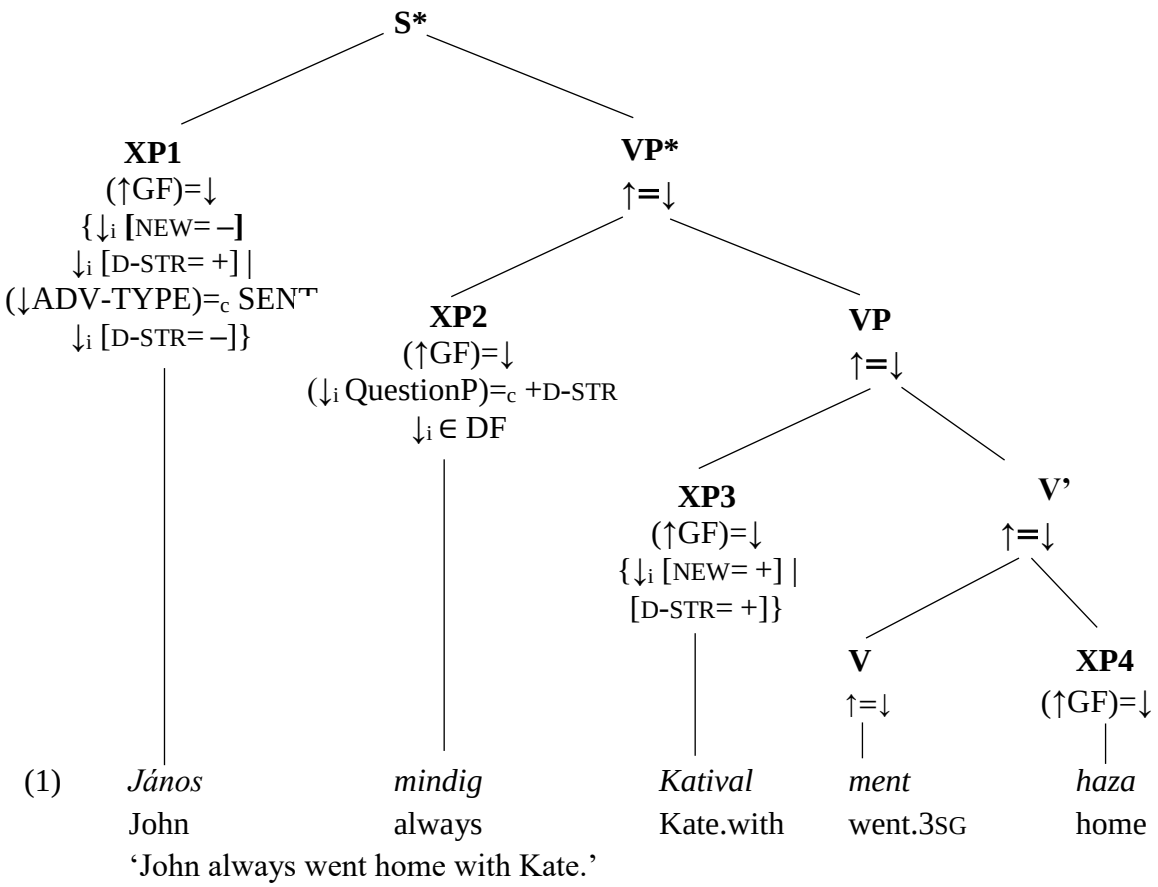

Figure 1.
Basic Hungarian sentence structure.

		+NEW	–NEW
+D-STRUCTURING	+CONTRASTIVE	Contrastive focus	Contrastive topic
	(–CONTRASTIVE)	Information focus	Topic
–D-STRUCTURING		Completive information	Background information

Table 1.
The proposed IS-taxonomy.

The sentence is headed by the exocentric S node. (The IP label is also possible if one assumes this category for Hungarian. However, see Laczkó (2014a:333-336) for an argumentation on the side of S.) This node dominates what is traditionally called the "topic-field" of the Hungarian sentence.[32] This is labelled as XP1 in Figure 1. The arrows in the annotations are

[32] The semantic properties of the topic field are not in the focus of this thesis, so they are not explicitly present here. It is generally assumed that topical entities are referential, although contrastiveness complicates the issue, as noted in connection with examples (43-45), Chapter 2. É. Kiss (2002:10) claims that topics in Hungarian must also be specific, so the indefinite *egy autó* ('a car') in (ia) must refer to a member of a previously introduced set. However, Gécseg & Kiefer (2009) point out that in the proper context, subjects similar to the ones in (ia) may

"metavariables". ↑ refers to the mother-node's functional structure, while ↓ refers to the node's own f-structure. So the (↑GF)=↓ should be read as "my grammatical function is the same as my mother node's grammatical function" and this notation indicates that any grammatical function (GF) may occupy this position, including adjuncts. Arguments are either Topics or Contrastive topics at information-structure (the feature CONTRASTIVE is not specified). Most adjuncts (ones with referential interpretation, e.g. temporal, locative, instrumental) behave in a parallel fashion. Sentence-adverbs (*luckily*, *hopefully*, etc.), however, seem to reject DISCOURSE-STRUCTURING IS-categories.[33] This may be seen from (2a), where it is impossible to interpret *valószínűleg* ('probably') as a Contrastive topic. Since restrictions are looser on CTs than they are on neutral topics (see example 45, Chapter 2), we may assume that *valószínűleg* is not a neutral topic either (it would also be incompatible with the aboutness-relation associated with topics). *Valószínűleg* may neither be a Focus (2b).

(2) a *#János valószínűleg Kateval ment haza, valószínűtlenül pedig Annával*
John probably Kate.with went.3SG home improbably conversely Anne.with
'Probably, John went home with Kate and improbably, he went home with Ann.'

b *#János valószínűleg ment haza (és nem biztosan)*
'It was probably that John went home (not surely).'

However, such adjuncts may still provide some new information (about the mental disposition/judgment of the speaker in (3) or be part of the background, so I assume that the –D-STRUCTURING categories are available for them.

(3) A: *Hol van (valószínűleg) János?*
where is probably John
'Where is John (probably)?'

introduce brand-new referents, as in (ib). This is possible because we can easily accommodate the referents by contextual inference.

(i) a *Egy autó megállt a házunk előtt.*
one car stopped.3SG the house.POSS.1PL in.front.of
'A car stopped in front of our house.'

b *Képzeld, mi történt tegnap.* (Guess what happened yesterday.)
Egy gyerek leesett a villamosról, de szerencsére nem sérült meg.
one child fell.off.3SG the tram.from but luckily not injured.3SG PV
'A child fell off the tram but luckily, he wasn't injured.'

[33] I suspect this is because their meaning is not compatible with discourse structuring moves ("Tell me about X!", questioning, question-splitting).

B: *János* *[valószínűleg]*$_{CI/BI}$ *az* *egyetemen* *van.*
John probably the university.on is
'John is probably at the university.'

This possibility is restricted to sentential adverbs with the help of a "constraining equation". As opposed to regular (defining) equations in LFG annotations, which simply specify a value (making it exist), constraining equations check whether the specifications hold, without making them exist themselves. So (↓ADV-TYPE)=$_c$ SENT is satisfied if the adverb occupying that position is specified as sentential as a lexical property.

The link between canonical topics and sentence adverbs that makes this constellation possible could be their capacity to restrict the domain for which the predicate holds. Both could be conceptualized as what Chafe (1976) (quoted in Maienborn 2001:229) calls "Chinese-style topics": they limit the applicability of the main predication to a certain restricted domain. […] Typically, it would seem, the topic sets a spatial, temporal, or individual framework within which the main predication holds." While standard "aboutness"-topics restrict the application of the comment to a certain entity, a sentential adverb restricts it in a way that is related to the cognition or the assessment of the speaker (e.g. a probability-judgment in 3). In the long run it may turn out that the disjunction in (1) is unnecessary and the position in question may simply be regarded as a topic-position. To do this, more work on the semantic aspects of information-structure should be done. Until that is carried out, I maintain this conservative position. Since a sentence can contain several such entities and they freely intermingle[34], this node is iterative.

The next position in the Hungarian sentence is the quantifier-field, labelled as XP2 in Figure 1. This is also iterative and the order of elements affects the meaning of the sentence, linear order being tied to scope.[35] This field is unspecified with respect to information-structure, as can be seen from the following examples.

(4) a A: *Hova* *mentek* *a* *gyerekek* *nyaralni?*
where went.3PL the kids be.on.vacation.INF
'Where did the kids go for a vacation?
B1: *[Minden* *gyerek]*$_{TOPIC}$ *a* *Balatonhoz* *ment* *nyaralni.*
every kid the Balaton.to went.3SG be.on.vacation.INF
'Every kid went to Balaton for vacation.'

[34] Certain preferences may apply, see Gécseg (2001), É. Kiss (2005).
[35] Kálmán (2001) identifies three subfields, the *is* ('too')-field, the *minden* ('every')-field and the *sok* ('many')-field.

B2: *[Néhány gyerek]$_{CT}$ a Balatonhoz ment (néhány pedig a*
some kid the Balaton.to went.3SG some conversely the
Mátrába).
Mátra.in
'Some kids went to the Balaton (and some went to the Mátra).'

b A: *Hova mentek mindig nyaralni a gyerekek?*
where went.3SG always be.on.vacation.INF the kids
'Where did the kids always go for vacation?
B: *A gyerekek [mindig]$_{BI}$ a Balatonhoz mentek nyaralni.*
'The kids always went to the Balaton for vacation.'

(5) a A: *János soha nem ment Katival haza?*
John never not went.3SG Kate.with home?
'Did John never go home with Kate?'
B: *Nem, János [MINDIG]$_{CF}$ Katival ment haza.*
no John always Kate.with went.3SG home
'No, John ALWAYS went home with Kate.'

b A: *Mikor ment haza János Katival?*
When went.3SG home John Kate.with
'When did John go home with Kate?'
B: *János [mindig]$_{IF}$ Katival ment haza.*
'John always Kate.with went home.'

c A: *Kivel ment haza János?*
with.whom went.3SG home John
'Who did John go home with?'
B: *János [mindig]$_{CI}$ Katival ment haza.*

Also, the +Q counterpart of Topics, Sorting keys occur here (6a). Questioning foci, however are not to be located here (6b). To enforce this restriction, I constrain question phrases here to be +DISCOURSE-STRUCTURING at information-structure.

(6) a *[Ki]$_{SORTING\ KEY}$ hova ment el?*
who where went.3SG away
'Who went where?'

b **[Ki]$_{QUESTIONING\ FOCUS}$ el-ment?*
who away-went.3SG

The next position is the much-discussed "focus-position" of Hungarian, which is Spec/VP in this framework.[36] It is easy to see if something occupies this position as such elements trigger the inversion of the verb and the preverb (preverbs normally occupy this position, immediately

[36] It had been Spec/VP in the Chomskyan analyses (e.g. É. Kiss 1992) as well before Brody's (1990) FocusP-analysis gained wide acceptance.

in front of the verb, see 6 above and 7-8 below). This is the primary place for Information foci (answers to wh-questions) and for Contrastive foci (e.g. corrections, emphasis, etc.). (8a) is also an example for the +Q counterpart of IF, Questioning focus occupying this slot.

(7) *János meg-vette az autót.*
John PV-bought.3SG the car.ACC
'John bought the car.'

(8) a A: *Mit vett meg János?*
what bought PV John?
'What did John buy?'
B: *János [az autót]$_{IF}$ vette meg.*
'John bought the car.'

b A: *János a motort vette meg?*
John the motorbike.ACC bought PV
'Did John buy the motorbike?'
B: *Nem, János [AZ AUTÓT]$_{CF}$ vette meg.*
no John the car.ACC bought.3SG PV
'No, John bought the CAR.

IF and CF are +NEW, +D-STRUCTURING. However, the preverbal constituent is not necessarily related to a discourse subquestion. For example, (7) could be an answer for a question like "What happened?", where the entire sentence is the answer. Also, bare nouns may be positioned here, in contexts like (9). These are still part of the pragmatic assertion, so they are +NEW. As such, they should be seen as instances of Completive information (+NEW, –D-STRUCTURING).

(9) a A: *Mit csinál János?*
what.ACC does John
'What is John doing?'
B: *János [fát]$_{CI}$ vág.*
John wood.ACC cuts
'John is cutting wood.' ('John is wood-cutting.')

This seems to suggest that the Spec/VP is the locus of +NEW IF-categories. This is not the entire picture though. As noted, in certain contexts, even Topical entities may be in Spec/VP, as in (10).

(10) A: *Ki Szentgyörgyi Albert?*
who Szentgyörgyi Albert
'Who is Szentgyörgyi Albert?

B: *[Szentgyörgyi Albert]*$_{TOPIC}$ *fedezte* *fel a* *C-vitamint.*
Szentgyörgyi Albert discovered.3SG up the vitamin.C
'Szentgyörgyi Albert (was the one who) discovered vitamin C.'

So it seems that the only IF-category excluded from the Spec/VP slot is Background Information, the i-structural category with full negative specification. As we have seem sentence adverbs cannot be topics, so *valószínűleg* is ungrammatical in (11a/B). It would be marginally acceptable as focus (11b/B).

(11) a A: *Ki* *Szentgyörgyi* *Albert* *valószínűleg?*
who Szentgyörgyi Albert probably
'Who is Szentgyörgyi Albert probably?'
B: *Szentgyörgyi Albert [valószínűleg]*$_{BI}$ *fedezte fel a C-vitamint.*

b A: *Én* *biztos* *vagyok* *benne,* *hogy* *Szentgyörgyi Albert* *fedezte*
I certain am that.in that(C) Szentgyörgyi Albert discovered.3SG
fel a *C-vitamint.*
up the vitamin.C
'I am certain that Szentgyörgyi Albert discovered vitamin C.

B: *Szerintem* *csak* *[valószínűleg]*$_{CF}$ *fedezte* *fel a* *C-vitamint*
in.my.opinion only probably disovered.3SG up the vitamin.C
'I think that he PROBABY discovered vitamin C.'

How such constructions should be analyzed is still under research. It seems from this discussion that the Spec/VP position in Hungarian is not particularly unique in terms of information-structure: there is no IS-category that is exclusively associated with this slot. Rather, what seems to be peculiar about the Hungarian "focus position" has to do with its semantics. These additional semantic restrictions are revolving around exhaustivity although the empirical and theoretical landscape is debated (for references, see footnote 26).

While I acknowledge the significance of these semantic restrictions and their necessity for the full characterization of Spec/VP, in my present analysis I do not include them, as information-structure is defined here as a level for representing those aspects of linguistic representation which concern not the information-content of a sentence itself, but rather the way it is presented. Exhaustivity is a concept clearly belonging to truth-conditional semantics, and is not simply a packaging phenomenon.

Finally, the main verb itself may either be Information/Contrastive focus (12) or Background information (13) in the sentence. That is, it is either the main assertion of the sentence, or entirely part of the background.

(12) a A: *Mit csinál János?*
what does John
'What is John doing?'
B: *[Eszik]$_{IF}$.*
eats.3SG
'He is eating.'

b A: *János eszik?*
John eats
'Is John eating?'
B: *Nem, [ISZIK]$_{CF}$.*
no drinks.3SG
'No, he is DRINKING.'

(13) A: *Mit eszik János?*
what.ACC eats John?
'What is John eating?'
B: *Halat [eszik]$_{BI}$ János.*
fish.ACC eats.3SG John
'John is eating fish.'

As already mentioned, the word-order in the postverbal area is generally free and different word-order permutations do not result in different interpretations. This part of the Hungarian sentence lies outside the scope of the dissertation, so I finish this section off and go on to show the properties of subordinate clauses.

3.2 Subordinate clauses in Hungarian[37]

According to É. Kiss (2002), the structure of subordinate clauses essentially parallels that of main clauses, so they contain the same structural positions. In her account, there are only two differences: one, subordinate clauses are introduced by the complementizer (*hogy,* 'that(C)'), and two, they are often associated with a pronoun. In this dissertation, I subscribe to the generally accepted view of É. Kiss (2002), that is, the internal structure of subordinate clauses is assumed to be similar to that of main clauses, as described in the previous section. A sketch is presented in Figure 2.

The point where debates are to be found in the literature is about the nature of the relationship between the matrix verb, the associate pronoun and the subordinate clause. The

[37] The scope of this dissertation only extends to subordinate clauses that are associated with some argument function. Consequently, cases like (i), where the predicate is fully saturated without a subordinate clause will not be discussed. Relative clauses are also set aside.

(i) *János boldog volt, (hogy nyert).*
John happy was that(C) won.3SG
'John was happy (that he won).'

questions are the following: What is the nature of the associate pronoun? How is the pronoun associated with the subordinate clause?

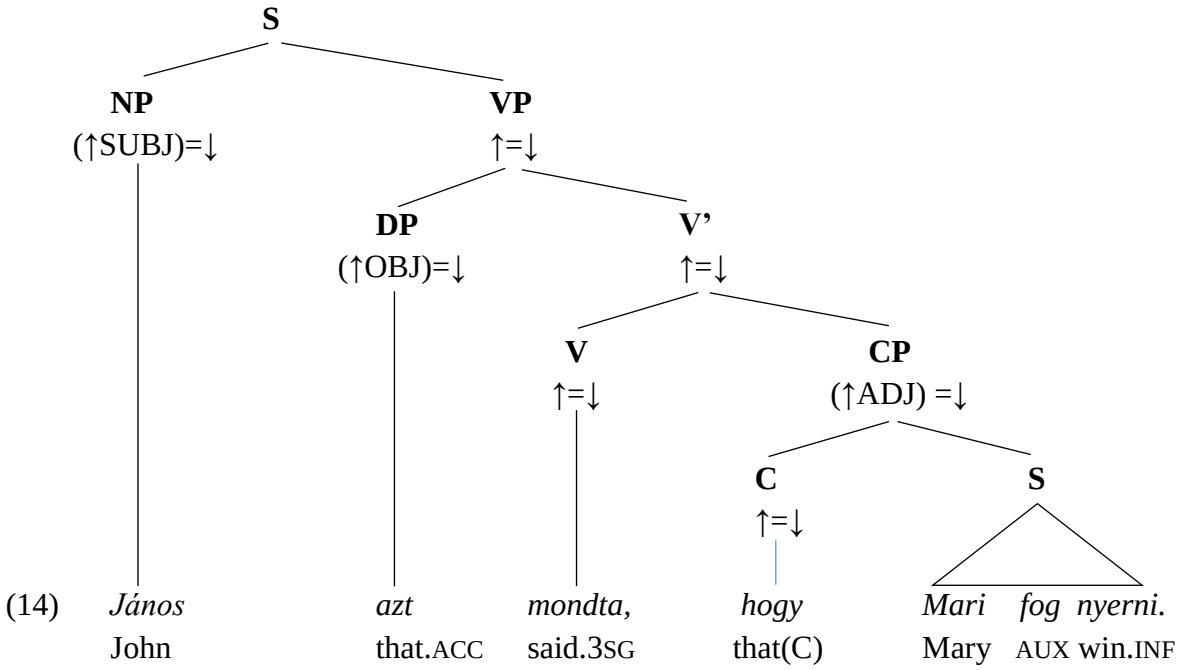

Figure 2.
A Hungarian sentence with a subordinate clause.

These questions are crucial for this dissertation since the proper analysis of the Operator fronting (OF) construction cannot be carried out without clarifying these issues. One example for OF is shown in (15b): some constituent which is thematically related to an embedded predicate surfaces in a matrix sentence operator position (the detailed discussion of OF will take place in 5.2). What is important here is that the verbs that can participate in OF all take subordinate clauses (15c-d). On the other hand, elements undergoing OF are in complementary distribution with the associate pronoun: if one is present, the other one cannot appear as in (16).

(15) a *Az-t mondtad, hogy jön János.*
that-ACC said.2SG that(C) comes John.
'You said that John comes.'

b *János(-t) mondtad, hogy jön.*
John(-ACC) said.2SG that(C) comes
'John you said that he will come.'

c **Az-t futottad, hogy jön János.*
that-acc ran.2SG that(C) comes John.

d **János-t futottad, hogy jön.*
John-ACC ran.2SG that(C) come'

(16) a *Jánost *(azt) mondtad *(azt), hogy jön.*

b

Párizsba	**(azt),*	*mondtad*	**(azt),*	*hogy*	*mész.*
Paris.to	that.ACC	said.2SG	that.ACC	that(C)	go.2SG

In the following sections, I will go into the details of the aforementioned issues. In brief, I will argue that contra Kenesei (1992/1994) the associate pronouns are not expletives but contentful demonstrative pronouns. When such pronouns are present, the subordinate clause itself is an adjunct. However, the pronoun is optional. Thus, when it is not present, the embedded clause itself picks up the grammatical function otherwise designated to the pronoun. As the complementizer is also optional, following LFG's Economy of Expression principle (see section 1.2), a CP projection for the subordinate clause is only assumed when *hogy* ('that(C')) is present. In other cases, the subordinate clause should be analyzed as an S. The constraints on the distribution of the pronoun and the complementizer then naturally follow from the proposed system and from general principles.

3.2.1 General properties of Hungarian subordinate clauses

The structure of Hungarian subordinate clauses is described in Kenesei (1992/1994) in detail so this section heavily draws from Kenesei's work with respect to empirical data, though the subsequent analysis will be considerably different. In the most basic form of the Hungarian subordinate clause, there is a main verb, an associate pronoun and a subordinate clause. The pronoun is a demonstrative one in form and bears some grammatical function, determined by the verb: it can be a subject (17a), an object (17b) or an oblique complement (17c).

(17) a

Az	*valószínű,*	*hogy*	*János*	*fog*	*nyerni.*
that	likely	that(C)	John	AUX	win.INF

'It is likely that John will win.'

b

Az-t	*mondtam,*	*hogy*	*János*	*fog*	*nyerni.*
that-ACC	said.1SG	that(C)	John	AUX	win.INF

'I said that John would win.'

c

Arra	*számítok,*	*hogy*	*János*	*fog*	*nyerni.*
that.onto	expect.1SG	that(C)	John	AUX	win.INF

'I expect that John will win.'

As can be seen from these examples, the demonstrative pronoun is usually in its distal form. However, in certain contexts the proximal version can also appear. Intuitively, this is related to the discourse-status of the information presented in the associated clause. *Is* ('too') signals in

(18) that the proposition had already been evoked in the discourse and in this case the CP may easily be dropped. If the CP is there, the absence of *is* ('too') would make (18) worse. A precise formulation of this idea awaits further research.

(18) *Én is ezt mondom, hogy János fog nyerni.*
I too this.ACC say.1SG that(C) John AUX win.INF
'I also say that John will win.'

Although such pronouns are commonly present in complex sentences containing subordinate clauses, they are in fact optional in many circumstances. So in sentences like (19a-b), the demonstrative could be absent, without a change in meaning. There is variation with regard to the droppabilty of pronouns bearing oblique case (as in 19c and 19d).

(19) a *Valószínű, hogy János fog nyerni.*
likely that(C) John AUX win.INF
'It is likely that John will win.'

b *Mondtam, hogy János fog nyerni.*
said.1SG that(C) John AUX win.INF
'I said that John would win.'

c **(Arra) számítok, hogy János fog nyerni.*
that.onto expect.1SG that(C) John AUX win.INF
'I expect that John will win.'

d *(Arról) meggyőződtem, hogy János a győztes.*
that.from ascertained.1SG that(C) John the winner
'I ascertained that John was the winner.'

In these examples, the pronoun is always preverbal. Moreover, it is associated with some "prominent" discourse-function, IF, CF or CT. Interpreting them as standard topics is not possible. Thus, sentence (20) is ungrammatical with a falling intonation, it must receive the fall-rise contour that characterizes contrastive topics. (The focus interpretation is excluded since the Spec/VP position is occupied by the preverb.)

(20) *Az-t meg-mondtam, hogy János fog nyerni.*
that-ACC PV-said.1SG that(C) John AUX win.INF
'That John will win, I did say.'

However, this could just be a side-effect of the event-structural properties of the verbs in question. É. Kiss (2005) observes that although in principle, any argument of a predicate may

be positioned in the topic-field, not all possibilities are equally neutral information-structurally. Take change-of-state verbs as an example. Sentence (21a) with the predicate *nyír* ('cut') is entirely natural with a standard falling intonation, in (21b) *a füvet* ('the grass.ACC') must be either focus (IF/CF) or Contrastive topic.

(21) a *Mari nyírja a füvet.*
Mary cuts the grass.ACC
'Mary is cutting the grass.'

b *[A füvet]*$_{\#TOP/CT/IF/CF}$ *nyírja Mari.*
'Mary is cutting the GRASS.'/ 'The grass, Mary is cutting.'

According to É. Kiss (2005) the reason for this is that depending on the event-type they express, predicates lexically select some of their arguments as "main news" and these must be in the "predicate part" of the sentence: the postverbal area or the focus position. For a change-of-state verb like *nyír* ('cut'), the theme is the "main news". So in (21a) it is postverbal, which is the default mapping. If it is not postverbal, like in (21b), the informativeness of the predicate part must be strengthened via some other means, like the focussing of the theme. Other methods are also available: focussing some other element (an adjunct in 22a), creating a contrastive topic (22b), or temporally modifying the predicate (22c).

(22) a *[A füvet]*$_{TOPIC}$ *tegnap nyírta le Mari.*
the grass.ACC yesterday cut.PAST.3SG down Mary
'Mary cut the grass yesterday.'

b *[A füvet]*$_{TOPIC}$ *nyírja Mari, de a bokrot nem.*
the grass.ACC cuts Mary but the bush.ACC not
'The grass, Mary is cutting, but the bush, she is not.'

c *[A füvet]*$_{TOPIC}$ *már nyírja Mari.*
the grass.ACC already cuts Mary
'Mary is already cutting the grass.'

The verbs under discussion behave similarly. The main news for them is whatever semantic role is associated with the respective grammatical functions in (23-26) (theme/proposition), so this is postverbal by default. If it is preverbal, their i-structural behavior parallels (22).

(23) a *Mari mondott egy történetet.*
Mary said.3SG a story.ACC
'Mary told a story.'

b *[Egy történetet]*$_{\#TOP/CT/IF/CF}$ *mondott Mari.*

(24) a *Mari fontolgatja a lemondását.*
Mary contemplates the resignation.POSS.3SG.ACC
'Mary is contemplating her resignation.'

b *[A lemondását]*$_{\#TOP/CT/IF/CF}$ *fontolgatja Mari.*

(25) a *Mari sérelmezi a cselekedeteidet.*
Mary resents the acts.POSS.2SG.ACC
'Mary resents your acts.'

b *[A cselekedeteidet]*$_{\#TOP/CT/IF/CF}$ *sérelmezi Mari.*

(26) a *Meg-győződtem János győzelméről.*
PV-ascertained.1SG John victory.POSS.3SG.FROM
'I ascertained that John won.'

b *[János győzelméről]*$_{IF/CF}$ *győződtem meg.*
c *[János győzelméről]*$_{\#TOP/CT}$ *meggyőződtem.*

The pronoun behaves in a similar fashion, with the addition that some verbs, e.g. *mond* ('say'), *gondol* ('think'), *állít* ('claim')) disprefer to be a neutral initial element in the "predicate part".[38]

(27) *#Mondtam/gondoltam/ állítottam azt, hogy János fog nyerni.*
said.1SG thought.1SG claimed.1SG that.ACC that(C) John AUX win.INF
Intended: 'I said/thought/claimed that John would win.'

Thus, these verbs require some additional weight in this configuration, e.g. by adding a quantifier (28a), a verum focus (28b), or filling the Spec/VP slot with a particle (28c). This is probably connected to the fact that while lexical noun objects are indeed possible for these verbs (as in 23, to be also discussed in 3.2.2.), they are quite restricted.

(28) a *Mindig mondtam azt, hogy János nagyon okos.*
always said.1SG that.ACC that(C) John very smart
'I have always said that John is very smart.'

b *El-mondtam azt, hogy szerintem János fog nyerni.*
away-said.1SG that.ACC that(C) in.my.opinion John AUX win.INF
'I told (them) that in my opinion John would win.'

c #Mondtam / MONDTAM azt, hogy János fog nyerni.
said.1SG that.ACC that(C) John AUX win.INF
'I DID say that John would win.'

If the verb is not of this kind, the pronoun may naturally be postverbal, just like in (29).

[38] I would like to thank Balázs Surányi for this observation.

(29) a *Fontolgatom azt, hogy lemondok.*
contemplate.1SG that.ACC that(C) resign.1SG
'I am contemplating that I resign.'

b *Sérelmezem azt, hogy korán hazamentetek.*
resent.1SG that.ACC that(C) early home.went.2PL
'I resent that you went home early.'

c *Meggyőződtem arról, hogy János győzött.*
ascertained.1SG that.from that(C) John won.3SG
'I ascertained that John won.'

What I conclude from this is that the placement and the discourse function of the pronoun associated with the subordinate clause is dependent on the semantic type of the verbs in question and is not a special property of the construction with the pronoun.

Another optional element in such sentences is the complementizer *hogy*. Following the principle of Economy of Expression, this means that the subordinate clause can be expressed either as a CP (if there is a complementizer) or as an S (when there is not). (30) here shows the complementizer-less versions of (17).

(30) a *?(Az) valószínű, János fog nyerni.*
that likely John AUX win.INF
'It is likely that John will win.

b *Azt mondtam, János fog nyerni.*
that-ACC said.1SG John AUX win.INF
'I said that John would win.'

c *Arra számítok, János fog nyerni.*
that.onto expect.1SG John AUX win.INF
'I expect that John will win.'

The droppability of the complementizer is subject to a number of lexical and structural constraints, as discussed in Kenesei (1992/1994). One of the most prominent restrictions is that "focusing in the main clause blocks the deletion of the embedded complementizer" (Kenesei 1994:336), as in (31b-c). However, it must be added that in the case of Information focus, only lexical noun foci block the omission of the complementizer. If the IF is a pronoun, it may be dropped (31a/B). The situation is similar with Contrastive topics (31d-e), but remember that CTs are always associated with a focus so the complementizer must be present, independent of the presence of the CT.

(31) a A: *Mit mondtál?*
what.ACC said.2SG
'What did you say?'
B: *[Azt]$_{IF}$ mondtam, (hogy) János fog nyerni.*
thatACC said.1SG that(C) John AUX win.INF
'I said that John would win.'

b A: *Kinek mondtad, hogy János fog győzni?*
who.DAT said.2SG that(C) John will win.INF
'To whom did you say that John would win?'
B: *[Marinak]$_{IF}$ mondtam, *(hogy) János fog győzni.*
Mari.DAT said.1SG that(C) John will win.INF
'I said to Mary that John would win.'

c A: *Annának mondtad, hogy János fog győzni?*
Anne.DAT said.2SG that(C) John will win.INF
'Did you say it to Anne that John would win?'
B: *Nem, [MARINAK]$_{CF}$ mondtam, *(hogy) János fog győzni.*
No Mari.DAT said.1SG that(C) John will win.INF
'I said to Mary that John would win.'

d *[Azt]$_{CT}$ MONDTAM, *(hogy) János fog nyerni, de azt nem*
that.ACC said.1SG that(C) John AUX win.INF but that.ACC not
hogy ma.
that(C) today
'I did say that John would win, but I did not say that he will win today.'

e *[Marinak]$_{CT}$ MONDTAM, *(hogy) János fog győzni de Annának*
Mari.DAT said.1SG that(C) John will win.INF but Ann.DAT
nem.
not
'To Mary, I did say that John would win, but I did not say it to Anne.'

It has been claimed that dropping both the pronoun and the complementizer leads to unacceptability (Jánosi 2014:104-107). Example (32) and the judgment are from her work.

(32) *Mondta, *(hogy) új autót vett.*
said.3SG that(C) new car.ACC bought.3SG
'He said that he had bought a new car.'

However, a closer investigation reveals that the generalization is premature. For example, Gábor Alberti (p. c., Budapest, 2015-06-18) has called my attention to the fact, also noted in Kenesei (1992:620/1994:337), that matrix verbs that take an interrogative complement readily allow dropping both the pronoun and the complementizer.

(33) *Kérdezte, mikor vettem új autót.*
asked.3SG when bought.1SG new car.ACC
'He asked when I had bought a new car.'

I think the weirdness of the complementizer-less version of (32) stems from the fact that without the pronoun, the matrix predicate *mond* is interpreted as an emphatic element in the sentence. In other words, in isolation, *mond* in (32) is interpreted as the focus of the sentences ("He DID say that he bought a new car"). If that is the case, the restriction showed earlier is the culprit: the main clause focus prevents dropping the complementizer. If a proper context is construed, where it is ensured that no emphasis is put on the predicate itself, dropping the pronoun and the complementizer becomes possible. Consider (34), which could be a reply to (33), or (35). In these cases, the information-content of the embedded clause is much more prominent, allowing for an unstressed main clause.

(34) *?Mondtam, nem vettem én újat, csak átfestettem a régit.*
said.1SG not bought I new.ACC just repainted.1SG the old.ACC
'I said that I hadn't bought a new one, I just repainted the old one.'

(35) A to B: *Fordulj balra.* [B turns right.] *Mondom balra.*
turn.IMP.2SG left.to say.1SG left.to
'Turn left. Left I say.'

Other predicates, like *kérdez* ('ask'), are less prone to this interpretational bias, and the result is acceptable sentences like (36). Other examples for the same phenomenon include *ígér* ('promise') *gondol* ('think'), *remél* ('hope') or *gyanít* ('suspect').

(36) *Ígérem,/ Gondolom,/ Remélem,/ Gyanítom, János fog nyerni.*
promise.1SG think.1SG hope.1SG suspect.1SG John AUX win.INF
'I promise/think/hope/suspect John will win.'

Nevertheless, there does seem to be a considerable amount of lexical variation in the omission possibilities. Nominal predicates do not allow the complementizer to be dropped (37a). A number of verbal predicates allow only either the pronoun or the complementizer to be dropped (37b), conforming to Jánosi's (2014) claim. No context could save the ones in (37b).

(37) a *Kár/ Szerencse, *(hogy) János nyert.*
pity luck that(C) John won.3SG
'It's a pity/luck that John won.'

b *(Azt) *jósoltam/ vallottam/ képzeltem, János nyert.*
that.ACC predicted.1SG testified imagined John won.3SG
'I predicted/ testified/ imagined that John won.'

A full account of the omission possibilities and constraints lies beyond the scope of this dissertation, but the phenomenon itself will be featured in section 5.2, where we are going to discuss the argument structural aspects of Operator Fronting. My general position is that dropping these elements is possible by default but syntactic (word class)/semantic (verb-type)/ information-structural (focus or contrast) factors may block the omission.

As a last general point about subordinate clauses, it has to be mentioned that with some verbs, it is possible to use another demonstrative pronoun, for example *úgy* ('so.DIST'). Sometimes it is the only option, or it may alternate with *azt* ('that.ACC'). We will return to these in the next section.

(38) a **Az/ Úgy tűnik, hogy János fog nyerni.*
that so.DIST seems that(C) John AUX win.INF
'It seems that John will win.'

b *Azt/ Úgy gondolom, hogy János fog nyerni.*
that.ACC so.DIST think.1SG that(C) John AUX win.INF
'I think that John will win.'

3.2.2 The expletive-demonstrative debate

There are two main approaches to the characterization of the pronoun and its relation to the embedded clause. According to the first one (put forward by Kenesei 1992/1994, also accepted by Lipták 1998 and Gervain 2002), the pronoun is an expletive and the real, semantic argument of the main verb is the propositional complement clause. The opposing view holds that the pronoun is contentful and it is associated with the complement clause via adjunction (Tóth 2000, Rákosi & Laczkó 2005) or complex NP-formation (É. Kiss (2002)).

Let us take a closer look at the chronologically first view, which holds that the pronouns in question are expletives. To evaluate it, first we need to make a short diversion, to obtain a brief overview of the literature on expletives in general. Then I will evaluate the Hungarian state of affairs in light of this overview.

3.2.2.1 About expletives in general

Expletives are semantically empty, pleonastic elements in a sentence. Their presence is motivated by some structural principle, for example the well-known Extended Projection Principle (EPP) of Chomskyan frameworks, which states that the structural subject position (Spec/IP) must be filled.

In the theory of pronouns developed by Cardinaletti & Starke (1994)[39], expletives are members of the class of "weak pronouns". Weak pronouns contrast with strong ones with respect to several properties: weak pronouns cannot be coordinated, they cannot be in theta-positions or peripheral positions (e.g. left/right dislocation, topicalization), they can refer to nonhuman entities (as opposed to strong pronouns), may have "reduced meaning" (e.g. the impersonal reading of *You can never know what the future brings*), and only weak pronouns can be phonologically reduced.

There are several constructions in English that have been hypothesized to have expletives in them: for instance extrapositions (39a-b)[40], presentational sentences (39c), weather-sentences (39d), and it-clefts (39e).

(39) a *It is obvious that John will* win.
b *It seems that John will win.*
c *There is a unicorn in the garden.*
d *It rains.*
e *It is John who will win.*

Intuitively, in all of these cases, the pronoun seems to have some sort of reduced/zero reference. This can be seen for example from the fact that they cannot be questioned or focussed. This also follows from the theory Cardinaletti & Starke (1994), as these constructions count as ones utilizing peripheral positions.

(40) a **What is obvious that John will win?*
b **ONLY IT is obvious that John will win.*

[39] Cardinaletti & Starke (1994) is a highly influential theory of pronouns, but see Manzini (2014) for a differing opinion.

[40] It has to be added that *obvious* in (39a) could also take a standard DP subject, as in (ia). A clausal subject is also possible (ib). This stands in contrast with *seem* (ii). For different perspectives on this, see Alrenga (2005) and Davies & Dubinsky (2009).

i) a *John's victory is obvious.* b *That John will win is obvious.*
ii) b **John's victory seems.* b **That John will win seems.*

(41) a **What seems that John will win?*

b **ONLY IT seems that John will win.*

(42) a **What rains?*

b **ONLY IT rains.*

(43) a **Where is a unicorn in the garden?*

b **ONLY THERE is a unicorn in the garden.*

(44) a **What is John who will win?*

b **ONLY IT is John who will win.*

In all of these cases, the pronoun is the subject of these sentences. This follows from the EPP, which is a general requirement on subject positions. However, it has been suggested (Postal & Pullum 1988) that the sentences in (45) contain the expletive pronoun *it* as an object:

(45) a *I regretted it every time that I had dinner with John.*

b *We demand it of our employees that they wear a tie.*

c *I still can't believe it that he is gone.*

The proper analysis of such sentences is debated. Rothstein (1995) argues that even though the semantics of these pronouns is bleached to a certain extent, they are in fact not expletives. This can be seen for example from the fact that removing the pronoun results in a slightly altered meaning. Thus, while (45a) means that every event of dinner was matched by an event of regret, the pronounless version could mean that there was only one regretting event (for example, some incident makes me reinterpret my evaluation of the past dinners with John, which may had seemed happy at those times). Another counter-argument is that Postal & Pullum's (1998) verbs do occur with uncontroversially semantic objects.

(46) a *I regretted my decision.*

b *We demand full compensation for our grievances.*

c *I still can't believe the story.*

Another suggestion for an object-expletive in English could be *there* in sentences like (47a). It is known that *believe* can take a non-thematic object in the case of "raising to object"[41] (47b). However, the fact that it behaves differently from a normal non-thematic object for example in

[41] I use "raising to object" as a traditional, descriptive label, not indicating theoretical commitment on my part.

there's inability to be followed by a manner-adjunct (47c), indicates that perhaps *there* should receive some other analysis. (The contrast between 47c and 47d is unnoticed in the literature, as far as my knowledge extends.)

(47) a *I believe there to be a boy outside.*
b *I believe John to be happy.*
c **I believe there strongly to be a boy outside.*
d *I believe John strongly to be happy.*

One possible account for the facts could be that *there* is not the main clause object, but is part of the embedded clause. This would explain why a main clause adverbial cannot follow it. However, *there* could be a subject in a passive sentence (*There was believed to be a boy outside*), and that is a strong indication of its status as an object. An alternative could be to maintain the object status of *there* in (47a), but argue that it is in some way different from non-thematic objects like *John* in (47b). Such an account could be built on the grounds that the non-thematic, expletive status of *there* has been called into question by researchers like Moro (1997) and Tortora (1997). Both of them investigate *there*'s behavior as a subject pronoun. Moro focuses on existential sentences like (39c) while Tortora puts the emphasis on predicates other than *be*, like *arrive* in (48).

(48) *There arrived four women at the station.*

Both Moro (1997) and Tortora (1997) discard the expletive analysis, but they replace it with different theories. Moro argues that *there* is a predicate, while Tortora advocates an analysis where it is a weak locative argument (weak in the sense of Cardinaletti & Starke 1994). Either of these approaches could serve as a starting point for an account, but the details of that account await further enquiry. Considering the available research about expletives, it seems reasonable to maintain the generalization that expletives occur as subjects.

Going on with the properties of expletives, a contrast is often mentioned between *there* in (39c) and weather-*it* (39d). Namely, Chomsky (1981) observed that while weather-*it* can bind a PRO in an adjunct clause (49a), *there* cannot (49b). Extraposition-*it* patterns with *there* in this respect.

(49) a *It$_i$ often clears up after PRO$_i$ raining heavily.*

b **There$_i$ is often a party after PRO$_i$ being a wake.*

c **It$_i$ is often obvious that Jack is a liar after PRO$_i$ being dubious that he is honest.*

These facts led Chomsky (1981) to the conclusion that weather-*it* is not an expletive, but it is referential in some generic sense. Chomsky labelled it as a "quasi-argument". The fact that existential *there* only possesses a person feature but not a number feature is in harmony with this view: given their status as syntactic dummies, expletives should be featurally impoverished. (The lack of number features can be seen for example from a pair like *there is a boy outside*/ *there are two boys outside*, where it is the logical subject (*boy*) which determines the number feature of *be*). Consequently, the quasi-argument view seems to be dominant today.[42]

The pronoun in *it*-clefts (39e) has received much less attention in the literature. It shares several properties with the already mentioned expletive-candidates: it cannot have a discourse function, cannot be questioned. So the default assumption in the literature seems to be that it is an expletive (e.g. É. Kiss 1998). However, Hedberg (2000) calls attention to the fact that the form of the introductory pronoun in a cleft sentence like (39e) is not fixed, but it is determined by the cognitive status of the associated clause. Under the appropriate context it can be replaced by some other pronoun (see 50a). This is unexpected under an expletive-analysis: expletives are purely syntactic objects, so they should be unaffected by semantic/pragmatic considerations. Other expletive-candidates are fixed in their form.

(50) a *I wasn't surprised by the massacre in China. [pause] This is not Iowa we're talking about. This is a different society.*

b **This is obvious that John will win.*

c **This seems that John will win.*

d **Here is a unicorn in the garden.*

e **This rains.*[43]

[42] For an alternative view (couched in Relational Grammar), where weather-*it* is treated as a "lexically selected expletive", see Alba-Salas (2004).

[43] As already mentioned, weather-predicates are often assumed to have quasi-argumental subjects. If that is true, we would expect to find some other subjects than *it*. While English examples are scarce, Bennis (1986) argues that the pronoun in Dutch weather-sentences can be replaced by various words, see e.g. (i). There are some Hungarian examples as well (ii).

i) *Het / De wind waait hard.*
it the wind blows strongly
'There is strong wind.' / 'The wind blows strongly.'

ii) *Esik (az eső).*
rains/falls the rain
'It (/the rain) is raining.'

What is evident from this discussion is that the category of "expletive" is not a homogeneous one, either empirically or theoretically. Nonetheless, certain generalizations seem to be plausible:

- Expletives are semantically empty, thus, they are not theta-role bearers.
- They reject discourse-related constructions (questions, foci, topics).
- They are required by some structural principle (e.g. EPP), to fill a certain position (to act as a dummy subject).
- They are featurally impoverished.
- Their form is unaffected by semantic/discourse considerations.

3.2.2.2 Expletives in Hungarian?

After this overview, we are now in the position of taking a closer look at the associate pronoun-subordinate clause construction of Hungarian. As already noted, according to Kenesei (1992/1994), verbs taking a subordinate clause in Hungarian strictly subcategorize for a CP as the realization of their propositional argument. Kenesei supports this view with examples like (51-54), in which replacing the CP with a semantically related DP results in ungrammaticality.

(51) a *(Az) szerencse volt, hogy idejekorán vettek fel kölcsönt.*
that luck was that(C) in.time bought.3PL up loan.ACC
'It was lucky that they had taken out a loan in time.

b **A kölcsön felvétele szerencse volt.*
the loan taking.out.POSS.3SG luck was

(52) a *(Az) jó volt, hogy már tegnap megérkeztetek.*
that good was that(C) already yesterday arrived.2PL
'It was good that you had already arrived yesterday.'

b **A tegnapi megérkezésetek jó volt.*
the yesterday arrival.POSS.2PL good was

(53) a *Mari azt hitte, hogy jókor szólalt meg.*
Mary that.ACC believed.3SG that(C) at.a.good.time spoke.3SG PV
'Mary believed that she had spoken at the right time.'

b **Mari a jókori megszólalást hitte.*
Mary the at.a.good.time speaking.ACC believed.3SG

(54) a *Azt mondta, hogy tudja a választ.*
that.ACC said.3SG that(C) knows the answer.ACC
'S/he said that s/he knew the answer.'

b **A válasz tudását mondta.*
the answer knowing.ACC said.3SG

Furthermore, Kenesei (1992/1994) proposes that subordinate clauses introduced by the complementizer *hogy* ('that(C)') cannot tolerate case-marking. So there is a "division of labor" between the subordinate clause and the pronoun: the first receives theta-marking, while the second carries the case. From this it follows that the pronoun, without a thematic role, is an expletive, forming a chain with the subordinate clause.

From an LFG-viewpoint, this means that these predicates have a lexical entry along the lines of (55), where the pronoun is an athematic argument. (55a) illustrates a subject-expletive (for a predicate like for *szerencse* ('luck') in 51), while (55b) is for an object-expletive (like *mond* in 54).

(55) a *predicate* <(COMP)> (SUBJ)

b *predicate* <(SUBJ)(COMP)> (OBJ)

The alternative view (Tóth 2000, É. Kiss 2002, Rákosi & Laczkó 2005) holds that the pronoun is not an expletive, but a real, referring entity and the clause is associated with it via complex NP-formation (É. Kiss) or adjunction (Tóth, Rákosi & Laczkó). One of the main arguments for this view is that the pronoun does not display the behavior of expletives in general, as discussed in the previous section. It may be questioned, explicitly focussed or quantified.

(56) a *Mit mondott Mari?*
what.ACC said.3SG Mary
'What did Mary say?'

b *CSAK AZT mondtad, hogy János fog győzni. (Azt nem, hogy ennyire fölényesen)*
only that.ACC said.2SG that(C) John AUX win.INF
'You only said that John will win. (And not that he will do so by such a large margin.)'

c *Azt is mondtad, hogy a verseny után elmegyünk ünnepelni.*
that.ACC too said.2SG that(C) the competition after away.go.1PL celebrate.INF
'You also said that after the competition, we'll go to celebrate.'

Also, as noted, expletives cross-linguistically occur as subjects, while for example in (56b), *azt* clearly functions as an object. However, Kenesei (1992/1994) proposes that the reason for this state of affairs is that expletives behave in different ways in typologically different languages. The aforementioned restriction (barring nonsubject functions) holds in configurational languages like English. In such languages expletives really just serve as fillers for specific syntactic slots (e.g. Spec/IP). In discourse-configurational languages, including Hungarian, they have a different role: they represent clauses in configurations which are unavailable for the clauses themselves, namely, when they would bear certain discourse functions.[44] The positions in (57a-b) are indeed unsuitable for *that*(C)-clauses in Hungarian.

(57) a **Csak hogy János fog győzni mondtad.*
only that(C) John AUX win.inf said.2SG

b **Hogy a verseny után elmegyünk ünnepelni is mondtad.*
that(C) the competition after away.go.1PL celebrate.INF too said.1SG

According to Kenesei (1992/1994), the source of the restriction is phonological: the focus and the verb must form a phonological phrase, and such a phrase cannot be larger than a finite clause. If this is correct, then this kind of expletive is different from standard ones in terms of its origin of motivation: syntax vs. phonology. However, it is not clear why a phonological restriction would necessarily result in inserting a semantically empty pronoun. It seems much more plausible that phonology should not care about the semantic nature of the pronoun.

It must be added that a leakage of this generalization is that from its perspective, one would expect that a clause will not appear preverbally in a position that can also be occupied by a pronoun. To put it differently, if the function of the pronoun is to represent a clause in a position barred for it, then the clause should not be able to surface in the barred position: we would expect complementary distribution. Although this is borne out in the case of foci, as shown in (58), both a pronoun or a clause may serve as a contrastive topic.

(58) a *[Hogy győzhet]$_{CT}$, NEM gondolta János.*
that(c) win.COND.3SG not thought.3SG John.
'That he could win, John didn't think (of).'

b [*Azt*]$_{CT}$ *NEM gondolta János, hogy győzhet.*
that.ACC not thought.3SG John that(C) win.COND.3SG
'That he could win, John didn't think (of).'

[44] "Rather than being required by the extended projection principle to fill in an empty subject, expletives in Hungarian can occur in quantifier field or topic positions where the clauses are blocked or have decreased acceptability" (Kenesei 1994:324).

What I am suggesting is that while the data in (57) may be explained along the line of Kenesei's (1992/1994) typological proposal, the alternative theory is not excluded by it: Hungarian may also utilize another strategy (real pronoun + adjunction/complex-NP formation) to bypass the restriction illustrated in (57), without the postulation of a new kind of expletive. In other words, the data in (57) are not conclusive, as they can be interpreted from several analytical perspectives. Hence, it is other empirical/theoretical/typological considerations that should decide the issue.

Another argument for the view where the pronoun is a genuine contentful demonstrative is the fact that the complement clause may be dropped, leaving the pronoun stranded. A sentence like (59) is problematic under the expletive-view, since the pronoun does not form a chain with a semantically contentful constituent. In the GB-framework of Kenesei (1992/1994), (59/B) would violate the Chain condition: every nominal must be a part of chain (which may be a one-membered a chain) which contains a case-position and a θ-position. There is no θ-marked position to license the supposedly expletive pronoun in the case position in the sentence (59/B).

(59) A: *Szerintem János okos.*
in.my.opinion John smart
'In my opinion, John is smart'
B: *Én is azt/ ezt mondom. (hogy János okos)*
I too that.ACC this.ACC say.1SG (that(C) John smart)
'I also say so.' ('that John is smart')

Furthermore, Kenesei's (1992/1994) typological generalization seems not to be very well supported by cross-linguistic data. Although there is a considerable body of research about non-configurational languages, to my best knowledge, no other language exhibits the pattern proposed by Kenesei (1992/1994). For instance, Finnish, a language related to Hungarian (which is also discourse-configurational), has an expletive which behaves in an entirely orthodox way: according to Nikanne & Holmberg (2002), Holmberg (2005) and Holmberg (2010), the pronoun *sitä* must be inserted to Spec/TP (and to nowhere else) if nothing else occupies it (60c). No object-expletives are reported in Finnish.

(60) a *Minulle sattui onnettomuus.*
to-me happened accident
'An accident happened to me.'

b **Sattui minulle onnettomuus.*
happened to-me accident

c *Sitä sattui minulle onnettomuus.*
EXPLETIVE *happened to-me accident*
'There happened an accident to me.'

Other non-configurational languages either have no expletives or they have expletives that behave like English expletives do.

Furthermore, Hungarian has another candidate for an element being expletive-like. As already mentioned in (38), some verbs associate the pronoun *úgy* ('so.DIST') with their subordinate clause. The Hungarian equivalent of *seem* is one such verb. Here, *úgy* behaves just like English expletives: it cannot be questioned (61b), focussed (61c), or quantified (61d) and it also cannot be omitted (61e). As no other subject is allowed in such sentences, it is also most probably the grammatical subject of the sentence (60f).

(61) a *Úgy tűnik, hogy János fog nyerni.*
so.DIST seems that(C) John AUX win.INF
'It seems that John will win.' Lit.: 'So seems that John will win.'

b **Hogy tűnik János?*
how seems John

c **ÚGY tűnik, hogy János fog nyerni, (és nem úgy. hogy Péter).*
intended: 'What seems is that John will win, not that Peter will do so.'

d *Úgy tűnik, hogy érdekes lesz a verseny. *Ezenkívül úgy is tűnik, hogy János fog nyerni.*
intended: 'It seems that the race is going to be interesting. What also seems is that John will win.'

e **Tűnik, hogy János fog nyerni.*

f **János úgy tűnik, hogy ő fog nyerni.*
John so.DIST seems that(C) he AUX win.INF

The form of this pronoun (formally, it is also a distal demonstrative) seems to be unaffected by discourse considerations, so the proximal counterpart is always quite degraded, even in contexts that otherwise license this form of the pronoun.

(62) *?#Nekem is így tűnik/látszik, hogy János fog nyerni.*
me.DAT too so.PROX seems that(C) John AUX win.INF
'It seems to me too that John will win.'

Some verbs which occur with the object accusative pronoun *azt,* can also alternatively select *úgy*. These verbs show a mixed behavior in these tests. On the one hand, they may be questioned

and replaced with a proximal counterpart. On the other hand, focussing and quantifying are still ungrammatical.

(63) a *Azt/ úgy gondolom, hogy János fog győzni.*
that.ACC so.DIST think.1SG that(C) John AUX win.INF
'I think that John will win.

b *Te hogy gondolod?*
you how think.2SG
'What do you think?' Lit.: 'How do you think?'

c **CSAK ÚGY gondolom, hogy János fog győzni, úgy nem, hogy nagy fölénnyel.*
only so.DIST think.1SG that(C) John AUX win.INF so.dist not that(C) large margin.with
intended: I think that John will win, and not that he will do so by a large margin.'

d **Úgy is gondolom, hogy János fog győzni.*
so.DIST too think.1SG that(C) John AUX win.INF
intended: What I also think is that John will win.'

e *Én is így gondolom, hogy János fog győzni.*
I too so.PROX think.1SG that(C) John AUX win.INF
'I also think that John will win.'

Another problem with the expletive analysis of the pronouns associated with Hungarian subordinate clauses is that they occur not only as grammatical functions associated with structural case (subject, object), but also as complements of predicates assigning inherent case. This is problematic for an expletive-analysis regardless of one's theoretical persuasion, even if one allows for object-expletives. Inherent case is always associated with the idiosyncratic meaning of the predicate.[45] From an LFG viewpoint, such elements bear the OBL(ique) grammatical function, which in LMT is specified as "semantically restricted" (see section 2.4). This is obviously incompatible with an analysis positing a lack of semantic content. Moreover, the CP-complements of such predicates are always replaceable with case-marked DPs. The apparent lack of such DPs in sentences like (51-54) is one of Kenesei's (1992/1994) main arguments for the expletive-analysis, so their systematic availability is surprising for such an approach. This had led Lipták (1998), who otherwise subscribes to the expletive-analysis in the case of nominative and accusative pronouns, to abandon this aspect of Kenesei's theory and refer to such inherently case-marked pronouns as "argumental referring words".

[45] Chomsky (1986) formulates this as the Inherent case condition: if A is an inherent case assigner, then A assigns case to an NP if and only if A theta-marks the NP.

(64) a *János büszke volt arra, hogy győzött.*
John proud was that.onto that(C) won.3SG
'John was proud that he had won.'

b *János büszke volt a győzelmére.*
John proud was the victory.POSS.3SG.onto
'John was proud of his victory.'

(65) a *János attól tartott, hogy veszít.*
John that.from was.afraid that(C) loses
'John was afraid that he might lose.'

b *János a vereségétől tartott.*
John the defeat.POSS.3SG.from was.afraid
'John was afraid of his (possible) defeat.'

Based on these considerations, the alternative theory of É. Kiss (2002), Tóth (2000) and Rákosi & Laczkó (2005) seems to be a viable alternative. As has been mentioned, in these approaches, the pronouns in question are genuine demonstrative pronouns, and the clauses themselves are associated with them via adjunction or complex-NP formation.

According to these proposals, the predicates, instead of having lexical entries like (55), have entries like (66).

(66) a *predicate* <(SUBJ)>

b *predicate* <(SUBJ)(OBJ)>

The first challenge for this approach is the pattern illustrated in (51-54), namely, that predicates occurring with a pronoun and a subordinate clause often reject a single DP subject or object. I would like to argue while there are indeed certain restrictions, the generalization as a whole does not hold water. My first observation is that even in Kenesei's original examples some judgments are debatable. For instance, (51b) seems to be relatively acceptable to me (it is even better if one takes a modified version like *nagy szerencse* ('big luck')). Also, a web-search results in several examples where *szerencse* ('luck') takes a DP subject.

(67) a *Merkel: micsoda szerencse volt a békés kelet-európai rendszerváltás.*
Merkel: what luck was the peaceful eastern.European change.of.system
'Merkel: what a luck was the peaceful change of the political system in Eastern Europe.'
from: http://hvg.hu/vilag/20140227_Merkel_micsoda_szerencse_volt_a_bekes_ke

b *Óriási szerencse volt számunkra a PC-k gyors elterjedése.*
huge luck was for.us the PC-PL fast spread.POSS.3PL
The fast spread of the PCs was a huge luck for us.
from: Hungarian National Corpus

c *Szerinte inkább véletlen szerencse volt a gép lelövése.*
according.to.him rather random luck was the plane shooting.POSS.3SG
'According to him, the plane's shooting down was sheer luck.'
from: Hungarian National Corpus

Secondly, the landscape of pronoun-distribution is more complex than the picture presented in Kenesei (1992/1994). There are predicates which can never occur with a pronoun, only with a clause. Such predicates are plausibly analyzed as subcategorizing for CPs. However, here the expletive-demonstrative question does not arise in the first place. *Muszáj* ('must') is a one-place predicate of this kind, while *szól* ('tell') is a two-place one. (*Muszáj* is mentioned in Kálmán 2001:170, about *szól*, see Jánosi 2013:61.)

(68) a *(*Az) muszáj (*az), hogy elgyere.*
that must that that(C) come.2SG. SBJV
'It is a must that you come.'

b *Szólt, hogy sikerült a vizsga.*
told.2SG that succeeded.3SG the exam
'S/he told us that s/he had passed the exam.'

c *(*Azt) szólta (*azt), hogy sikerült a vizsga.*

With other predicates, the pronoun is optional but its presence or absence clearly modifies the meaning of the sentence. *Van* ('be') is one such predicate: without *az*, the sentence has an existential reading, with it, the sentence describes a particular situation.

(69) a *Van, hogy János győz.*
be.3SG that(C) John wins
'It happens that John wins.'

b *Az van, hogy János győzött.*
that be.3SG that(C) John won.2SG
'The situation is that John won.'

That the meaning difference between (69a) and (69b) is not simply because of the focus discourse function associated with the pronoun is evidenced by the fact that a distinct focussed version of (69a) actually exists, but with a different pronoun, *olyan* ('such.DIST').

(70) a *Csak olyan van, hogy János győz.*
only such.DIST be.3SG that(C) John wins
'Only that John wins happens.'

If the presence of *az* causes such a difference in meaning, it is unlikely that it should be analyzed as an expletive. If *az* in (69b) is not an expletive, we then expect it to behave like normal pronouns and unlike expletives. This prediction is borne out: a direct question, using the corresponding question word, *mit* ('what') can only be formed about (69b), and not (69a). Hence, only B1 is a proper response in (70).

(71) A: *Mi van?*
what is
approx.: 'What's up?'/ 'What is the situation?'
B1: *Az van, hogy János győzött.*
that be.3SG that(C) John won.3SG
'(The situation is that) John won.'
B2: *#Van, hogy János győz.*
be.3SG that(C) John wins
'It happens that John wins.'

Interestingly, *úgy* ('so.DIST'), a pronominal element which I have argued to behave much like ordinary expletives, can also follow the existential sense of *van*: (72) is synonymous with (69a). However, as expected if *úgy* is an expletive, no question can be formulated about it, see (73). *Olyan* cannot be questioned either, which leaves us uncertain about its status.

(72) *Van úgy, hogy János győz.*
be so.DIST that(C) John wins.
'It happens that John wins.'

(73) **Hogy/Milyen van?*
how/like.what be

What this could mean that there are at least two separate lexical entries for *van*, one which subcategorizes for a CP (as in Kenesei's 1992/1994 proposal), and a separate one that accepts a DP subject (which could be the pronoun). So there is no "division of labor" between the clause and the pronoun: *van* in (69a) takes the clause as a subject and no expletive is needed in the structure. This situation is similar to the one with *szól* ('tell') in (69b), where the clause satisfies the subcategorization requirements of the predicate, without the help of an expletive. In both cases, inserting a pronoun is impossible. From this perspective, the main difference between *van* and *szól* is that *van* has an alternative lexical entry (with a different meaning), taking a DP

subject, which seemingly saves the sentence. However, in reality, it is not saved, but another construction is used.[46]

Another group of predicates may occur with a DP or a CP without a change in meaning. For example, *fontos* ('important') may take a subject pronoun and a clause or just a DP. Thus, it directly contrasts with *szerencse* ('luck') in (51a-b).

(74) a *(Az) fontos (az), hogy János győzött.*
that important that that(C) John won.3SG
'It's important that John won.'

b *Fontos János győzelme.*
important John victory.POSS.3SG
'John's victory is important.'

There are also a lot of examples where an object pronoun (*azt*) can be replaced with a single DP. (75)-(81) represents the capacity of subordinating verbs taking CPs and DPs alternatively, a pattern that is claimed to be ungrammatical by Kenesei (1992/1994). These examples stand in stark contrast with (51)-(54).

(75) a *A miniszter azt fontolgatja, hogy lemond.*
the minister that.ACC contemplate.3SG that(C) resign.3SG
'The minister is contemplating that he may resign.

b *A miniszter a lemondást fontolgatja.*
the minister the resignation.ACC contemplate.3SG
'The minister is contemplating about resignation.'

(76) a *Azt jósolom, hogy János fog nyerni.*
that.ACC predict.1SG that(C) John AUX win.INF
'I predict that John will win.'

b *János győzelmét jósolom.*
John victory.POSS.3SG.ACC predict.1SG
'I predict John's victory.'

(77) a *Valótlanul állítottuk azt, hogy jó idő lesz.*
wrongly claimed.1PL that.ACC that(C) good weather be.FUT
'We wrongly claimed that there will be good weather.'

[46] Alrenga (2005) uses a similar argumentation to distinguish between two senses of *happen* and *appear*. For example, *appear* in (ia) is synonymous with *seem*, while in (ib) it means approximately "became visible".

i) a *It appears that John will win.* b *The sun appeared on the horizon.*

b *Valótlanul állítottuk a jó időt.*
wrongly claimed.1PL the good weather.ACC
'We wrongly claimed that there will be good weather.' Lit: 'We wrongly claimed good weather.'
from: http://index.hu/belfold/2015/05/27/idojaras_elorejelzes_meteorologia/

(78) a *Az orvos azt tanácsolja, hogy sokat mozogjak.*
the doctor that.ACC recommend.3PL that(C) lot.ACC move.IMP.1SG
'The doctor recommended that I exercise a lot.'

b *Az orvos sok mozgást tanácsol.*
the doctor lot movement.ACC recommend.1SG
'The doctor recommends a lot of exercise.'

(79) a *Furcsállom azt, hogy János vesztett.*
find.strange.1SG that.ACC that(C) John lost.3SG
'I find it strange that John had lost.'

b *Furcsállom János vereségét.*
find.strange.1SG John defeat.POSS.3SG.ACC
'I find John's defeat strange.'

(80) a *Azt firtatta, hogy miért távoztam korán.*
that.ACC asked.3SG that(C) why left.1SG early
'S/he pumped me for why I had left early.'

b *A korai távozásom okát firtatta.*
the early leave.POSS.1SG reason.POSS.3SG.ACC asked.3SG
'S/he pumped me for the reason of me leaving early.'

(81) a *Azt sérelmezem, hogy korán távoztatok.*
that.ACC resent.1SG that(C) early left.2PL
'I resent that you had left early.'

b *A korai távozásotokat sérelmezem.*
the early leave.POSS.2PL.ACC resent.1SG
'I resent you leaving early.'

It is true that some prominent subordinating verbs like *mond* ('say') and *hisz* ('believe') are not grammatical with an object that is the result of a nominalization of a clause, like in (53) and (54), but in the light of (75)-(81) that seems to be a lexical restriction on them, rather than a substantive generalization.

Furthermore, even *mond* is capable of taking DP objects, if these are not clause-nominalizations, but simple nouns with a propositional meaning.

(82) *Mondott nekem egy viccet/ hírt/ történetet/ három dolgot.*
said.3SG me.DAT one joke.ACC news.ACC story.ACC three thing.ACC
'S/he told me a joke/ a piece of news/ story/ three things.'

One might argue that (82) represents another lexical entry. Although my semantic intuition would not support this idea, it would not be very far-fetched to hold such a claim. Actually, this is the line of argumentation that I have pursued in connection with *van* ('be') earlier. However, I think that a good case can be made against it in the case for *mond* ('say'). The objects in (82) are manifestations of the same object function that can also be fulfilled by the pronoun. Evidence for this comes from coordination facts.

To understand the reasoning, let us take the English verb *believe* as an introductory example. It can occur with a thematic object as in (83a) or a nonthematic one, as in (83b).[47] (83b is a so-called raising-to-object sentence, see section 6.3 for details). The relevant LFG-style lexical entries are shown in (83a') and (83b').

(83)	a	*I believe the story.*	a'	*believe* <(SUBJ)(OBJ)>
	b	*I believe John to be happy.*	b'	*believe* <(SUBJ)(XCOMP)>(OBJ)

Although they contain an object, the lexical entries cannot be mixed, so the objects cannot be coordinated (see also Rothstein 1995:505).

(84) a **I believe the story and John to be happy.*

Mond ('say') and *hisz* ('believe') also has a usage like *believe* in (83b), where they take a nonthematic object and a non-finite complement.

(85) *Jánost* *mindenki* *okosnak* *mondja/* *hiszi.*
John.ACC everyone smart.DAT say.3SG believe.3SG
'John is said/believed to be smart by everyone' Lit.: 'Everyone says/believes John to be smart.'

What I propose is that while the usage in (85) indeed represents a separate lexical entry (which is like (83b'), an object like the one in (82) and a pronoun in a subordinating sentence are manifestations of the OBJ function of the same lexical entry.

Thus, I expect that the object of (82) and a pronoun-clause complex can be coordinated, but an object like the one in (85) and a single DP or a pronoun-clause complex cannot. This

[47] I do not claim that the lexical entries are entirely unrelated. What I say is that on the syntactic level of f-structure, they clearly have different subcategorizations, while the connection may be represented at the argument-structure.

prediction is borne out, see (86a). Similar sentences may be construed about the other examples in (75)-(81).

(86) a *Tomi éppen mondta a viccet és azt, hogy hol*
Tom just said.3SG the joke.ACC and that that(C) where
hallotta, amikor elment az áram.
heard.3SG when away.went the electricity
'Tom was telling the joke and telling about where he had heard it, when a power outage occurred.'

b *Valótlanul állítottuk a jó időt és azt, hogy lehet*
wrongly claimed.1PL the good weather.ACC and that.ACC that(C) possible
majd kirándulni.
then make.a.trip.INF
'We wrongly claimed that there will be a good weather and that one can make a trip.'

c *Fontolgatom a visszavonulásomat és azt, hogy*
contemplate.1SG the resignation.POSS.1SG.ACC and that.ACC that(C)
ezt hamar megteszem .
this.ACC soon do.1SG
'I'm contemplating about my resignation and that I do this soon.'

d *Furcsállom János vereségét és azt, hogy ez*
find.strange.1SG John defeat.POSS.3SG.ACC and that.ACC that(C) this
mintha nem is érdekelné.
as.if not even interest.COND.3SG
'I find John's defeat and that it doesn't seem to bother him strange.'

e *Sérelmezem a korai távozásotokat és azt, hogy nem is*
resent.1SG the early leave.POSS.2PL and that.ACC that(C) not even
köszöntetek el.
greeted.2PL away
'I resent you leaving early and that you hadn't even said goodbye.'

On the other hand, the lexical entry illustrated in (85) cannot be coordinated with either a pronoun-clause complex or a simple DP object.

(87) a **Jánost okosnak és azt mondtam, hogy győzni fog.*
John.ACC smart.DAT and that.ACC said.1SG that(C) win.INF AUX

b **A viccet/ hírt/ történetet és Jánost okosnak mondtam.*
the joke.ACC news.ACC story.ACC and John.ACC smart.DAT said.1SG

Hisz ('believe') is not really productive in contemporary Hungarian with a simple DP object. However, in an artistic/archaic style, it is capable of taking a DP object (88a-b). Occasionally,

one may come across contemporary examples as well, showing that the structure is not entirely obsolete (88c). If one is willing to take these examples into consideration, *hisz* behaves exactly like *mond* in (85) and (86): the single DP object can coordinate with a pronoun-clause complex but the athematic object cannot, see (88) and (89).

(88) a *?Apám hitte a szavak igazát.*
father.POSS.1SG believed.3SG the words truth.ACC.POSS.3SG
'My father believed (in) the words' truth.
from a Hungarian pop song, http://www.zeneszoveg.hu/dalszoveg/173/zoran/apam-hitte-zeneszoveg.html

b *?Hiszem a római katolikus anyaszentegyházat.*
believe.1SG the roman catholic holy.church.ACC
'I believe (in) the Holy Catholic Church'
from: the Compendium of the Catechism of the Catholic Church
http://www.vatican.va/archive/compendium_ccc/documents/archive_2005_compendium-ccc_hu.html

c *Hiszem Isten jóságát.*
believe God goodness.ACC.POSS.3SG
'I believe (in) God's goodness.'
from: interview with Géza Röhrig in Heti Válasz (Hungarian political magazine), 2015-June-18.

(89) a *Hiszem apám igazát / Isten jóságát*
believe.1SG father.POSS.1SG truth.ACC.POSS.3SG God goodness.ACC.POSS.3SG
és azt, hogy ez a helyes cselekedet.
and that.ACC that(C) this the right deed
'I believe my father's truth/ God's goodness and that this is the right deed.

b **Jánost okosnak és apám igazát/ Isten jóságát hittem.*

In the light of such counterexamples, I would like to argue that Kenesei's (1992/1994) proposal, which is based on the apparent scarcity of DP complements of subordinating verbs, cannot be maintained. As the typological picture seems to weigh rather against than for the expletive-analysis of the pronouns in question, I conclude that an alternative theory, where these demonstratives (*az* and its case-marked derivatives) are referring, is to be preferred.

In É. Kiss's (2002) version of this approach, the pronoun and the clause form a complex noun phrase. Although it is not stated explicitly, most probably É. Kiss assumes that the pronoun-clause complex is base-generated postverbally and the pronoun may be moved to preverbal operator positions afterwards. This framework predicts that "movement" out of subordinate clauses is ungrammatical since complex noun phrases are islands (this is Ross's 1967 Complex Noun Phrase Constraint, CNPC). See (89a) and (90a) below.

As Rákosi (2006b) notes what is problematic about this account is it cannot straightforwardly account for cases when "movement" does take place, namely, when the pronoun is not explicitly present (90b and 91b)[48]. É. Kiss (2002) herself also acknowledges this.

(90) a **János mondtad azt, hogy jön.*
John said.2SG that.ACC that(C) comes

b *János mondtad, hogy jön.*
John said.2SG that(C) comes
'(Of) John you said that he will come.'

(91) a **Mitől mondtad azt, hogy fél János?*
what.from said.2SG that.ACC that(C) fears John?

b *Mitől mondtad, hogy fél János?*
what.from said.2SG that(C) fears John?
'What did you say that John fears?

To overcome this problem, É. Kiss (2002:253) has to stipulate "that a projection containing no phonologically realized material is transparent for subjacency. Then the noun phrase subsuming the argument clause would activate the CNPC only when its nominal head is spelled out phonologically."

As already mentioned, this kind of data is also problematic for Kenesei's (1992/1994) theory, and for a similar reason: both É. Kiss (2002) and Kenesei (1992/1994) would posit zero pronominal heads in (90b) and (91b). In a Chomskyan framework, this would mean that the extraction possibilities are determined only at PF (den Dikken 2010, footnote 6). Without further elaboration, such an approach is unprecedented and thus has reduced plausibility. A further burden for Kenesei's approach is that expletive *pro*s are theoretically problematic: a fundamental issue from a GB/MP perspective is that such elements would contribute neither semantic nor phonological data to the linguistic computation (PF or LF). [49]

There is no need for such stipulations in the proposal of Tóth (2000), to which Rákosi & Laczkó (2005) also subscribe. In these frameworks, a verb like *mond* takes a simple DP object, and the clause itself is an adjunct. (90a) follows, since adjuncts are also islands. This is attested in Hungarian as well.

[48] To keep the picture clear from interference, I only illustrate this with examples where the fronted element bears its original case. Such sentences are analyzed as involving some movement-like dependency (e.g. functional identification in LFG) in every account, while the case-switched version (*János-t*(ACC*) mondtad, hogy jön*) is plausibly thought of as a "base-generated" structure, see section 5.2.

[49] See Bieberaurer (2008) for a detailed argumentation against expletive *pro*s.

(92) a *Jöttem, hogy János lásson engem is.*
came.1SG that(C) John see. SBJV.3SG me too.
'I came so that John can see me too.'

b **János jöttem, hogy lásson engem is.*
John came.1SG that(C) see. SBJV.3SG me too

c **Kit jöttél, hogy lásson János?*
whom came.2SG that(C) see. SBJV.3SG John
intended: 'Whom did you come so that John can see?'

The verb in (92a) does not subcategorize for a proposition at any level, so the clause must be an adjunct, expressing purpose. If we try to extract the subject of the clause, the sentence becomes ungrammatical, as "movement" out of adjuncts is impossible. The same explanation goes for (90a) and (91a).

As for (90b) and (91b), I assume along with Tóth (2000) and Rákosi & Laczkó (2005) that the clause itself functions as the object of the predicate and as such, it is an argument, from which extraction can take place.

Let us conclude this section with some remarks about the place of Hungarian subordinating constructions in the cross-linguistic palette. The idea that a propositional theta-role may be assigned to a pronoun is not unprecedented in the literature. In Dutch, Hoekstra (1983) and Bennis (1986) argued that *het*, which had often been taken to be an expletive, is in fact a referring pronoun. *Het* occurs as a subject of weather-verbs (93a) or as subject/object of some verbs taking propositional complements (93b-c). This second use of *het* is directly comparable to the Hungarian situation.

(93) a *Het regent.*
it rains
'It rains.'

b *Het wordt betreurd dat Jan ziek is.*
it is regretted that(C) John ill is
intended: 'It is regretted that John is ill.'

c *Jan betreurde het dat hij ziek was.*
John regretted it that(C) he ill was.
'John regretted (it) that he had been sick.'

Just like in the case of Hungarian associate pronouns, *het* in (93b-c) is optional. However, when it is present, extraction is impossible from the subordinate clause (94b, 94d). Hoekstra's (1983) and Bennis's (1986) explanation for this is the same as our explanation was for (90)-(91): when

the pronoun is present, it is the argument of the main predicate and the clause itself is an adjunct, thus an island.

(94) a *Wat wordt door iedereen betreurd dat Jan gelezen heeft?*
what is by everyone regretted that John read has
intended: 'What is regretted by everyone that John read?

b **Wat wordt het betreurd dat Jan gelezen heeft?*
what is it regretted that John read has

c *Wat betreurde jij dat hij gezegd had?*
What regretted you that(C) he said had
intended: 'What do you regret that he has said?'

d **Wat betreurde jij het dat hij gezegd had?*
what regretted you it that he said had

The situation is similar in German. Berman (2001) analyzes the pronoun in (95a) as being the object argument of *sagen* ('say'). Just like in Hungarian and Dutch, the presence of the pronoun is optional, but when it is present, it blocks extraction out of the embedded clause (95b).

(95) a *weil er (es) gesagt hat, dass Hans krank ist*
because he it said have that(C) Hans ill is
'because he said that Hans is ill.'

b *Was hat er (*es) gesagt, dass er gelesen hat?*
what has he it said that he read has
'What did he say that he read?'

Finally, the object pronouns in (96) which were claimed to be expletives by Postal and Pullum (1988) but are argued to be real pronouns by Rothstein (1995) also show this pattern.

(96) a *What do you believe (*it) that John will do?*

b *A full compensation, I strongly demand (*it) that I get.*

The emerging pattern may also be related to research done in the area of CP-licensing. Since Stowell (1981) it has been noted that there are two basic ways of integrating sentential arguments:[50] as direct arguments or as appositions to a nominal head (see also Synder 1992). These patterns seem to be a subtype of this latter relation, although it is yet to be investigated

[50] This may be elaborated in terms of CP-licensing or the licensing of propositional arguments. In a Mininalist framework a null C head is hypothesized even in the absence of an overt head. In LFG, in accordance with the Economy of Expression, such cases are treated as bare IPs or exocentric S heads.

what the exact relationship is between adjunction and apposition. What seems to be unique about the Hungarian pattern is that the apposition/adjunction configuration occurs across the board, with a wide range of verbs of saying and cognition, while in other languages this is much more limited.

3.2.3 Patterns of Hungarian subordinate clauses

Let us present the overall picture that emerges out of this view of subordinate clauses. Some verbs realize their propositional argument exclusively with a CP. The CP may either bear SUBJ (97a), or COMP (97b) grammatical function. The lexical entries with the relevant information are shown in (97a'-b').

(97) a *(*Az) Muszáj, *(hogy) elgyere.* a' *muszáj* <(SUBJ)>
that must that(C) come.SBJV.2SG
'You must come.'

b *Szólt, *(hogy) jön.* b' *szól* <(SUBJ)(COMP>
told that(C) comes
'He told (us) that he comes.'

Other verbs can take either a nominal or a clause as a realization of their propositional argument. Functionally, these may be either subjects, objects or obliques. Categorially, a nominal may be realized as a lexical noun (98) or a demonstrative pronoun (99). By default, a clause may be either a CP (if the complementizer is present) or an S (if the complementizer is absent). If the propositional argument is realized as a demonstrative, a clause is an adjunct. If the clause is present but a nominal is not, then the clause is the SUBJ/OBJ/OBL argument of the predicate (100).

(98) a *A vereséged nagy kár számunkra.* *a'* *kár* <(SUBJ)>
the defeat.POSS.2SG great pity for.us
'Your loss is a great pity for us.'

b *A lemondásomat fontolgatom.* b' *fontolgat* <(SUBJ)(OBJ)>
the resignation.ACC.POSS.1SG contemplate.1SG
'I'm contemplating my resignation.'

c *A lemondásomon gondolkodok.*
the resignation.POSS.1SG.on thinking.1SG
'I'm thinking about my resignation.'

c' *gondolkodik* <(SUBJ)(OBL)>

(99) a *Az kár, hogy vereséget szenvedtél.*
that pity that(C) defeat.ACC suffered.2SG
'It's a pity that you have been defeated.'

b *Azt fontolgatom, (hogy) lemondok.*
that.acc contemplate.1SG that(C) resign.1SG
'I'm contemplating (the issue) that I may resign.'

c *Azon gondolkodom, (hogy) lemondok.*
that.on thinking.2SG that(C) resign.1SG
'I'm thinking about (the issue) that I may resign.'

(100) a *Kár, hogy vereséget szenvedtél.*

b *Fontolgatom, hogy lemondok.*

c *Gondolkodom, hogy lemondok.*[51]

To round this chapter up, let us present the simplified c- and f-structural analyses of the (98b), (99b) and (100b), typical representatives of the subject matter of this chapter.

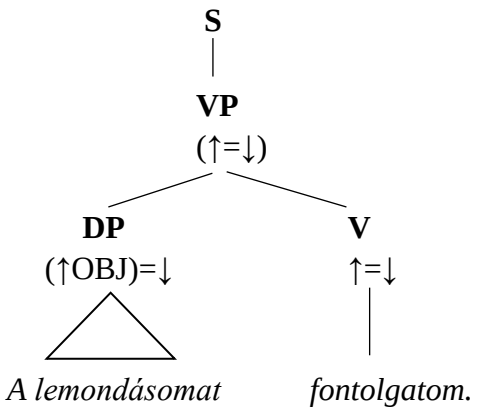

Figure 3a.
C-structure of (98b).

[51] It is possible that the CP is an adjunct in this case, as *gondolkodik* ('think') may be used intransitively. Similar verbs seem to require the oblique arguments to be nominals (see the examples below). This suggests that the realization of the OBL in these cases is restricted to casemarked NPs or DPs.

(i) **(Attól) tartok, hogy nem én fogok győzni.*
that.from be.afraid.1SG that(C) not I AUX win.INF
'I am afraid that I may not win.'

(ii) **(Azzal) számolok, hogy én fogok győzni.*
that.with count.1SG that(C) I AUX win.INF
'I expect that I will win.'

$$\begin{bmatrix} \text{PRED} & \textit{fontolgat} \langle(\text{SUBJ})(\text{OBJ})\rangle \\ \text{SUBJ} & [\text{PRED} \quad \textit{pro}] \\ \text{OBJ} & [\text{PRED} \quad \textit{lemondás}] \end{bmatrix}$$

Figure 3b.
F-structure of (98b).

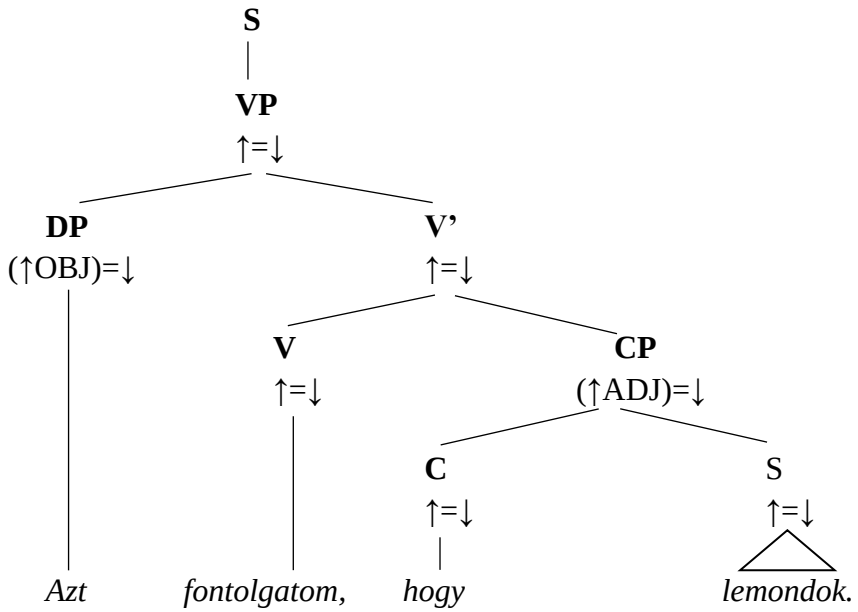

Figure 4a.
C-structure for (99b).

$$\begin{bmatrix} \text{PRED} & \textit{fontolgat} \langle(\text{SUBJ})(\text{OBJ})\rangle \\ \text{SUBJ} & [\text{PRED} \quad \textit{pro}] \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \textit{az} \\ \text{CASE} & \text{ACC} \\ \text{ADJ} & \begin{bmatrix} \text{PRED} & \textit{lemond}\langle(\text{SUBJ})\rangle \\ \text{SUBJ} & \text{PRED} \quad \textit{pro} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Figure 4b.
F-structure of (99b).

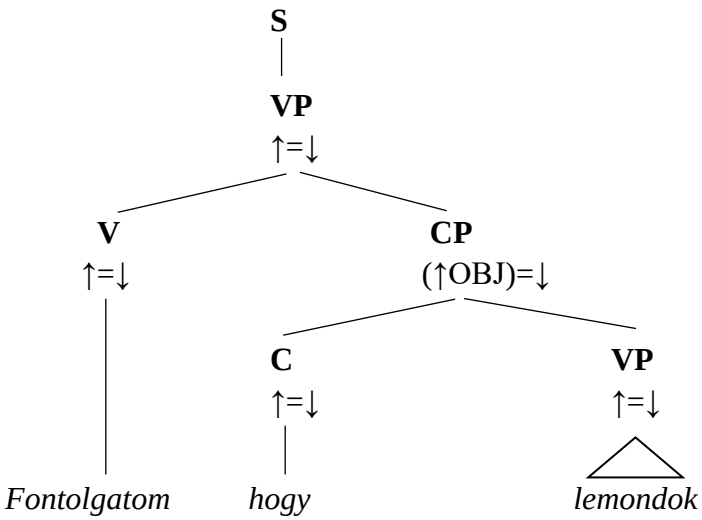

Figure 5a.
C-structure for (100b).

$$\begin{bmatrix} \text{PRED} & \textit{fontolgat} \langle (\text{SUBJ})(\text{OBJ}) \rangle \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \textit{pro} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \textit{lemond} \langle (\text{SUBJ}) \rangle \\ \text{SUBJ} & \text{PRED} \quad \textit{pro} \end{bmatrix} \end{bmatrix}$$

Figure 5b.
F-structure of (100b).

3.3 Conclusion of chapter 3

In this chapter I have provided a general overview of Hungarian sentence-structure. I have taken Laczkó (2014a, 2014b, 2015) as a baseline and elaborated it in two areas. First, information-structural considerations were added, utilizing the framework elaborated in Chapter 1, section 2.6. Second, I also investigated the structures of subordinate clauses. I have shown how the arguments of these verbs are mapped to grammatical functions and categories. I have argued that the optional *az* ('that') pronoun and its case-marked and deictic variants in Hungarian sentences with subordinate clauses should receive an analysis where they are contentful pronouns (demonstratives), and not expletives. This approach was shown to be preferable on theoretical, empirical and cross-linguistic grounds as well.

CHAPTER 4

CONSTRUCTIONS ON THE ENGLISH LEFT PERIPHERY

Although English is considered to be a configurational language, it does not mean that the left-periphery of the sentences is inactive for discourse purposes. The three structures that we are going to investigate are Topicalization (TOP), Clause-initial adjuncts (CIADJs) and Left-dislocation (LD). As these labels may be used for a variety of constructions in different languages, I add a language-specific tag in the abbreviations in order to clearly identify them. (1) provides an illustration of each type.

(1) a *John, I like.* → TOP-Eng
b *Yesterday, John met Mary.* → CIADJ-Eng
c *John$_i$, I like him$_i$.* → LD-Eng

In each of these cases, some element is placed at the left-periphery of the sentence and they bear some sort of discourse prominence. The syntactic status of the clause-initial element is different in each case. In (1a), an argument (an object) is fronted. In (1b), the fronted element is an adjunct. In (1c), the syntactic status of the fronted element itself is not easy to determine, but it is coreferent with an object pronoun.

In this chapter, I will investigate these left-peripheral constructions. In each subchapter first I will show that the constructions each have distinct syntactic and information-structural properties and then I will propose an LFG-based analysis for each of them.

4.1 Topicalization in English

The term "topicalization" originates in Ross (1967). Here it is defined as to those constructions in English where semantically obligatory elements, such as an argument or a predicate is located at a clause-peripheral position on the left edge (clause-initial adjuncts will be discussed in the next section).

4.1.1 Properties of TOP-Eng

(1a) has already illustrated TOP with a direct object NP. However, TOP-Eng is not restricted to this functional-categorial constellation. As (2) shows it can occur with a range of functions and grammatical categories.

(2) a *To John, I gave an apple.* (OBL, PP)
b *That John would win, I would have never said.* (COMP, CP)
c *To arrive in time, John at least tried.* (XCOMP, IP/CP[52])
d *Surrender, John never will.* (PRED, VP)
e *Happy, John will never be.* (PRED, AP)

Topicalization can be rightfully called a long-distance dependency, as the canonical position of the fronted element can be embedded at arbitrary depth. (Stylistic and processing factors may put a limit to this, but not syntax *per se*.)

(3) a *John, Mary said that she saw.*
b *John, Mary said that John believed that she saw.*

Although the depth of the canonical position is unlimited, the path to the embedded position can be constrained by syntactic factors. Dalrymple (2001:392-394) discusses some of these (the following examples are hers). One such constraint is that an intervening non-bridge[53] predicate can disrupt the dependency.

(4) **John, Mary whispered that she saw.*

Furthermore, Ross's (1967) island constraints also affect Topicalization. (5) illustrates this with a Complex Noun Phrase-island (5a), a subject-island (5b), an adjunct-island (5c) and a wh-island (5d).

[52] In the Chomskyan tradition, infinitival *to* is standardly assumed to be heading an IP (or more recently a TP). LFG is less committed on this issue. Falk (2001:139-140) argues that *to* is a complementizer for non-finite clauses.

[53] Traditionally, a "bridge verb" is one that allows extraction from a complement clause. So for example, while *say* and *think* are bridge verbs, *whisper* and *yell* are not.

(i) *Who did you say/think/*whisper/*yelled Bill saw?*

(5) a **John, I doubt the claim that Mary saw.*

b **John, that you saw surprised me.*

c **John, we think that David laughed when we selected.*

d **John, I can't guess why you like.*

It must be noted that in contrast with (5c) TOP-Eng may target an argument inside an untensed adjunct, as shown in (6) where the fronted phrase is the prepositional object of the locative adjunct.

(6) *?That room, Chris teaches his classes in.*

Facts like this indicate that the topicalized constituent maintains a strong relationship with its canonical position. The presence of the topicalized constituent can also be detected with its interaction with binding theory:

(7) a *Himself$_i$, John$_i$ likes.*

b **Him$_i$, John$_i$ likes.*

c *His$_i$ mother, John$_i$ likes.*

d **John$_i$, he$_i$ likes.*

(7) is classic binding-theoretic data. Of course, the technical implementation of these generalizations is different in different frameworks. Minimalism relies on tree-structures, while LFG mainly utilizes f-structures, but the basic idea is the same: a reflexive must be bound in its binding domain, a pronoun must be free in its binding domain and an R-expression must be free everywhere. The pattern in (7) is clear if it is assumed that the topicalized phrases in some sense are also tied to their canonical positions. (For a detailed discussion on binding theory in LFG, see Dalrymple 1993, for an introduction, see Falk 2001, chapter 7).

Deploying multiple TOP-Engs in a sentence causes ungrammaticality.[54] Combining them with questions leads to similar results.

(8) a **An apple, John, I gave.*

b **An apple, whom did you give?*

c **Whom, an apple did you give?*

[54] There has been some debate in the literature about the grammaticality of multiple topicalizations. For a discussion about such cases, see Breul (2004:199-205).

d **Did an apple, you give John?*

e **An apple, did you give John?*

Let us now turn to the information-structural properties of TOP-Eng. For a start, let us review the IS-taxonomy that I have argued for in 2.6. Given this, a precise formulation of the information-structural properties of TOP-Eng will be possible.

		+NEW	–NEW
+D-STRUCTURING	+CONTRASTIVE	Contrastive focus	Contrastive topic
	(–CONTRASTIVE)	Information focus	Topic
–D-STRUCTURING		Completive Information	Background information

Table 1.
My proposed information-structural taxonomy.

Despite its name, the construction is not about standard, neutral topics (Prince 1999). Instead, the topicalized constituent may actually have two distinct functions: it can be interpreted as some kind of focus or a topic of a particular kind. What I would like to argue for is that TOP-Eng is a structure that is tied to the +CONTRASTIVE feature, so topicalized elements are interpreted either as Contrastive focus or Contrastive topic: the set of alternatives becomes active in the discourse at the point the sentence containing the contrastive element is uttered. This position is not unprecedented (see e.g. Chafe 1976 or Molnár & Winkler 2010), but the concise summary from this perspective has not been provided for it in the literature.

Regarding the focus-like reading, Choi (1997), referring to Ward (1988), asserts that the fronted phrase actually refers to two discourse elements: one, a set or a scale, and two, a specification of a value or an element in that set or scale. In (9a) this would mean that the sentence evokes a set of dog names that one may chose and picks *Fido* from that set. The evoked alternative set is lacking in the non-preposed version (9b).

(9) a *Fido I named my dog.*

b *I named my dog Fido.*

Zimmerman (2008) notes that for a wh-question, a topicalized answer is only acceptable if it is against the expectations of the hearer, not as a neutral answer (10/B2). Being unexpected entails

that there is some other entity to which the expectedness is compared. So being unexpected may be interpreted as being a surprising pick from a set of alternatives.

(10) A: *What did you eat in Russia?*
B1: *Caviar*.
B2: *#Caviar we ate!*

A frequent context for this use of TOP-Eng is corrections, which is obviously contrastive in the sense that the speaker rejects a wrong answer and at the same time provides a correct one. Thus, multiple discourse questions are involved.

(11) A: *What did you eat in Russia? Pelmeni?*
B: *No, caviar we ate!*

In the discourse-tree representation I adopted in 2.6.2 (11) may be represented as Figure 1.

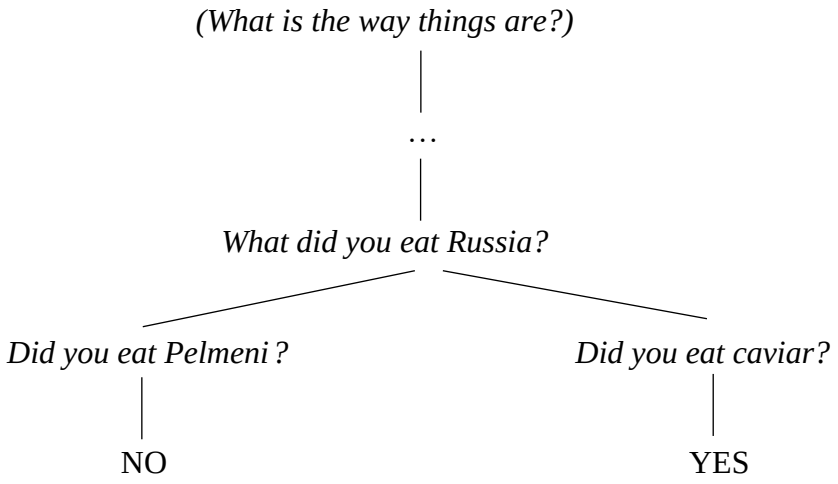

Figure 1.
Discourse-tree for (11).

Birner & Ward (1998) mentions further contexts where a TOP-Eng structure is used. For example, it is used in "proposition assessment", that is, when the speaker expresses some emphatic stance with regards to the uttered proposition. The assessment may be affirmation (12a), suspension (12b) or denial (12c).

(12) a *At the end of the term I took my first schools; it was necessary to pass, if I was to stay at Oxford, and <u>pass I did</u>.*
b *Mark submitted his report late, if <u>submit it he did</u>.*

c *The NBA's new collective-bargaining agreement sounds as though it was written by the same people who put together the Internal Revenue Service's long form. Simple it is not.*

In each of these cases the proposition is contrasted to some alternative state of affairs: (12a) evokes the scenario of not passing, (12b) the scenario of not submitting the report and (12c) a number of other characteristics that the agreement could be assessed with. These interpretations are actually closer to the CT interpretation as there is no clear indication of the truth-status of the evoked subquestions. For example, (12c) may be posited the discourse-tree representation shown in Figure 2.

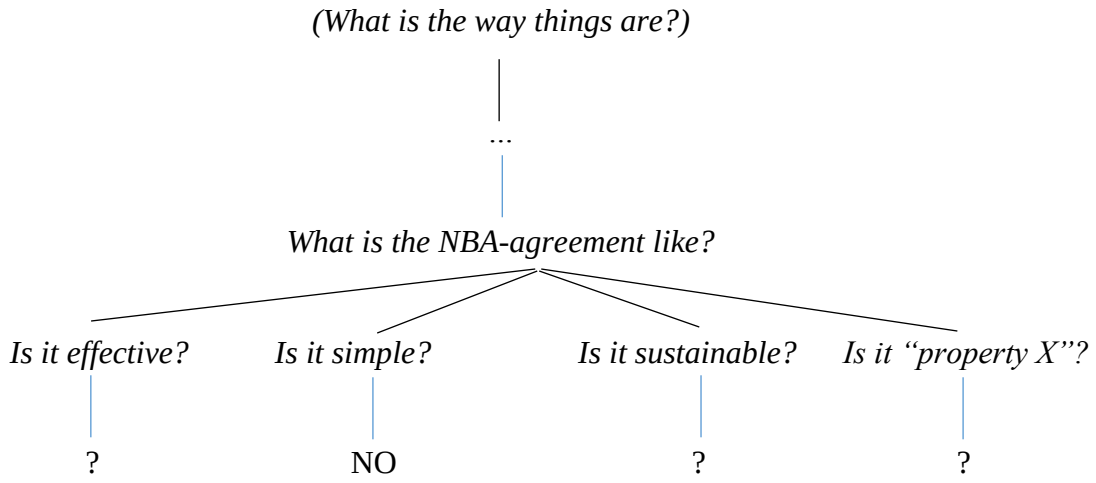

Figure 2.
Discourse-tree for (11c).

It may happen that the evoked alternative is not an opposing situation, but a lesser version of the event denoted by the predicate. Birner & Ward (1998) labels this as "scalar affirmation". This is an indication that the event denoted by the predicate happens to a large degree (occupies a high position on a hypothetical scale which measures the intensity of the event). One example that Birner & Ward (1998) mention is (13), which expresses that riding took place not just in an ordinary manner.

(13) *Asked what he thought about during today's race on a sultry day, [Tour de France winner Greg LeMond] said: "I didn't think. I just rode." Ride he did.*

I must add that Birner & Ward (1998) actually argues against the contrastive nature of TOP-Eng. However, they point their refutation against Chafe's (1976) conception of contrast, which is "an assertion on the part of a speaker that one of 'a limited number of candidates' is 'correct'". Birner & Ward (1998) points out that there are many instances of TOP-Eng, where an

interpretation resting on a "correct selection" is implausible. However, my notion of contrast is less specific than Chafe's, mine involves the evocation of discourse-salient alternatives (which are modelled as being linked to subquestions). Thus, it is possible for me to subscribe to the contrastive interpretation of TOP-Eng, even in the light of Birner & Ward's (1998) criticism towards Chafe (1976).

The claim that Topicalization marks contrastive categories also sheds some light on why it can be used with nonreferential expressions, demonstrated in (2d-e). As discussed in 2.6.2.1, the referentiality requirements on Contrastive topics are much looser than the ones on regular topics.

In conclusion, Topicalization behaves as a regular long-distance dependency, marking contrastive information-structural categories.

4.1.2 Analysis of TOP-Eng

Let us summarize the properties of TOP:

- the fronted element may be an argument or a predicate
- the functional and categorial status of the fronted element is flexible
- the fronted element maintains strong ties to its canonical position (seen in island- and binding-data)
- at most one instance of TOP is possible
- the fronted element is interpreted contrastively at information-structure

The analysis I advocate is based on Dalrymple (2001, Chapter 14). Since at most one phrase can be topicalized, a unique position is needed for such elements. Since topicalized entities follow complementizers and precede subjects in complex sentences as (14) attests, a natural place for them is an IP-adjoined position.

(14) *John said that Mary, he likes.*

TOP's sensitivity to islands and its "reconstruction" with respect to binding proves that a strong link with the canonical position should be maintained. This strong link has been modelled as movement in Chomskyan frameworks: the topicalized phrase is base-generated clause-internally and it undergoes A-bar movement to some left-peripheral position. LFG's functional identification serves essentially the same purpose. Functional identification immediately

accounts for the binding data, as the fronted element is fully present in the embedded f-structure as well.

Based on the data and such considerations, the following ID-rules are needed for TOP:

(15) IP → XP IP
$\downarrow \in (\uparrow \text{UDF})$
$(\uparrow \text{TOPPATH}) = \downarrow$
$\downarrow_i[\text{D-STR} = +]$
$\downarrow_i[\text{CONTRASTIVE} = +]$

A couple of notes are in order regarding (15). The XP node stands for the variety of grammatical categories that may participate in TOP-Eng (NP, DP, VP, AP, PP, AdvP). TOPPATH is a shorthand for the possible path of identification. Dalrymple (2001:396) defines it as follows:

(16) $\text{TOPPATH} \equiv \{\text{XCOMP} \mid \text{COMP} \mid \underset{(\rightarrow \text{LDD} \neq -)}{\text{OBJ}}\}^* \{(\text{ADJ} \in) \underset{(\rightarrow \text{TENSE})}{(\text{GF})} \mid \underset{\neg(\rightarrow \text{TENSE})}{(\text{GF})}\}$

In the words of Dalrymple (2001:396), "this expression allows the within-clause grammatical function of the UDF to be arbitrarily deeply embedded inside any number of properly constrained XCOMP, COMP, or OBJ functions, and optionally to appear as an untensed member of the ADJ set of such a function, or as an argument of the ADJ." So this specification covers all cases of argument and predicate fronting, as well as those cases of clause-initial adjuncts where the adjunct is interpreted contrastively (TOP with an adjunct).

In the third line of (15) we find a number of constraints about the possible path to the original position of the topicalized element. This is the TOPPATH IN (16). The (→LDD ≠ –) notation is meant to capture the fact that nonbridge verbs block the association of the topicalized phrase with its canonical position. It should be decoded as: "the path does not contain an element which is negatively specified for the LDD (long-distance dependency) feature" (nonbridge verbs are negatively specified in this respect as a part of their lexical entry). The (→TENSE) notation means that the OBJ function that the TOPPATH goes through must have a tense value. The function of this is to ensure that those verbs that realize their CP as an OBJ can also participate in TOP. This is needed because CPs may actually bear either the COMP function or they may be OBJs, as mentioned in section 2.4 (see also footnote 4, section 2.4). Such objects have a tense value. They contrast with objects acting as heads of complex noun phrases, which lack such a value. These are islands and cannot participate in TOP. Finally, ¬(→TENSE) makes sure that the topicalized entity can appear as a grammatical function only

inside an untensed ADJUNCT function (recall the contrast between 5c and 6). These constraints block island-violations. The annotation about information structure ensures that the topicalized phrase is interpreted as one of the contrastive categories, CT or CF.

As noted earlier, the arrows in the annotations are "metavariables". ↑ referring to the mother-node's functional structure, while ↓ referring to the node's f-structure. So ↓∈ (↑UDF) in (15) should be read as "this node is a UDF in the functional structure of the IP". As an example, Figure 1a and 1b show the constituent- and functional structure of a sentence like (1a) (*John, I like*). For brevity, I only annotate the parts of the c-structure that are directly relevant for the current issues.

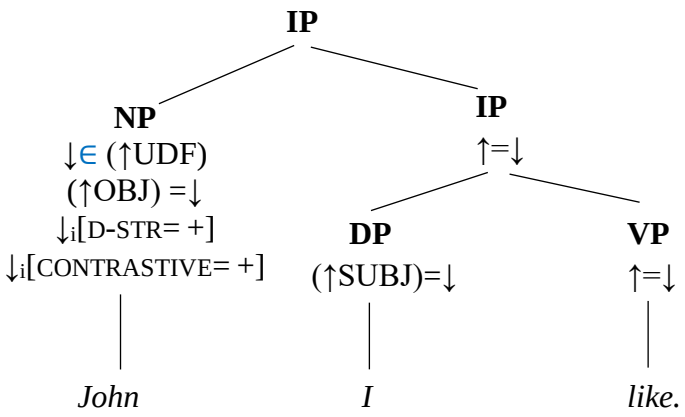

Figure 1a.
Constituent-structure of (1b).

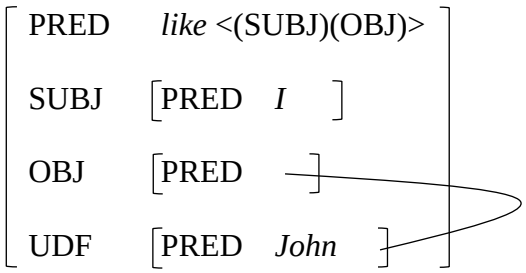

Figure 1b.
Functional structure of (1a).

What is not clear at this point is what causes the uniqueness of this IP-adjoined position. That only one functionally identified element may be IP-adjoined seems to be part of some wider conspiracy with parametric variation in languages, see Engels (2012:171-176), who gives an Optimality-theoretic[55] account of the data. Such ambitions are beyond the scope of this dissertation so I leave this aspect of the research for future investigations.

[55] See Prince & Smolensky (1993).

4.2 Clause-initial adjuncts in English

While TOP-Eng may involve adjuncts, in this section I will show that clause-initial adjunct of the kind shown in (1b) may participate in a different construction as well. A deeper investigation to be presented here will reveal several crucial differences, as regards both syntax and information-structure.

It should be taken into consideration that that the class of adjuncts is huge, and they do not behave in a homogenous manner. I will restrict this discussion to those aspects of such adjuncts that are directly relevant for the purposes of this dissertation and put many fine details aside. For such details, the interested reader is referred to the references mentioned in this section.

4.2.1 Properties of CIADJs-Eng

From a syntactic point of view, the most important distinction that we have to make is that adjuncts differ according to what we might call as their canonical position. While certain adjuncts are "base-generated" at the left periphery and have no ties to any clause-internal position, other adjuncts maintain ties to such positions (Frey 2003). This is mirrored in what part of the event expressed by the sentence these adjuncts modify. Following Frey (2003), I will refer to the first type of adjuncts as "frame/event external adjuncts", while I will call the second type "event-internal adjuncts". As an illustration, consider (17).

(17) a *In New York, there is always something to do.*
b *These days it rains a lot.*
(18) a *In the box, John found a hammer.*
b *With his metal detector, John found some interesting items in the garden.*

In (17a), *in New York* and in (17b), *these days* provide a spatial/temporal frame for the event. That is, they do not modify the contents of the event, but restrict the domain for which the claim holds. On the other hand, *in the box* in (18a) specifies the place of the hammer, *with a metal detector* in (18b) specifies the tool of the process. These provide more information about some internal part of the event.

(18) shows that the event external-event internal distinction is not based on grammatical category or semantic type (as both *in New York* and *in the box* are locative PPs), but on interpretation. As we will see, this distinction has consequences for the syntax and information

structure of these adjuncts. Event-internal adjuncts show syntactic properties that liken them to TOP-Eng: they are sensitive to island-constraints (19a) and show principle C effects (19b).

(19) a **In the box, we heard the claim that John had found a hammer.*

b **In Ben$_i$'s office, he$_i$ lay on the desk.*

Haegeman (2003) and Frey (2003) also demonstrates that CIADJ-Engs are interpreted at a distance ("long adverbial fronting" in Haegeman's words) pattern with TOP-Engs (argument fronting) with respect to a number of syntactic phenomena. For instance, an event-external adjunct may alleviate a "*that*-trace" effect (20a), but an event-internal (20b) may not.

(20) a *Who do you think that yesterday had a great time?*

b **Who do you think that in the box found a hammer?*

These CIADJs also seem to follow TOP with respect to their distribution: only one may be present and they cannot co-occur with TOP, as in (21), see also Engels (2012:172).

(21) a **In the box, with a flashlight John found a hammer.*

b **[In the box/With a flashlight], a hammer *[in the box/with a flashlight], John found.*

Frame/event external adjuncts show a dissimilar behavior in these respects. In (22) we find no principle C violation (19b and 22 are from Frey 2003).

(22) *On Ben$_i$'s birthday, he$_i$ took it easy.*

Unlike an event-internal CIADJ, these adjuncts may co-occur with TOP, either preceding or following it (23).

(23) a *Probably, Linda, you have met.*

b *Linda, probably, you have met.*

Regarding information-structure, we first should note that any clause-initial adjunct may, but does not have to be interpreted contrastively. Apart from some cases, CIADJs are not necessarily interpreted as contrastive. So sentences in (17) and (18) may be uttered without

evoking alternatives in the discourse. So (17a) does not necessarily suggest that there are other places where one cannot do anything or (18a) does not necessarily implicate that there is another container where John found something other than a hammer. This is a critical difference in comparison with TOP.

It should be added however that there are CIADJs which only allow the contrastive reading. For example, there are some adjunct-types that highly disprefer a neutral topic interpretation. For example, manner or measure adjuncts are often ungrammatical clause-initially. Example (24a) is from Ernst (2002). The sentence improves if the adverb receives "strong contrastive stress" (Ernst 2002: 470), that is, if it is an instance of TOP with an adjunct (24b).

(24) a **Tightly, she must hold on to the railing.*

b *TIGHTLY she must hold on to the railing.*

I think the reason for this must be semantico-pragmatic rather than syntactic in nature. In particular, following Shaer (2004:388), we can assume that "the degraded acceptability of a sentence containing a fronted adverbial may have its source in the difficulty of inferring the relation of the adverbial not only to its host sentence but also to previous discourse." That is to say, although the CIADJ-construction would assign a Topic interpretation to these adjuncts, they cannot assume this role because they are non-referential. The +CONTRASTIVE feature eases the referentiality requirement, as noted earlier.

Returning to referential CIADJs like (17) and (18), I follow Maienborn (2001) and Ernst (2002:399-402) in that these cases of CIADJs are topic-like entities. In (17a) *in New York*, as a frame-setter introduces a discourse topic, in a general discussion about cities. Similarly, *these days* in (17b) introduces a temporal frame. (18a-b) could be imagined in the context of a story, so the box or the metal detector are most likely the links to the previous discourse. Accordingly, it is plausibly viewed as an entity introducing a new subtopic. It follows that the context should be one in which the hearer can accommodate such a new subtopic, like (25a). (25a) also illustrates that the CIADJ is not necessarily contrastive, as no pragmatic tension can be detected when no alternative is provided. If the context is not one in which the *box* is a plausible link to the discourse, the sentence is weird. In (25b), even though it is plausible that one may find a box in a cellar, it is still hard to establish the *box* as a plausible subtopic (even if the definite article is replaced with an indefinite one to neutralize the identifiability-effect of *the*).

(25) a *John found several containers in the cellar: a box, a bag and a jar. He wanted to see what's inside. In the box, he found a hammer. (…) In the bag, there was another hammer.*

b *John went into the cellar. #In a box, he found a hammer.*

Additionally, just like the Hungarian "topic-position" (as in Figure 1 in section 3.1), sentential adverbs may also occupy this slot, as the sentential adverb *probably* does so in (23). Following the remarks made in 3.1, I classify such cases as Completive or Background information.

4.2.2 Analysis of CIADJs-Eng

As the contrastive uses of Clause-initial adjuncts are covered by the analysis provided for TOP, this section is only concerned with non-contrastive CIADJs. It has been established earlier that CIADJs may be divided into at least two subcategories as far as their syntax is concerned. Frame-setters/event external adjuncts fully belong to the left periphery, while event internal adjuncts maintain ties to their clause-internal interpretational site.

Let us first approach the event-internal type. An alternative IP-adjoined position with annotations is provided for the event-internal type in (26) and (27).

(26) IP → AdvP/PP IP
↓∈(↑UDF)
↓∈(↑ADJPATH)
↓$_i$ [NEW= –]
↓$_i$ [D-STR= +]

(27) ADJPATH ≡ {XCOMP | COMP (→LDD ≠ –) | OBJ (→TENSE) }* ADJ

Here the phrase-structural category is restricted to AdvP and PP, in order to prevent inappropriate categories from being parsed as adjuncts (e.g. nominals in LD). The ADJPATH is a modified version of the TOPPATH: it encodes similar restrictions (to capture the similar syntactic behavior of TOP and event-internal CIADJs) but is restricted to adjuncts. So the fronted adjunct may be embedded at an arbitrary depth, through XCOMPs, COMPs, and tensed OBJ functions.

The one major modification is that there is no disjunction in the second half of the notation. This is because although extracting an argument out of an adjunct is possible (as in 6), extracting an adjunct from an adjunct is ungrammatical. (28 is to be construed as *on the third floor* being an adjunct for *that room*.)

(28) **On the third floor, Chris teaches his classes in that room.*

If this configuration is not possible for the adjunct for distance or semantico-pragmatic reasons (as in 24a), they may be analyzed using the TOP annotation, introduced earlier, where the +CONTRASTIVE feature makes non-referential entities acceptable.

Another difference is that these CIADJs are specified at information-structure as having the feature-specification of Topics: –NEW, +D-STRUCTURING.

Frame setter/event external adjuncts behave differently: as we saw earlier, they do not "reconstruct", see the data in (22). This suggests that the integration of such adjuncts into the sentence is generally looser. According to Minimalist approaches, such entities are "base-generated" in some high position of the syntactic tree. This high position could be some CP- or IP-adjoined position, or in a designated TopicP, depending on the details of the analysis.

There is also an alternative approach. Shaer (2004), building on the ideas of Haegeman (1991), claims that such clause-initial adjuncts are "syntactic orphans": they are not integrated into the phrase structure. In other words, they are not proper parts of the sentence.

One of the main arguments of Shaer (2004) is that certain polarity items like *ever* in (29a) are not permitted even if a supposedly base-generated CIADJ would in principle license them (*ever* needs to be under the scope of *only*). If the CIADJ is undoubtedly part of the sentence, as in (29b), the polarity item is grammatical. One might conclude that inversion itself is the critical factor in the licensing of polarity items like *ever*, but heavy contrastive stress (as in Topicalization) may also mitigate ungrammaticality to a certain extent, see (29c).

(29) a **Only in New York, John could ever have fun.*
b *Only in New York could John ever have fun.*
c *?ONLY IN NEW YORK, John could ever have fun.*

However, Engels (2012), referencing Haegeman (1995) points out that the facts in (29) may have alternative explanations, e.g. the non-projection of the relevant feature of *only* outside the PP. Whatever the exact formulation of the constraints are, the complete exclusion of these adjuncts from the f-structure of the sentences would make their semantic association with the sentence problematic, so I side with the more conservative approach, where they are proper parts of the sentence.

In light of the discussion, I propose that the following annotation is possible for such CIADJs.

(30) IP → AdvP/PP IP
↓∈(↑ADJ)
{↓$_i$ [NEW= –]
↓$_i$ [D-STR= +] |
(↓ADV-TYPE)=$_c$ SENT
↓$_i$ [D-STR= –]}

These adjuncts are not extracted: they modify at the sentence-level so they are not identified with any clause-internal function. At information-structure, they are either Topics (in case of referential expressions) or Completive/Background information, in the case of sentence-adverbs (*probably*, *luckily*, etc.).

The analysis of a sample sentence with a frame-setting and an event-internal adjunct is shown below.[56]

(31) *Yesterday, for an hour I was looking for a taxi.*

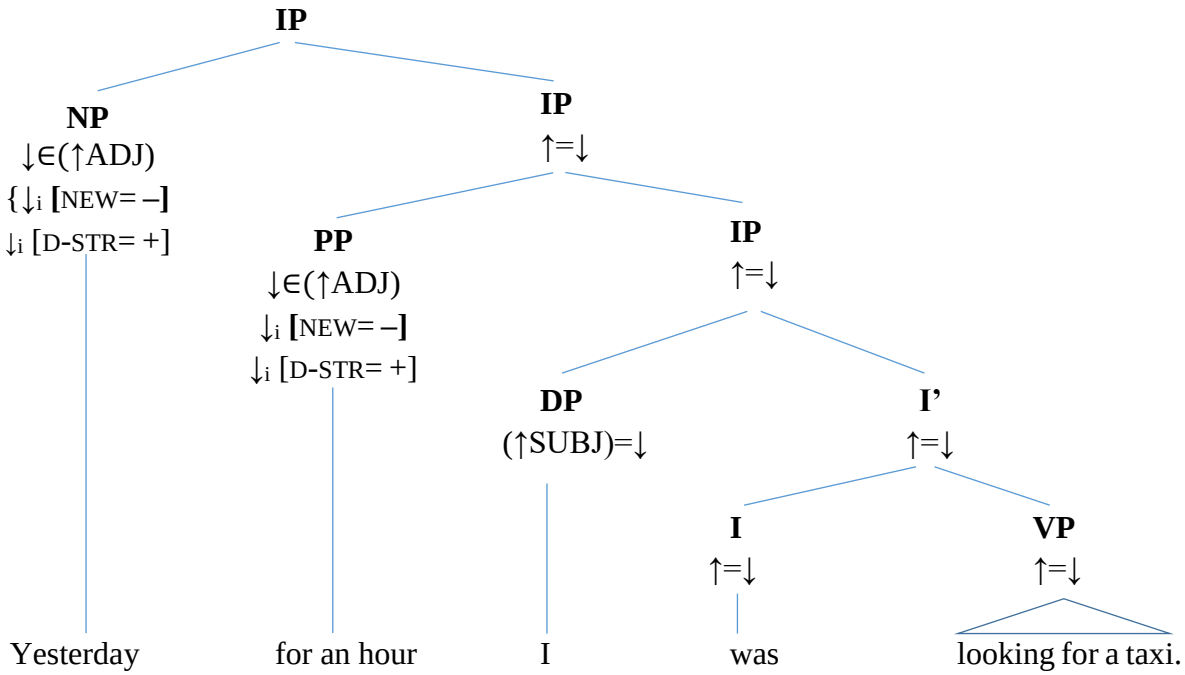

Figure 2a.
Constituent-structure of (31).

[56] Although I take the ordering of adjuncts to be free as far as syntax is concerned, other factors such as semantic or scopal considerations do impose constraints, see Ernst (2002), Engels (2012:177-188).

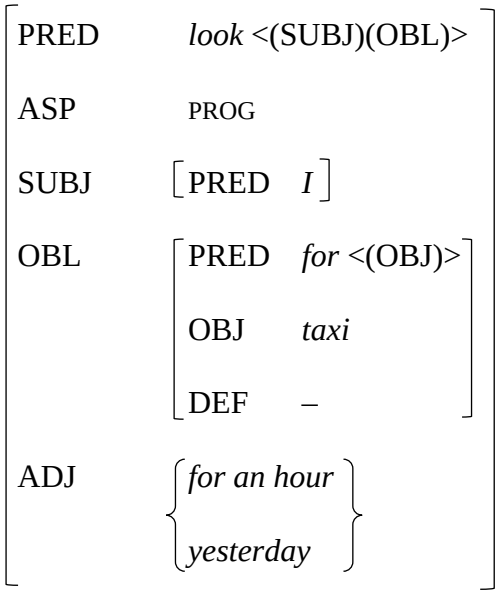

Figure 2b.

Functional-structure of (31).

4.3 Left-dislocation in English

The term "Left-dislocation" (LD) was first used by Ross (1967) to label English sentences where there is some element at the left periphery of the sentence (I will descriptively refer to these as the "left dislocated" elements) and also there is a co-referent resumptive pronoun later in the clause. As we have illustrated in (1c) (*John$_i$, I like him$_i$*), left-dislocation minimally differs from TOP in the fact that here we usually find a pronoun in the canonical position of the fronted constituent. However, just like in the case of CIADJs, if we look deeper, a wide range of other differences surface.

4.3.1 Properties of LD-Eng

First, in contrast with TOP, the fronted element in LD-Eng must be a referential expression. For this reason, the examples with a fronted VP, AP and CP in (32) are only marginally acceptable, as opposed to the natural-sounding examples in (2).

(32) a *???Surrender, we will never do so.*

b *???Happy, Tom will never be like that.*

c *???That Tom was a movie star, we would have never guessed that.*

From this it also follows that quantified or indefinite nominals are quite degraded in LD-Eng, as already noted in Rodman (1974). Rodman actually marks the examples in (33) with a *, but Shaer (2009) shows that some examples of this sort may be found in naturally occurring texts, e.g. see (34). Still, the phenomenon seems to be quite restricted.

(33) a ???*A boy, I saw him.*

b ???*Someone, he is coming.*

c ???*Everybody, they are doing it*.

(34) *And most folks, they don't seem like they notice that you've answered that same question 400 times that hour.*

Secondly, the effects related to island-constraints that were observed with TOP-Eng are absent in the case of LD-Eng. It is clear that syntactic factors are irrelevant for the dependency between the left-dislocated element and its pronominal counterpart.

(35) a *John, I doubt the claim that you like him*. (CNPC)

b *John, I can't guess why you like him.* (wh-island)

c *John, we think that David laughed when we selected him.* (adjunct-island)

d *John, that you saw him surprised me.* (subject island)

e *John, I whispered that I saw him.* (nonbridge verb)

Thirdly, the binding-patterns are also different. The fronted reflexive (36a) and the pronoun in (36b) remains unbound, which leads to ungrammaticality for the former case. This pattern is the opposite of what was seen in (7) earlier with TOP-Eng. This suggests that there are no "reconstruction"-effects in LD-Eng.

(36) a *??Pictures of himself$_i$, John$_i$ likes them.* (Principle A)

b *Pictures of him$_i$, John$_i$ likes them.* (Principle B)

That sentences like (36a) are not entirely unacceptable could be a result of some poorly-understood processes that make variable binding possible even when the necessary syntactic configurations do not hold. Actually, such claims have also been made in connection with LD-Eng, see e.g. Vat (1981), who reports that (37a) is not entirely ruled out. See also (37b), a sentence without any left-dislocation, from Shaer & Frey (2003). Under a strict syntactic view of binding, the sentences in (37) should be ungrammatical.

(37) a *?His$_i$ first article, I think [every linguist]$_i$ would consider it a failure.*

b *?When his$_i$ boss is happy, [every office-worker]$_i$ is happy too.*

Lastly, while in TOP-Eng, the case of the fronted constituent is always the one that it would get in its canonical position (38a), in LD-Eng, the two may be different (38b).

(38) a *Me/*I, John likes.*

b *Me, I like beer.*

In principle, the connection between the left-dislocated element and the pronoun can be thought of as one involving syntactic or mere pragmatic relationship. I think that the latter option seems more likely though, since one can find examples where, given the proper context, there is no pronoun present in the construction (these are "unlinked topics"). The following examples are from Lambrecht (2001).

(39) a *Tulips, do you have to plant new bulbs every year?*

b *Austin, at least you can sit near the AC.*

The relation between LD-Eng and other fronting constructions in a single sentence is not clear-cut. According to Grohmann (2003) LD-Eng always has to precede TOP, as in (40). Shaer (2009), on the basis of examples like (41), debates this.

(40) a *Mary, John, she likes.*

b **Mary, John, he likes.*

(41) a *Now this junk, my father, he was always collecting.*

b *Now my father, this junk he was always collecting.*

Another debated issue is to what extent LD-Eng can be regarded as a root-clause phenomenon. According to most standard accounts, LD-Eng is ungrammatical in subordinate clauses, but once again, the opposite claim is also present in the literature. Note that there is no debate about the embeddability of TOP or CIADJ (42b-c).

(42) a *?John said that Mary, he likes her.*

b *John said that Mary, he likes.*

c *John said that in New York, there is always something to do.*

Related to this is the fact that while TOP and CIADJ allow syntactic dependencies to arch over them (43c-d)[57], LD-Eng blocks such phenomena (43a-b). This is probably partly because of the just mentioned fact that LD-Eng is somewhat marked in embedded positions. The long-distance dependency exacerbates the already dispreferred configuration.

(43) a **How would you say that Robin, he would solve this problem?*

b **This is a problem that Robin, he wouldn't solve.*

c *How do you think that this problem, Robin would solve?*

d *How do you think that in New York, Robin would behave?*

These last three properties (ordering, embeddability, intervening) all seem to point to one direction: LD-Eng is a more peripheral structure in some sense than TOP and CIADJs. As such it always has reduced grammaticality when the structure occupies a sentence-internal position.

The main syntactic properties having been surveyed, let us move to the information-structural properties of left-dislocation. Prince (1998) claims that there are three basic functions for LD-Eng: island-amnesty, simplifying discourse processing, signaling a "poset-inference"

In the first use, it is actually applied as covert Topicalization. The speaker would like to use TOP, but faces a syntactic obstacle, e.g. an island, and thus is forced to put a resumptive pronoun in the canonical position of the initial element. One such example is shown in (44).

(44) *Tom, the story about *(him) was funny.*

As such uses are clearly forced by core syntax and have nothing to do with information-structure, I exclude them from the scope of this section.

The second function of LD-Eng is "simplifying discourse processing." According to Prince (1998) this means that by using LD-Eng, people remove discourse-new entities from positions that are dispreferred for them. Prince's (1998) example for this is the following segment:

(45) *My sister got stabbed. She died. Two of my sisters were living together on 18th Street. They had gone to bed, and this man, their girlfriend's husband, came in. He started fussing with my sister and she started to scream. <u>The landlady, she</u> went up, and he laid her out. So sister went to get a wash cloth to put on her, he stabbed her in the back.*

[57] This is not entirely unrestricted though, see Browning (1996:252-253).

According to Prince (1998), *the landlady* in its original position would be a subject and subjects are generally dispreferred as discourse-new entities. One can also approach this from the perspective of Lambrecht's (1994:185) "Principle of the separation of reference and role": do not introduce a referent and talk about it in the same clause.

On the axis of referential givenness, it could also be added that such Left-dislocated elements tend to occur if the topic is shifted, that's is, when a not yet topical entity is promoted to the topic-status. This is also the conclusion of Frey (2005), who shows that LD-Eng cannot be used to simply continue a given topic, it must "break the discourse continuity and start a new discourse unit" (Frey 2005:23). See (46):

(46) *I heard some news about John. #John, Mary kissed him.*

That LD-Eng can be used to introduce a brand-new entity can also be shown from the fact that it is conceivable that someone, looking for a particular *Tom,* enters a room and utters the following sentence, containing an LD-Eng:

(47) *Tom, where is he?*

The same could hardly be conceivable with TOP as the discourse-conditions for evoking contrast are much stricter. There must be a discourse question already under discussion, to which the contrast and the set of alternatives can be related.

The third use of LD-Eng according to Prince (1998) is to trigger an inference on the part of the hearer that the entity represented by the initial NP stands in a salient partially-ordered set relation to some entity or entities already evoked in the discourse-model. Partially ordered sets, "posets" are "defined by a partial ordering R on some set of entities, e, such that, for all e-1, e-2, and e-3 that are elements of e, R is either reflexive, transitive, and antisymmetric or, alternatively, irreflexive, transitive, and asymmetric" (Prince 1998). In essence, this means that the left-dislocated entity has some set relation with other elements.

Prince (1998) sees these functions as separate entities. However, subsequent research suggests that there may be a way to have a unified view of functions 2 and 3 (as was stated, the first function is set aside).

Gregory & Michaelis (2001), based on a corpus study on TOP and LD-Eng suggest that the overarching function of LD is that of "topic promotion", that is, to bring entities into the discourse. They have compared all the LD-Eng tokens with all the TOP tokens and have found three factors that back this claim up.

First, they examined the givenness of LD-Engs, compared to TOPs. They used Gundel, Hedberg & Zacharski's (1993) cognitive statuses to determine the referential givenness of an element in the discourse. The authors found that LD-Eng has relatively low givenness in the discourse, as opposed to TOP-Engs. The latter had higher activation status, which is expected if they are always contrasted to some discourse elements, as was established in the previous section. LD-Eng was used to bring new topics into the discourse, which is in accordance with its function of topic-promotion.

Gregory & Michaelis's (2001) second target for investigation was the anaphoricity of left-dislocated and topicalized entities. They categorized tokens according to the type of the anaphoric link that the fronted element had to the discourse (from highest to lowest): directly mentioned, the entity is member of a set that has been mentioned, none. They found that LD-Engs tended to have low anaphoricity, which is expected of a topic-promoter.

Gregory & Michaelis's (2001) final factor was topic persistence. They measured to what extent the fronted elements in LD-Eng and TOP tend to remain topics of the subsequent discourse. They found that LD-Eng has a high topic persistence, as opposed to TOP. This is in line with what we have discussed in connection with these structures: LD-Eng is a topic promoter, so one expects that the entity introduced by it is going to be talked about. We do not necessarily have such expectations for contrasted elements introduced by TOP.

Thus, it is plausible to claim that LD-Eng is a construction that is linked to the Topic information-structural category. It may have contrastive uses, as in Prince's (1998) poset use, but unlike TOP-Eng, this contrastivity is not compulsory, as can be clearly seen in (45).

The link of LD-Eng and topics in general is further supported by a corpus study by Snider & Zaenen (2006) who found that there is a positive correlation between LD-Eng and animacy. This is expected if LD-Eng is Topic-related, since cross-linguistic investigation often observes a connection between animacy and topichood (and topic-marking).[58]

4.3.2 Analysis of LD-Eng

Left-dislocation in English (LD-Eng) resembles frame-setting CIADJ in that it seems to be loosely integrated into the sentence. It is unaffected by islands, induces no Principle C effects as we have seen in the previous section. Based on this one might assume that an analysis like the one posited for CIADJ would be suitable for LD as well, where they are "base-generated" in a left-peripheral node, and are not functionally plugged into the internal part of the sentence.

[58] See Dalrymple & Nikolaeva (2011, chapter 1 and references therein) for an outline of correlation between animacy and topichood in the context of object-marking in a number of languages.

However, I think it is undesirable to collapse the analysis of CIADJ and LD-Eng. The reasons are the following.

While CIADJ are easily assigned the grammatical function ADJUNCT, it is not at all clear what GF should be assigned to the nominals that occur in LD-Eng. This makes a situation that is difficult to handle with LFG's functionally annotated phrase-structure rules. Assigning no GF is not a particularly appealing option given the central role of f-structure in LFG.

LD-Eng also differs from CIADJ in a number of ways syntax-wise. While CIADJs are flexibly placed around TOP, LD-Eng seems to be marked in post-TOP positions. Although the data is not entirely straightforward, it does contrast with CIADJ, where there is no doubt about the positional flexibility. The situation is similar with respect to embeddability. The embeddability of LD is questionable, while CIADJ (and TOP) are fully licensed in subordinate clauses. It seems that LD is "radically" left-peripheral. From this perspective it seems plausible that LD-Eng may block or disrupt syntactic dependencies reaching over it.

What this suggests is that LD-Eng is even less integrated than CIADJ. In my opinion, this is because the fronted elements in LD-Eng are "intruders" in the phonetic string of the sentence. They are unattached to the syntax of the containing sentence at any syntactic level, they function as a fully independent entity in relation to the host. As such, placed sentence-internally they disrupt the inner structure of the sentence that they are phonetically part of. Essentially they are syntactic orphans in the sense of Haegeman (1991) and Shaer (2009).

This approach is feasible in LFG as the phonetic string in itself could be regarded as a representational level in LFG. It is the acoustic data, which the function π maps to the constituent-structure (the level where constituency is represented) and the function β maps to the prosodic structure (where intonation, stress, etc. are represented, see Mycock & Lowe 2013). As LD-Eng-constituents are independent entities in the host string, they and the matrix sentence are mapped from the phonetic string to entirely different c- and f-structures.

It is true, the f-structure of left-dislocated element would be rather fragmentary. *John* in (1c) (repeated here as 48a) would most probably only have a bare PRED feature, but some sort of analysis for such fragments is independently needed for constructions like vocatives or elliptic answers.

(48) a *John, I like him.*

b *Hey, John, what do you see?*

c A: *What may I serve you?*
B: *Tea.*

Any agreement-like phenomena like number matching in (42a) follow from general discourse-considerations, like pronoun-antecedent relationship. Also, since there is no syntactic link between the fronted element and any sentence-internal correlate, case-mismatches of the sort illustrated in (38) are not unexpected.

Reconstruction effects are also expected to be absent, as in (36) and (37). That some reconstruction may still be possible could be the result of some poorly-understood processes at play. Shaer & Frey (2003:476) speculate that what happens is "semantic subordination of the expression containing the variable to the expression containing the quantificational expression".

As these independent elements may project to information-structure in their own right, they may still assume the Topic role that they usually have in discourse. Also, they may be referred back to by a pronoun in the host sentence, to form an explicit link, but this is not mandatory, in the case of unlinked topics (39), the relationship is mere pragmatic inference. Consequently, no syntactic formalization is necessary for the analysis of LD-Eng.

4.4 Conclusion of chapter 4

In this chapter I presented an LFG-based account of the three clause-initial, discourse-related constructions of English: Topicalization, Clause-initial adjuncts and Left-dislocation. Of the three, TOP-Eng is the most integrated structure: it is a functional dependency, marking contrastive topics/foci at information-structure. Event-internal CIADJs are integrated like TOP-Eng, but these are not necessarily contrastive. Other, frame setting CIADJs are not tied to sentence-internal positions, they fully belong to the left-periphery as Topics. Finally, LD-Eng, also a Topic-marking device at information-structure is a "syntactic orphan", it is not parsed as a part of the host-sentence. Consequently, it is radically extra-sentential, barred from clause-internal positions.

As result of the discussion in this chapter, the following phrase structure of the English left-periphery emerges. As indicated throughout the chapter, I take the positions of these fronted elements to be IP-adjunctions. Frame-setting/event external adjuncts may occur either before or after topicalized phrases. These phrases, TOP-Engs share a node with event-internal CIADJs. They are integrated via versions of a functional identificational equation. Also, their i-structural specification differs, requiring only TOP-Eng to be contrastive.

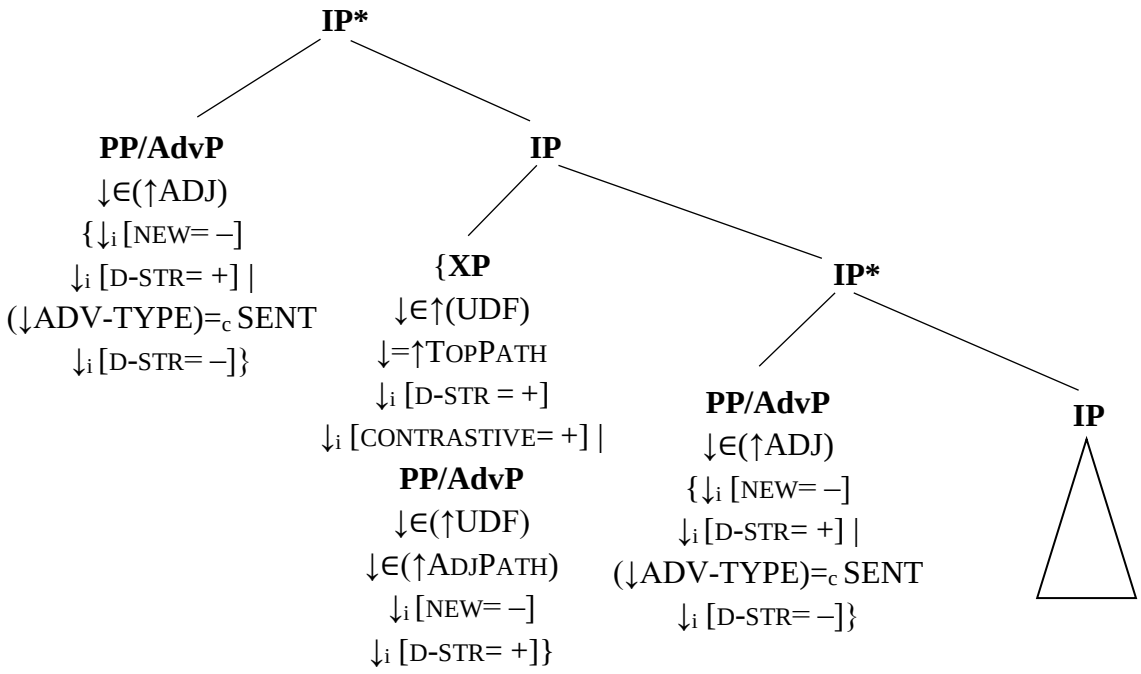

Figure 3.
The English left periphery.

CHAPTER 5

CONSTRUCTIONS ON THE HUNGARIAN LEFT PERIPHERY

In this section we turn our attention to Hungarian and investigate some left-peripheral constructions in this language. In harmony with the goals of the thesis, each construction has repercussions for the information-structure of the sentence.

The first part of this chapter is about configurations that are reminiscent of left-dislocation structures in English. I follow the literature in calling this structures "left-dislocation". However, it should be born in mind that I do not want to suggest that the analysis of these Hungarian sentence-types is parallel to that of LD-Eng. In fact, as we will see, the Hungarian LD actually has several subtypes, each of which should receive its own analysis and only one of these analyses is akin to the analysis of English LD.

The second part of the chapter focuses on a construction which has already been mentioned briefly in 3.2, Operator fronting (OF). OF also falls into distinct subtypes, with one that resembles TOP-Eng in that it behaves as a long-distance, "movement"-like dependency. The other pattern has no counterpart in the English constructions that have been surveyed. That version of OF is a long-distance anaphoric dependency, called "prolepsis" and it will be argued to be sharing properties with what is called "control" in syntactic theory.

5.1 Left-dislocation in Hungarian

In the previous section, I have used the term "Left-dislocation" for those English constructions where there is some clause-initial nominal entity that is often co-referent with some clause-internal pronoun.

The Hungarian constructions that we are to investigate also show this pattern, which justifies the label for it, although note that while the clause-internal pronominal counterpart was argued to be optional in LF-Eng, the Hungarian constructions (LD-Hun) are defined on the basis of the pronoun, so we will not have optionality here.

LD-Hun[59] as a term has actually been used for three distinct structures in the literature. These, as we will see, display different syntactic and information-structural properties. They are going to be labelled as follows:

- Contrastive topic left dislocation (CTLD-Hun)

  (1) *Jánost, azt Kati szereti.*
  John.ACC that.ACC Kate likes
  'As for John, Kate likes him.'

- Non-contrastive left dislocation (NCLD-Hun)

  (2) *Erre János az fogta magát, és elszaladt.*
  then John that took.3SG himself.ACC and away.ran.3SG
  'Then John, he ran away.'

- Contrastive focus left dislocation (CFLD-Hun)

  (3) *A könyvet, AZT olvastam a szobában (és nem az újságot).*
  the book.ACC that.ACC read.1SG the room.in and not the newspaper.ACC
  'The book, I read THAT in the room (and not the newspaper).'

Of these types, CTLD-Hun was investigated in Lipták (2011), NCLD-Hun in Lipták (2012) in a minimalist setting. She analyzes both of them as the dislocated element and the pronoun are base generated as sisters in a VP-internal position and are subsequently moved into a clause initial, specific projection (ContTopP and TopP respectively). Den Dikken and Surányi (to appear) also discuss CTLD-Hun, from the perspective of possible Minimalist analyses. CFLD-Hun has not been a subject of research, it has only been mentioned by Kenesei et al. (1998). A systematic survey of LD-Hun has not been undertaken.

It is a common property of all three LD-Huns illustrated in (1)-(3) that they can only occur at the left periphery. The presence of the resumptive pronoun at the right edge of the clause causes ungrammaticality.

(4) **Kati szereti [azt, Jánost/ Jánost, azt].*
Kate likes that.ACC John.ACC John.ACC that.ACC

In the following three sections, I survey the properties of the subtypes of LD-Hun and then in 5.1.4 I will present my analysis for them.

[59] I would like to express my gratitude to Zsuzsanna Gécseg, who helped me with a number of valuable comments as I was researching LD-Hun.

5.1.1 Properties of CTLD-Hun

In (1), the left-dislocated phrase is *János* (‘John’) the subject matter of *Kati*’s love, and the resumptive pronoun is the demonstrative *azt (*‘that.ACC’). It has been noted that due to prescriptivist considerations, some speakers may have reservations about referring to a person with a demonstrative pronoun and prefer to use a personal pronoun in such LD-Huns.

(5) *Jánost, őt Kati szereti.*
John.ACC he.ACC Kate likes
‘As for John, Kate likes him.’

The natural interpretation of the sentence is where *János* is a contrastive topic, so (1) and (5) indicate that for other people, a different state of affairs holds (for example, as for Peter, Ann loves him). Hence the name CTLD-Hun. The intonation is accordingly the fall-rise one, typical of contrastive topics. This is not a very surprising state of affairs given that the possible presence of a demonstrative is a diagnostic of contrastive topics (Kálmán 2001:41). This is reinforced by the fact that the elements that are incompatible with contrastive topics for semantic reasons remain equally ungrammatical in the CTLD-Hun-constructions as well. (6) is semantically anomalous because if more than six students did attend the talk, it is not possible that less than six did not (for details, see Gyuris 2009).

(6) *#Hatnál több diák, az eljött az előadásra.*
six.by more student that came the lecture.to
‘#As for more than six students, they did attend the lecture.’

Conversely, in CTLD-Hun, non-referential elements are grammatical. This would not be the case if the fronted element was a regular topic, as the referentiality-constraints are eased in the presence of contrastiveness.

(7) a *Állatot, azt nem tartok.*
animal.ACC that.ACC not keep.1SG
‘As for animals, I don’t keep one.’

b *Kevés fiút, azt Mari hívott meg.*
few boys.ACC that.ACC Mary invited.3SG PV
Mary invited few boys (contrasted to e.g. John, who invited many).

In our target structure, where the resumptive pronoun is present, its adjacency to the host element is a matter of variation among speakers. Lipták (2011) claims that a strict adjacency is mandatory (so 8 is ought to be ungrammatical), while Baloghné Nagy (2013) finds (8) acceptable. Interestingly, Lipták (2012) contrasts with her earlier account, marking such structures grammatical.

(8) *?Jánost, tegnap azt mindenki látta a koncerten.*
John.ACC yesterday that.ACC everyone saw.3SG the concert.on
intended: 'As for John, everyone saw him at the concert.'

The form of the resumptive is fairly restricted. It must be a demonstrative or a personal pronoun and it shows connectivity effects with the left-dislocated constituent: they must have identical case-, number- and deictical features.

(9) a **Jánost, az Kati szereti.*
John.ACC that.NOM Kate likes.3SG

b *A lányokat, *azt/ azokat tegnap láttam.*
the girls.ACC that.ACC those.ACC yesterday saw.1SG
'As for the girls, I saw them yesterday.'

c *Azt a fiút, *ezt/ azt tegnap láttam.*
that.ACC the boy.ACC this.ACC that.ACC yesterday saw.1SG
'As for that boy, I saw him yesterday.'

However, the situation is not as simple as it may seem from (9). One must take care and not just blindly follow the case-marking on the left-dislocated entity when producing CTLD-Hun. For example, if the left-dislocated entity is an adverb, the case-marked forms of *ez*/*az* ('this'/'that') cannot refer back to that, since they can only pick nominal antecedents. The syntactically and semantically suitable pro-form, *úgy* ('so') has to be used.

(10) *Szép-en, *az-on/ úgy csak Kati dolgozik.*
nice-on that-on so only Kate works
'(As for) nicely, only Kate works like that.'

As a general rule, it can be established that those pro-forms should be picked as resumptives in CTLD-Hun that are used to refer to the respective types of entities in discourse, independent of CTLD-Hun use. In other words, CTLD-Hun involves a genuine anaphoric dependence. The syntactic constraints are imposed on the top of these semantic/pragmatic considerations.

This may be exposed clearly if we left-dislocate place names. Often there are two ways to refer to such entities: the appropriately case-marked form of *ez*/*az* (‘this’/’that’), or a special spatial pro-form. Accordingly, we can use either of these forms in CTLD-Hun, see (11).

(11) a *A házban, abban/ ott senki nem volt.*
the house.in that.in there nobody not was
’As for the house, there was nobody.’

b *A kertbe, abba/ oda gyakran kiment.*
the garden.to that.to there often out.went.3SG
’As for the garden, he often went out there.’

c *A konyhából, abból/ onnan gyakran jöttek jó illatok.*
the kitchen.from that.from from.there often came good scents
’As for the kitchen, good scents often came from there.’

However, when referring to certain types of places (e.g. cities), the case-marked forms cannot be used, only the spatial pronoun is suitable. An example can be seen in (12), where although the city itself bears the *–re* (‘onto’) ending in the meaning of *to*, when referring back to the city, only *oda* (‘there’) is appropriate.[60]

(12) A: *Szeged-re mész?*
Szeged-onto go.2SG
‘Are you going to Szeged?’
B: *Igen, *arra/ oda megyek.*
yes that.onto there go.1SG
‘Yes, I’m going there.’

This pattern is preserved at CTLD-Hun, as attested by the following examples: it is always the generally used pronoun that occurs. If a case-marking is featured in several uses (for example, in 13e-f), the functionally appropriate form is used.

(13) a *Szegedre, *arra/ oda szívesen megyünk.*
Szeged.onto that.onto there gladly go.1PL
‘As for Szeged, we gladly go there.’

b *Debrecenben, *abban/ ott sokan élnek.*
Debrecen.in that.in there many live.3PL
‘As for Debrecen, many people live there.’

[60] As Alberti Gábor (p.c.) notes, the choice is dependent on the locative argument being underlyingly an adverbial or a prepositional phrase.

c *Izlandon, *azon/ ott szívesen laknék.*
Iceland.on that.on there gladly live.COND.1SG
'As for Iceland, I would gladly live there.'

d *Vidékre, *arra/ oda nem költöznék.*
countryside.onto that.onto there not move.COND.1SG
'As for the countryside, I wouldn't move there.'

e *Az egyetemen, *azon/ ott sokan dolgoznak.*
the university.on that.on there many work.3PL
'As for the university, many people work there.'

f *Az egyetemen, azon/ *ott sokan dolgoznak.*
the university.on that.on there many work.3PL
'As for the university, many people work on it.' (many people work on the plans, construction, etc. of the university)

An interesting consequence of this is that in certain cases, the semantic and formal requirements are in conflict, so there is no perfect form for the resumptive pronoun. One such case is (14).

(14) *Olyan helyekre, ?oda/ ?olyanokra/ ?azokra nem megyünk.*
such place.PL.onto there such.PL.onto those.onto not go.1PL
'As for such places, we won't go there.'

The problem in (14) is that the grammatically most appropriate form would **odákra* ('there.PL.onto'), which is semantically and formally matching) does not exist. The other possible forms are either semantically or formally conflict with the left dislocated constituent. What is evident from our discussion regarding the form of the resumptive is that one cannot impose rigid syntactic criteria on it; semantic considerations play the crucial role.

The remaining properties of CTLD-Hun follow from the general properties of contrastive topics: they occur in the "topic field" of the main- or subordinate clauses, intermingled with non-contrastive topics.

(15) *János Marit, azt Tamással reggel látta. (Katit délután.)*
John Mary.ACC that.ACC Thomas.with morning saw.3SG Kate.ACC afternoon
'As for Mary, John, with Thomas saw her in the morning. (They saw Kate in the afternoon.)'

(16) *Erika azt mondta, hogy Júliával azzal gyakran találkozik.*
Erika that.ACC said.3SG that(C) Julia.with that.with often meets.3SG
(*Tiborral ritkán.)*
Tibor.with rarely
'Erika said that as for Julia, she meets her often. (She meets Tibor rarely.)'

Although stylistically marked, it is grammatically acceptable to have multiple instances of CTLD-Hun in a sentence. The following example is from Den Dikken and Surányi (to appear). (Stylistic or processing factors may put a limit to the process, but not syntax *per se*.)

(17) *Jánost azt Péterrel azzal még nem kevertem össze.*
John.ACC that.ACC Peter.with that.with yet not mixed.up.1SG
'John, with Peter I have never mixed him up.'

If the host-elements and the pronouns are separated in a configuration like (18), the order of the pronouns preferably follow the order of the hosts, but the alternative order is not ruled out either.

(18) a *Jánost Péterrel azt azzal még nem kevertem össze.*
John.ACC Peter.with that.ACC that.with yet not mixed.up.1SG

b *Jánost Péterrel azzal azt még nem kevertem össze.*
Both: "John, with Peter I have never mixed him up.'

Like contrastive topics in general in Hungarian, constituents in CTLD-Hun "reconstruct", so they are systematically interpreted as if they occupied a lower position in a syntactic tree. So they avoid Principle A-violations (19a), and they are capable of taking narrow scope (19b).

(19) a *Magát azt szereti János.*
himself.ACC that.ACC likes John
'(as for) Himself, John likes.'

b *Az pro$_i$*[61] *anyját azt mindenki$_i$ szereti.*
the mother.POSS.3SG.ACC that.ACC everyone likes
'Everyone$_i$ likes his$_i$ mother.'

5.1.2 Properties of NCLD-Hun

NCLD-Hun is a relatively less-researched sentence-type of Hungarian. The only deeper investigation of it has been carried out by Lipták (2012). Apart from this, NCLD-Hun is mentioned by Baloghné Nagy (2013), but apart from the lack of contrastivity, no other

[61] The "pro" here represents the unpronounced possessor. I use it in the string for clarity purposes, not because I posit a zero c-structural element. Such unpronounced entities are only present in f-structure within the LFG-framework (due to Economy of Expression).

properties are stated by her. For convenience, let us repeat the example sentence from (2), as (20).

(20) *Erre János az fogta magát, és elszaladt.*
then John that took.3SG himself.ACC and away.ran.3SG
'Then John, he went and ran away.'

A difficulty of the characterization of NCLD-Hun is the fact that without context, intonation and knowledge of the speaker's communicative intention, a sentence like (20) is actually ambiguous between the noncontrastive NCLD-Hun and the contrastive CTLD-Hun readings. So, given the proper conditions, (20) may be interpreted contrastively (e.g. in contrast to John, Kate tried to do something about the situation). What facilitates the non-contrastive reading is the lack of an obvious alternative and the informal phrase *fogta magát* (literally 'grabbed himself'), but these are not infallible methods. In all the examples in this section, a non-contrastive interpretation is intended, with an accordingly "flat" intonation. (Emphasizing the flat intonation is also the reason behind the lack of the comma between the left dislocated element *János* and the resumptive pronoun.)

Regarding formal features, NCLD-Hun is a more flexible construction like CTLD-Hun, although case-identity is required. Number-identity is optional. Examples (21c-d) are from the Hungarian National Corpus:

(21) a **Ezután a könyv azt visszavittem a könyvtárba, majd*
then the book.NOM that.ACC back.took.1SG the library.to then
hazamentem.
home.went.1SG

b *Ezután a könyveket ?azt/ azokat visszavittem a könyvtárba,*
then the books.ACC that.ACC those.ACC back.took.1SG the library.to
majd hazamentem.
then home.went.1SG
'So I took the books back to the library and went home.'

c *…jól van, az adatokat azt fölírom rendeléskor.*
right be.3SG the data .ACC that.ACC up.write.1SG order.at
'Right, I take a note of the data at the ordering.'

d *Ha már okozott esetleg következményeket is, az ilyen eseteket*
if already caused.3SG perhaps consequences.ACC too the such cases.ACC
azt már mindig nehezebb kikezelni…
that.ACC already always difficult.COMPARATIVE treat.INF
'If it perhaps has some consequences, these cases are more difficult to treat.'

The distribution of NCLD-Hun is identical to CTLD-Hun, so it can occur both in main- and in subordinate clauses, intermingled with topics.

(22) a *Azt mesélték, hogy erre János az fogta magát és elszaladt.*
that.ACC told.3PL that(C) this.to John that took.3SG himself.ACC and away.ran.3SG
'It was told that then John, he ran away.'

b *Jánossal Zoltán az Marit táncolni hívta.*
John.with Zoltán that Mary.ACC dance.inf asked.3PL
'With John, Zoltán asked Mary to dance.'

Like in CTLD-Hun, the separability of the left dislocated element and the resumptive is a matter of variation. I take (23) to be grammatical.

(23) *János tegnap az fogta magát és elszaladt.*
John yesterday that took.3SG himself and away.ran.3SG
'John, yesterday he ran away.'

There are no "reconstruction"-effects in NCLD-Hun, just like there is no reconstruction with regular topics.

(24) a **Ezután magát azt János megszerette.*
then himself.ACC that.ACC John got.to.like.3SG
Intended: 'Then John got to like himself.'

b **Ezután az pro$_{*i/j}$ anyja az mindenkit$_i$ megölelt.*
then the mother.POSS.3G that everyone.ACC hugged.3SG
'Then his$_{*i/j}$ mother hugged everyone$_i$.'

In contrast with CTLD-Hun, not every grammatical function and word class is equally acceptable in NCLD-Hun. So far I have exemplified the construction with subjects and objects, which are both grammatical, but if we use non-referential elements like adjectival predicates in it as in (25), we get marked results. (Of course, on the CTLD-Hun-interpretation the sentence is fine.)

(25) **Editet szépnek annak/ olyannak gondolom.*
Edit.ACC pretty.DAT that.DAT such.DAT think.1SG
'I consider Edit pretty.'

(25) makes sense under the assumption that at information structure, NCLD-Hun is intuitively a non-contrastive topic-marking device. Neutral topics may not be non-referential, so the ungrammaticality of (25) falls out from general considerations, and is not related to NCLD-Hun itself. This intuition gets further support if we investigate the interaction of NCLD-Hun with copular sentences. Traditionally (e.g. Declerck 1989), these are classified into five subtypes:[62]

(26) a *János okos (volt).*[63] → Predicational
John smart was
'John is/(was) smart.'

b *A hajnalcsillag az esthajnalcsillag (volt).* → Equative
the morning.star the evening.star was
'The morning star is/(was) the evening star.'

c *Az a férfi (ott) János.* → Identificational
that the man there John
'That man (there) in John.' (In a deictic context.)

d *A demokrácia egy olyan rendszer (volt), ahol a népakarat érvényesül.* → Definitional
the democracy one such system was where the people's.will reign.3SG
'Democracy is a system where the will of the people rules.'

e *A legjobb jelölt János (volt).* → Specificational
the best candidate John was
'The best candidate is/was John.'

NCLD-Hun may occur with the subject of each of these types (as seen in 27, with the added pronouns *az* ('that')), except for the specificational subtype. (26e) is only grammatical on the contrastive reading. That is, when the sentence is parsed as CTLD-Hun.

(27) a *János az okos (volt).*

b *A hajnalcsillag az esthajnalcsillag (volt).*

c *Az a férfi az (ott) János.*

d *A demokrácia az egy olyan rendszer (volt), ahol a népakarat érvényesül.*

e *#A legjobb jelölt az János volt.* (grammatical on the contrastive reading)

[62] For a recent overview of the state of affairs pertaining to copular clauses, see Mikkelsen (2011).
Also, I would like to thank the audience of LingDokKonf 19. for calling my attention to the relationship of copular clauses and NCLD.

[63] In Hungarian, the copula is zero in present tense. For more on Hungarian copular clauses, Laczkó (2012) in LFG and Hegedűs (2013) in minimalist framework.

The problem with (27e) lies in the semantics of specificational copular sentences. The initial constituents in (27a-d) are regular referential aboutness topics, the sentence says something about them. This is not the case in the specificational sentences (27e). The constituent *a legjobb jelölt* (‘the best candidate’*)* is not an entity that is the subject of predication, but it is a predicate itself, introducing a variable (“the best candidate is X”). It is not a referential entity, so it is not a topic. Not being topic-entity, it is not a suitable host for the NCLD-Hun pronoun. The other nominative entity is also not suitable for the pronoun, as it identifies the variable introduced by the predicate (“X=János”). As such, its interpretation is that of an information focus.

5.1.3 Properties of CFLD-Hun

The CFLD-Hun represents an unmapped territory of the Hungarian language. It receives only one mention in the literature, by Kenesei et al. (1998). Let us repeat our example from (3) as (28).

(28) *A könyvet, AZT olvastam a szobában (és nem az újságot).*
the book.ACC that.ACC read.1SG the room.in and not the newspaper.ACC
‘The book, I read THAT in the room (and not the newspaper).’

As indicated by the label of the construction, here the resumptive pronoun resides in the preverbal focus-position of the sentence (c.f.: ***A könyvet, AZT a szobában olvastam***). Phonologically, there is a noticeable pause in pronunciation between the fronted element and the rest of the clause.

It might occur to us that CFLD-Hun is simply a variant of CTLD-Hun or that there is some general “left-dislocating” mechanism at work, which targets contrastive information-structural categories. This would not be implausible given that I have argued in section 2.6.1 that contrastivity is a basic feature of information structure. Also, English Topicalization was also argued to involve +CONTRASTIVE discourse functions. However, as we will see, the syntactic differences are too numerous between the two structures so a unified treatment for them is dispreferrable.

As far as formal features are concerned, there is more flexibility here compared to CTLD-Hun. The case and number of the left-dislocated element is not necessarily identical withthose of the resumptive pronoun. The deictic features on the other hand should be identical here as well.

(29) a *?A könyv, János AZT vitte vissza a könyvtárba.*
the book.NOM John that.ACC took.3SG back the library.to
'The book, John took IT back to the library.'

b *?A könyveket, János AZT vitte vissza a könyvtárba.*
the book.PL.ACC John that.ACC took.3SG back the library.to

c *#Azt a könyvet, EZT vittem vissza a könyvtárba.*
that.ACC the book.ACC this.ACC took.3SG back the library.to

The occurrence of CFLD-Hun is much more restricted than CTLD-Hun. In contrast with CTLD-Hun, which freely intermingles with topics, CFLD-Hun is preferably string-initial. Preceding elements cause the sentence to degrade. From this it follows that CFLD-Hun is marked in embedded contexts.

(30) a *?Mari a könyvet, AZT vitte vissza a könyvtárba.*
Mary the book.ACC that.ACC took.3SG back the library.to

b *?János azt mondta, hogy a könyvet, AZT vitte vissza a könyvtárba.*
John that.ACC said.3SG that(C) the book.ACC that.ACC took.3SG back the library.to

The usage of the resumptive pronoun is also markedly different in CFLD-Hun. In Hungarian, to refer to a person with a demonstrative pronoun is only felicitous sentence-internally (as in CTLD-Hun, NCLD-Hun). If the pronoun and the referent person are in different sentences, demonstratives are considered rude and personal pronouns ought to be used.

(31) A: *Láttad már ma Ádámot?*
saw.2SG already today Adam.ACC
'Have you seen Adam today?'
B: *Nem, őt/ #azt még nem láttam.*
no him that.ACC yet not saw.1SG
'No, I haven't seen #that/him today.'

As noted this principle is not operational in the case of the other LD-Huns, one can freely use demonstratives, without the negative pragmatic effect. However, in CFLD-Hun the restriction is in place, one should use a personal pronoun.

(32) *Jánost, ŐT/ #AZT hívtam fel tegnap.*
John.ACC him that.ACC called.1SG up yesterday
'John, I called HIM/#THAT yesterday.'

Finally, the CFLD-Hun differs from CTLD-Hun in terms of binding and scope as well. Compared to the fully grammatical forms in (19), the parallel structures with CFLD-Hun are definitely marked. In other words, no "reconstruction"-effects can be observed. (Although the effects discussed in connection with LD-Eng and binding may blur the picture a bit, see 36 in Chapter 4).

(33) a *??Magát, ŐT szereti János.*
himself him.ACC loves.3SG John

b *??Az pro$_i$ anyját, ŐT szereti mindenki$_i$.*
the mother.POSS.3SG.ACC her.ACC likes everyone

Considering all these, it seems highly plausible that CFLD-Hun constitute a structure distinct from other versions of LD-Hun. The exact nature of the distinction will be fleshed out in the next section, where I preset my analysis of LD-Hun.

5.1.4 Analysis of LD-Hun

Having surveyed the properties of LD-Hun, now we are in the position of positing an analysis for these Hungarian structures. This section will follow a top-down approach. I will start out with some discourse/pragmatic considerations. I will then outline some general views about the nature of different types of LD-Hun and finally I get down to the technicalities of the analysis, in the framework of LFG.

To start with, let us ask a basic semantic/pragmatic question about the "signature-elements" of LD-Hun, the pronouns themselves: what is their role in the structure? This is not evident at first sight, as they are optional in all LD-Hun constructions. Nothing happens if they are omitted, as in (34), where the pronoun-less versions of (1)-(3) are shown.

(34) a *Jánost Kati szereti.*
John.ACC Kate likes
'As for John, Kate likes him.'

b *Erre János fogta magát, és elszaladt.*
then John took.3SG himself.ACC and away.ran.3SG
'Then John ran away.'

c *A KÖNYVET olvastam a szobában (és nem az újságot).*
the book.ACC read.1SG the room.in and not the newspaper.ACC
'I read THE BOOK in the room (and not the newspaper).'

However, pronouns do contribute semantically and pragmatically to sentences, so there should be contexts where their presence/absence makes a difference. As it turns out, there are such contexts. Consider (35) first, taken from Gécseg & Kiefer (2009). They use it to illustrate that topics in Hungarian are not necessarily specific, contra É. Kiss (2002:10), as *valaki* ('somebody') in (35) may have a non-specific reading (see also footnote 32 in section 3.1.).

(35) *Valaki bekopogott az ajtómon tegnap.*
somebody in.knocked.3SG the door.POSS.1SG.on yesterday
'Somebody knocked at my door yesterday.'

If a LD-Hun pronoun is added to such a sentence, it becomes infelicitous if uttered out of the blue.

(36) *Elmondom mi történt tegnap. Valaki (#az) bekopogott az ajtómon.*
tell.1SG what happened.3SG yesterday somebody that in.knocked the door.POSS.1SG.on
'I tell you what happened yesterday. Somebody knocked at my door.'

Another instance where the pronoun makes a difference is when it is attached to an indefinite nonpronominal expression, like the example in footnote 33, repeated here as (37). Here the pronoun may not be added if the indefinite is not specific.

(37) *Elmondom mi történt tegnap*. ('I tell you what happened yesterday.)
Egy gyerek (#az) leesett a villamosról, de szerencsére nem sérült meg.
one child that fell.off.3SG the tram.from but luckily not injured.3SG PV
'A child fell off the tram but luckily, he wasn't injured.'

Nonetheless, both *valaki* ('somebody') and *egy gyerek* ('a child') may be acceptable in contexts where the referents are members of contextually given sets, when they are specific.

(38) a *Sok embert meghívtam. Valaki az eljött, valaki nem.*
lot person.ACC invited.1SG somebody that came.3SG somebody not
'I invited a lot of people. Some came, some didn't.'

b *Több gyerek játszott a homokozóban. Egy gyerek az*
several kid played.3SG the sandpit.in one child that
kitalálta, hogy épít egy homokvárat.
came.up.with.3SG that(C) build.3SG one sandcastle.ACC
'Several kids were playing in the sandpit. One of them came up with the idea of building a sandcastle.'

From this it might seem that the import of the pronoun is that it imposes a specificity restriction on the host element. This conclusion however is premature, as generics may readily be supplemented with an LD-pronoun. (39) is grammatical and felicitous, regardless of contrastivity.[64]

(39) a *A delfin az egy okos állat.*
the dolphin that one smart animal
'The dolphin is a smart animal.'

b *Egy gyerek az nem tudhat mindent.*
one kid that not know.COND.3SG everything.ACC
'A kid cannot know everything.'

It is also clear from (38) and (39) that it is also not indefiniteness that causes the infelicity of (36) and (37). It appears that simply referencing a semantic or syntactic feature will not enable us to pin down the conditions for the pronouns in LD-Hun. We must turn to pragmatics and investigate the pattern in (36)-(39) from the perspective of the usage of the pronouns themselves. I would like to propose that what governs the distribution is the referential givenness-conditions on the usage of the pronominals.

Recall from section 2.6.1.1 that referential givenness is about "a relation between a linguistic expression and a corresponding non-linguistic entity in the speaker/hearer's mind" (Gundel 1999). One approach about the characterization of referential givennes is the Givenness Hierarchy of Gundel, Hedberg and Zacharski (1993) (already mentioned in section 4.3.1). They distinguish between six cognitive statuses that a referring expression may have. The statuses are the following, from lowest to highest. The definitions are from Gundel et al. (2010), the examples are from Gundel, Hedberg and Zacharski (1993):[65]

(40) a Type identifiable: associate type representation.
I couldn't sleep last night. A dog (next door) kept me awake.

[64] I would like to thank Balázs Surányi for calling my attention to examples like (38b) and (39b).

[65] Note that the higher statuses entail the lower ones, e.g. anything that is "referential" is also "type identifiable". For details, see the references.

b Referential: associate unique representation by end of sentence.
I couldn't sleep last night. This dog (next door) kept me awake.

c Uniquely identifiable: associate unique representation by end of NP.
I couldn't sleep last night. The dog (next door) kept me awake.

d Familiar: associate representation in memory.
I couldn't sleep last night. That dog (next door) kept me awake.

e Activated: associate representation in working memory.
I couldn't sleep last night. That kept me awake.

f In focus: associate representation in focus of attention.
I couldn't sleep last night because of your dog. It kept barking.

I propose that the distal demonstrative that is used in LD-Hun by default may be felicitously used if the referent is at least "referential" in this sense: a unique entity may be associated to the referent of the pronoun in the mind of the interlocutors. This subsumes the notion of specificity which is a favorable aspect as von Heusinger (2011) argues that specificity should be understood as "referential anchoring". For example, when using a specific indefinite (as in 38) the hearer must establish a permanent representation of the referent. What I would like to add to this is that in the case of generics, the referent may be construed as a unique class of entities. That is, the concept of being "referential" in the sense of the Givenness Hierarchy generalizes over specific indefinites and generics. However, nonspecific indefinites are only type-identifiable, as is clear from (40a). This explains the infelicity of (36) and (37) as the fronted elements in these cases fail to single out an entity.

Now we can turn to the syntactic analysis of LD-Hun. My basic claim is that a line has to be drawn between the status of CFLD-Hun on one side and CTLD-Hun and NCLD-Hun on the other. CFLD-Hun should be treated like LD-Eng, with the left-dislocated element being a "syntactic orphan": it is not properly integrated into the host sentence. In contrast, the other two are proper parts of the sentence they are attached to.

To back up the claim of CFLD-Hun being extra-syntactic, let us review its properties. Many of these should sound familiar from our discussion of LD-Eng from previous sections.

(i) The connectivity-effects are weak between the fronted element and the resumptive pronoun, there need not be number- or case-identity, just like in LD-Eng. Since the fronted element bears no syntactic connection with the pronoun, strict featural matching is not required. However, general discourse considerations are still operational, so wildly inconsistent specifications are going to be infelicitous (e.g. referring to a distal expression with a proximal pronoun, as in (41b). However, this

is no more surprising than the strangeness of (42), for which obviously no one would think that a syntactic characterization would be necessary.

(41) a *?A könyv, János AZT vitte vissza a*
the book.SG.NOM John that.ACC took.3SG back the
könyvtárba.
library.to
'The book, John took THAT back to the library.'

b *#Azt a könyvet, ezt vittem vissza a*
that.ACC the book.ACC this.ACC took.1SG back the
könyvtárba.
library.to
'That book, I took THIS back to the library.'

(42) *Look at [that car]$_i$! #This$_i$/That is so beautiful!*

(ii) Its occurrence is preferred in a sentence-initial position. While the sentences in (43) cannot be starred, there is some degradation compared to a sentence-initial version. The degree of the markedness could be subject to inter-speaker variation. This is exactly the situation with LD-Eng, for which I also subscribe to an "orphan"-analysis.

(43) a *??Mari a könyvet, AZT vitte vissza a könyvtárba.*
Mary the book.ACC that.ACC took.3SG back the library.to

b *??János azt mondta, hogy a könyvet, AZT*
John that.ACC said.3SG that(C) the book.ACC that.ACC
vitte vissza a könyvtárba.
took.3SG back the library.to

(iii) When referring to people, a personal pronoun must be used. Since referring to a person with a demonstrative is only felicitous sentence-internally in Hungarian, it is expected that the extra-sentential string-initial personal name in CFLD-Hun can only be referred to with a personal pronoun.

(44) *Jánost, ŐT/ #AZT hívtam fel tegnap.*
John him that.ACC called.1SG up yesterday
'John, I called HIM/#THAT yesterday.'

(iv) Weak reconstruction-effects. Since reconstruction is a sentence-internal process, it is expected not to save sentences like (45) (though remember from section 4.3 that

variable binding may still be possible via some extrasyntactic process (e.g. Shaer & Frey's (2003) suggestion of "semantic subordination").

(45) a	*??Magát,*	*AZT*	*ugatta*	*meg*	*a*	*kutya.*
	himself.ACC	that.ACC	barked.3SG	PV	the	dog
b	*??A pro$_i$*	*kutyáját,*	*AZT*	*szereti*	*mindenki$_i$.*	
	the	dog.POSS.3SG.ACC	that.ACC	likes	everyone	

Thus, the properties of CFLD-Hun all make sense if we assume that the fronted element is extra-sentential, like LD-Eng. We could think of it as if the speaker first makes an announcement of some salient entity and then makes some assertion about it. The real structure behind it could be more faithfully indicated with a punctuation like (46).

(46)	*Jánost*	– *tegnap*	*ŐT*	*hívtam*	*fel.*
	John.ACC	yesterday	him	called.1SG	up
	'John – yesterday I called HIM up.'				

CTLD-Hun and NCLD-Hun contrast with CFLD-Hun in that they are proper parts of the host-sentence. That is why they show different behavior in the aforementioned tests: they show connectivity or reconstruction-effects, freely occur in sentence-internal positions, can use demonstratives to refer to people. CTLD-HUN also shows "reconstruction"-effects.

Their integration into the sentence may be implemented with additional annotations for the Hungarian sentence-structure introduced in chapter 3. My proposal is shown below, with a sample sentence. The pronoun resides in the "topic field" of the sentence, marked with LD-Hun PRON in Figure 1. The necessary annotations for it are shown in separately in (47), to avoid clutter.

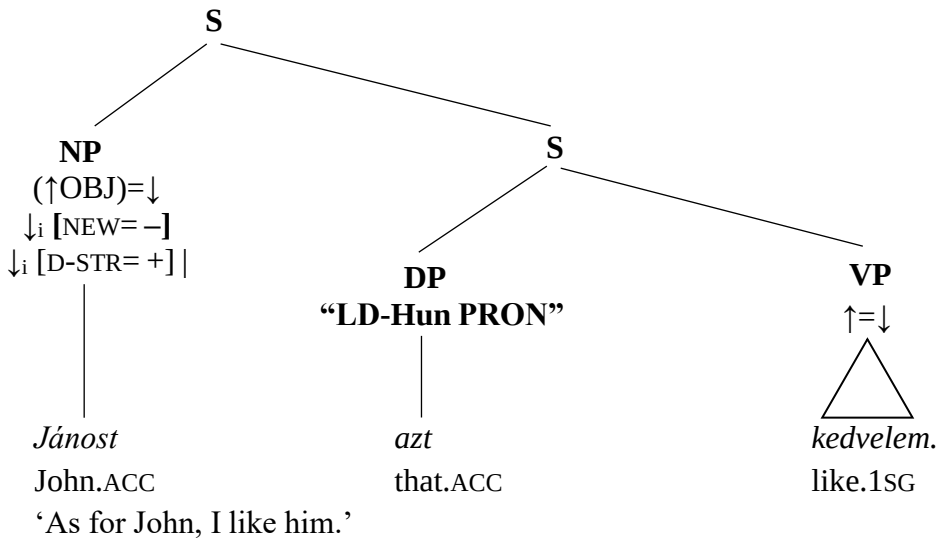

Figure 1.
CTLD-Hun/NCLD-Hun in the Hungarian clause.

Depending on whether we talk about CTLD-Hun or NCLD-Hun, the annotation of the pronoun (in the place of “LD-Hun PRON”) is the following. (47a) shows the appropriate annotations for CTLD-HUN, while (47b) is for NCLD.

(47) a ↓∈ ((↑GF_α) ADJUNCT)
(↓PRED FN)=$_c$ PRO
(↓PRON TYPE) =$_c${DEM | PERS}
(↓INDEX) = $_c$ (↑GF_α INDEX)
{(↓ CASE) =$_c$ (↑GF_α CASE)
(↓ NUM) =$_c$ (↑GF_α NUM) |
¬ (↓ CASE)
¬ (↓ NUM)}

b ↓∈ ((↑GF_α) ADJUNCT)
(↓PRED FN) =$_c$ PRO
(↓PRON TYPE) =$_c$ {DEM | PERS}
(↓INDEX) = $_c$ (↑GF_α INDEX)
(↓CASE) = $_c$ (↑GF_α CASE)

Let us elaborate on the annotations. As the pronouns are optional, they are functionally regarded as adjuncts of some grammatical function (identified as GFα) of the main clause.[66] Note that this does not make any commitment toward the adjacency of the host and the pronominal, only requires that the pronoun have an antecedent in the sentence.

As noted the =$_c$ in the annotations is a “constraining equation”. So for example (↓PRED FN)=$_c$ PRO checks whether the element occupying that slot is a pronoun. If not, the sentence is ungrammatical so this prevents the free insertion of nonpronominal constituents with such specifications.

The pronoun itself must be a demonstrative or a personal one. I take the choice to be free, subject to speaker choice and general semantic restrictions. The referent of the pronoun

[66] This is comparable to Lipták (2011), who argues them to be appositive modifiers.

must be identical with the grammatical function that the pronoun is adjunct of. This is encoded as tagging the GF with the α-subscript, so the specific GF is kept constant throughout the annotation. In other words, the subscript α is meant to ensure that the reference, the case and the number features of the pronoun are tied to the same grammatical function. (So it is formally prevented that e.g. the pronoun associated with the OBJ is co-referent with the SUBJ or formally matches some OBL).

In CTLD-Hun the case and number features of the pronoun must match the case and number feature of the host, or alternatively a caseless, numberless pronoun is used. This happens for instance in the case of *itt*/*ott* ('here'/'there') or *így*/*úgy* ('so.PROX'/'so.DIST'), see examples (10) and (11). The choice is not constrained here, as it should fall out from general principles governing pronoun selection. In the case of NCLD-Hun, only the case feature is set to match, allowing for mismatches of the sort illustrated in (21).

What is not resolved here (and in other analyses in the literature) is how Binding Theory interacts with LD-Hun. The pronouns are bound, which should be a Principle B violation. One is tempted to say that such uses of pronouns, when they are configurationally licensed are exempt from the standard restrictions of Binding Theory. Such a distinction may be drawn along the lines of "deep" and "surface" anaphora (Hankamer & Sag 1976, see also den Dikken & Surányi to appear). The hosts themselves are not exempt from such considerations: reversing the order of the coreferent pronoun and an R-expression leads to a Principle C violation and ungrammaticality.

(48) **Azt$_{i/}$ Őt$_i$ Jánost$_i$ kedvelem.*
that.ACC him.ACC John.ACC like.1SG

5.2 Operator fronting in Hungarian

The other discourse-related long-distance dependency under scrutiny in Hungarian is what I label "Operator fronting" (OF). The name reflects my aim to provide a concise theory- and analysis-neutral description of the phenomenon. In the previous literature, the structure has also been called "Focus raising" (Kenesei 1992, Lipták 1998, Gervain 2004), "Operator raising" (Gervain 2002) and "Long operator movement" (É. Kiss 2002). These names for the phenomenon mirror the Chomskyan frameworks that the analyses are couched in. The framework of this dissertation is Lexical-Functional Grammar, so I have tried to find a label which has the least amount of implicature about the analysis. "Movement" and "Raising" explicitly suggest that the analysis employs some kind of dislocation and are obviously

Chomskyan terms. “Fronting” still has some derivational flavor to it, but it is at least not an exclusively Chomskyan term, it is more theory-neutral.

An example for Operator Fronting is shown in (49a), where the locative phrase *Párizsba* (‘to Paris’) is a complement of the embedded verb, yet it occurs in the main clause, sentence-initially, in a fronted position. It also bears some discourse function (hence the “operator” in the name)[67]. For contrast, (49b) shows a standard subordinate clause, with no fronted material.

(49) a *Párizsba mondtad, hogy mész.*
Paris.into said.2SG that(C) go.2SG
‘To Paris you said that you will go.’

b *Azt mondtad, hogy Párizsba mész.*
that said.2SG that(C) Paris.into go.2SG
‘You said that you will go to Paris.’

5.2.1 Properties of Operator Fronting

As was shown in (49), the essence of OF is that some element that would normally be assumed to be belonging to the finite subordinate clause, surfaces in the matrix clause. In (49), the affected element is an oblique function (a prepositional phrase, in c-structural terms), but there is a range of other options as well. In (50), it is demonstrated that the fronted element could also be a subject (50a), an object (50b) or an adjunct (50c) of an embedded clause.

(50) a *János mondtad, hogy jön a partira.*
John said.2SG. that(C) comes the party.onto
‘(Of) John you said that he will come to the party.’

b *Egy kutyát mondtad, hogy láttál.*
one dog.ACC said.2SG that(C) saw.2SG.INDEF
‘A dog you said that you saw.’

c *Tegnap mondtad, hogy sokat ettél.*
yesterday said.2SG. that(C) lot.ACC ate.2SG
‘Yesterday you said that you ate a lot.’ (intended: the eating was yesterday)

A note is in order here. Fronting subjects, obliques or adjuncts may result in ambiguous structures, as sometimes these elements may also be interpreted in such a way that they belong to the matrix clause. It is clearly observable in (50c), where in principle *yesterday* could also

[67] The information-structural import of the “fronting” is ambiguous. That’s why I use topicalization in some of the translations, which also has several interpretational options, as has been discussed in 4.1.

refer to the time of saying. In (50a), if *say* had 3rd person singular conjugation, the sentence would be ambiguous between the meaning "(someone) said that John will come to the party" analogous to the given meaning in (50a), or the unremarkable interpretation "John said that he comes to the party", as the conjugation could be equally triggered by the *pro*-dropped 3SG subject or *János*. If there is an overt subject plus a nominative fronted element as in (51), the sentence is judged degraded by native speakers (Szűcs 2010), because of the ambiguity and processing difficulty caused by the two initial nominative constituents.

(51) *?János Kati mondta, hogy jön.*
John Kate said.3SG that(C) comes
'John (of) Kate said that he will come. / (Of) John, Kate said that he will come.'

The case is similar with obliques. In (52), *Párizsban* ('in Paris') could refer to the location of the running (the OF interpretation) or to the location of saying.

(52) *Párizsban mondtad, hogy futottál.*
Paris.in said.2SG that(C) ran.2SG
'Of Paris you said that you ran there.' / 'You said in Paris that you ran.'

Although such cases may be interesting from a language processing point of view, they will not be addressed in this dissertation. I just note that this complication exists so one has to be careful about the construction of examples and try to avoid problematic ones if possible (e.g. by choosing an oblique that cannot be interpreted in the matrix clause, as in 49a).

The pronoun associated with subordinate clauses in Hungarian (see 49b) cannot be present when OF takes place.

(53) *(*Azt) Párizsba mondtad (*azt), hogy mész.*
that.ACC Paris.to said.2SG that.ACC that(C) go.2SG

The distance between the fronted element and its standard position is not limited to the immediate subordinate clause, it can span across several clauses.

(54) *Párizsba mondtad, hogy hallottad, hogy megy János.*
Paris.to said.2SG that(C) heard.2SG.DEF that(C) goes John
'To Paris you said that you heard that John will go.'

There are a couple of interesting variations to be observed about OF. The first one is a possible "case-switch": the fronted constituent may bear the case assigned by the matrix predicate, even if that case does not correspond to the function of the fronted element in the embedded clause. For instance, (55b) shows a fronted embedded subject, (55c) shows an embedded oblique and (55d) shows an embedded possessor. These may all bear the accusative case assigned to the OBJ argument of the main verb. In (55c-d), a resumptive pronoun must surface in the embedded clause. As noted in den Dikken (2010), such resumptive pronouns are also marginally possible with fronted accusative subjects (55b)[68]. The resumptive pronoun is impossible if the fronted element bears its original case (e.g. 55a).

(55) a *János mondtad, hogy (*ő) jön a partira.*
John.NOM said.2SG that(C) he comes the party.to
'(Of) John you said that he will come to the party.'

b *János-t mondtad, hogy (?ő) jön a partira.*
John-ACC said.2SG that(C) he comes the party.to

c *Párizs-t mondtad, hogy oda utazol.*
John-ACC said.2SG that(C) there travel.2SG.
'(About) Paris you said that you will go there.'

d *János-t mondtad, hogy az (?ő) órája elveszett.*
John-ACC said.2SG that(C) the his watch.POSS.3SG lost
'(Of) John you said that his watch got lost.'

If an embedded indefinite object is fronted to the matrix object position, the main verb may show definite conjugation, triggered by the embedded clause (which counts as a definite object), or it may show indefinite conjugation, corresponding to the fronted object. So in this case the conjugation distinguishes between two possible configurations.

(56) *Két almát mondtad/ mondtál, hogy vettél.*
two apple.ACC said.2SG.DEF said.SG.INDEF that(C) bought.2SG.INDEF
'Two apples you said that you bought.'

It has not been noted in the literature that the case-mismatch presented in (55) is actually not restricted to main clause objects and accusative-marking. As discussed in section 3.2, subordinate clauses may also be associated with inherently case-marked pronouns. The fronted element may also replace such a pronoun, assuming its case.

[68] Since Hungarian is a pro-drop language, personal pronouns are avoided if independent factors do not necessitate them. Hence the marked nature of (55b-d). Adding emphatic stress to the resumptive improves such sentences.

(57) a *Arról gondolkodtam, hogy János jön a partira.*
that.from thought.1SG that(C) John comes the party.to
'I was thinking about John coming to the party.'

b *Jánosról gondolkodtam, hogy jön a partira.*
John.from thought.1SG that(C) comes the party.to
About John I was thinking that he comes to the party.

(58) a *Attól tartok, hogy János jön a partira.*
that.from be.afraid.1SG that(C) John comes the party.to
'I'm afraid of John coming to the party.'

b *Jánostól tartok, hogy jön a partira.*
John.from be.afraid.1SG that(C) comes the party.to
'Of John I was afraid that he come to the party.'

The second interesting variation is about agreement matching/mismatching and it was first observed by Gervain (2002): if the fronted constituent that has been case-switched contains a quantifier or a numeral, the embedded verb may show either singular or plural agreement. This is unexpected, since such elements in a simple (non-subordinated) sentence invariably trigger singular agreement (59d). This is shown with a fronted embedded subject bearing accusative case in (59a-b). Oblique-marked case-switched elements pattern similarly (59c). If the fronted element bears its original nominative case, only singular agreement is permitted (59e). From this it seems that it is the case-switching that enables the number-mismatch.

(59) a *Az összes lány-t mondtad, hogy jön.*
the every girl-ACC said.2SG that(C) come.3SG
'(Of) every girl you said that they come.'

b *Az összes lány-t mondtad, hogy jönnek.*
the every girl-ACC said.2SG that(C) come.3PL

c *Az összes lányról gondolkodtam, hogy jön/ jönnek.*
the every girl-from thought.1SG that(C) come.3SG come.3PL
'(Of) every girl I was thinking that they come.

d *Az összes lány jön / *jönnek.*
the every girl come.3SG come.3PL

e *Az összes lány mondtad, hogy jön / *jönnek.*
the every girl.NOM said.2SG that(C) come.3SG come.3PL

Thus, taking the two axes of variation with fronted elements (original/switched case, matching/mismatching agreement), there are four possible combinations, out of which the one with original case and mismatching agreement is impossible. About the three remaining

combinations, now the question is whether people judge them uniformly or there is some dialectal/idiolectal variation.

According to Gervain (2002), there are two groups of speakers. One group accepts both matching (singular in 59a) and mismatching (plural in 59b) agreement on the embedded verb, but only switched case (accusative in 59a) on the fronted element. That is, they accept sentences like (59a) and (59b). The other group accepts both original (nominative) and switched (accusative) case, but only with matching (singular) agreement, so they accept sentences like (59a) and (59c). Jánosi (2013) debates this and claims that while there could be some speaker variation, no consistent speaker groups can be distinguished. I agree with Jánosi (2013): no distinct dialects will be posited, so OF will be treated as a unitary phenomenon with regard to acceptability patterns.

There are also further differences between "case-switched" fronted elements and one bearing their original case. Gervain (2002) shows that the case-switched version may bypass island-violations, for example the complex-NP-constraint:

(60) a *Az elnök-öt mondtad, hogy hallottad a hírt, hogy jön.*
the president-ACC said.2SG that(C) heard.2SG the news.ACC that(C) comes
'(Of) The president you said that you heard the news that he will come.'

b **Az elnök mondtad, hogy hallottad a hírt, hogy jön.*
the president.NOM said.2SG that(C) heard.2SG the news.ACC that(C) comes

Another example for this is shown in (61). This is based on the analysis of subordinate clauses where in presence of an associate pronoun, the CP itself is an adjunct and thus, an island for extraction.

(61) a *János-t mondtad, hogy azt hallottad, hogy jön.*
John-ACC said.2SG that(C) that.ACC heard.2SG that(C) comes
'Of John you said that you heard that he comes.'

b **János mondtad, hogy azt hallottad, hogy jön.*
John.NOM said.2SG that(C) that.ACC heard.2SG that(C) comes
'Of John you said that you heard that he comes.'

Also, Gervain (2009) shows that no strong crossover effects[69] are present in the case-switched version, in contrast to the case-retaining one.

(62) a *Két szomszéd-ot panaszoltál egymásnak, hogy zajos/ zajosak.*
two neighbor-ACC complained.2SG each other.to that(C) noisy.SG noisy.PL
'Two neighbor you complained about each to each other that they are noisy.'

b **Két szomszéd panaszoltad egymásnak, hogy zajos.*
two neighbor.NOM complained.2SG each other.to that(C) noisy.SG

Furthermore, Gervain (2009) observes that the opposite pattern may be observed with reconstruction-effects: it yields ungrammaticality with the (switched) accusative version, but not if the case is retained.

(63) a **Két rossz hír-t egymásról mondtál, hogy letaglózta a fiúkat.*
two bad news-ACC each other.about said.2SG that(C) saddened.3SG the boys.
'Two pieces of bad news about each other you said that saddened the boys.'

b *Két rossz hír egymásról mondtad, hogy letaglózta a fiúkat.*
two bad news.NOM each other.about said.2SG that(C) saddened.3SG the boys.
'Two pieces of bad news about each other you said that saddened the boys.'

Fronting idiom-chunks is impossible with the accusative version, but marginally possible with the nominative.

(64) a **A gépszíj-at mondtad, hogy elkapta Jánost.*
the driving.belt-ACC said.2SG that(C) caught.3SG John.ACC
'(Of) The driving belt you said that it caught John./ *John has to work a lot lately.'

b *??A gépszíj mondtad, hogy elkapta Jánost.*
the driving.belt.NOM said.2SG that(C) caught.3SG John.ACC
'(Of) The driving belt you said that it caught John./ John has to work a lot lately.'

In all the examples considered so far, the fronted element was preverbal. However, it is possible to place a case-switched fronted element postverbally (65b-c). Then there is a verum focus on

[69] Strong Crossover (SCO) is the label for the phenomenon where a movement of a wh-phrase "crosses over" a pronoun that binds it.

(i) **Who$_i$ does he$_i$ think you saw t_i?*

the verb. As discussed in 3.2, there is a similar pattern with the associate pronouns of standard subordinate clauses, with which fronted elements are in complementary distribution (65a). Such variation is impossible with non-case-switched fronted constituents.

(65) a *?De hiszen te MONDTAD azt, hogy jön János.*
but you said.2SG that that(C) comes John
'But you DID say, that John comes.'

b ?*De hiszen te MONDTAD János-t, hogy jön.*
but you said.2SG John-ACC that(C) comes
'But (about) John, you DID say that he comes.'

c *?De hiszen te tartasz Jánostól, hogy jön.*
but you be.afraid.2SG John.from that comes
'But you ARE afraid of John, that he comes.'

c **De hiszen te MONDTAD János, hogy jön.*
but you said.2SG John.NOM that(C) comes

In the cases where a fronted subject bears a switched accusative case in the matrix clause, the main verb was *mond* ('say'). This was done for expository reasons, as this is the most commonly used verb in this construction. But this type of OF is obviously not restricted to *mond*. At the very least, it assumed to work with so-called bridge verbs. According to Kenesei (1992), the licensing verbs are *akar* ('want'), *szeretne* ('would like'), *mond* ('say'), *hisz* ('believe'), *képzel* ('imagine'), etc. In É. Kiss (2002:253) we find "*akar* ('want'), *szeretne* 'would like', *kell* ('need'), *szabad* ('may'), *lehet* ('is possible'), *nyilvánvaló* ('is obvious'), *valószínű* ('is likely'), as well as verbs of saying and verbs denoting mental activities, among them *mond* ('say'), *ígér* ('promise'), állít ('claim'), *gondol* ('think'), *hisz* ('believe'), etc."

Before investigating the list of verbs to some more extent, a baseline generalization can be made: the verb must be one that occurs with a subordinate clause, associated with a pronoun.

(66) a **Azt futottuk, hogy János megerősödjön.*
that ran.1PL that(C) John strengthen.SBJV
'We ran every day so that John strengthened.'

b **János(t) futottuk, hogy megerősödjön.*
John(.ACC) ran.1PL that(C) strengthen.SBJV

What the main accounts in the literature (Kenesei 1992, Lipták 1998, É. Kiss 2002, Gervain 2002) tend to agree on is that it is the set of bridge verbs that licenses OF, they only differ in which particular examples they highlight.

Let us explore the landscape in more detail.[70] As a starting point, let us take a look at the set of bridge verbs, as listed by Kálmán (2001), which is the largest list in the literature for Hungarian. Kálmán's (2001) criterion is that a verb is a bridge verb if it can occur in a construction labelled "fake question-word question" by Kálmán (2001). It is a construction where two question words are present: *mit* ('what.ACC') in the main clause and another one in the subordinate clause. According to Kálmán (2001), the construction is triggered by a locality condition: in Hungarian, the question word in front of a main verb is always a local complement of the main verb. From this it follows that extraction, as illustrated in (67a) should be impossible in Hungarian. The "fake question-word question"-construction is then a way to bypass this restriction (67b). Although I do not think either the locality condition or the subsequent analytical idea of "fake-question words" is a proper characterization of the state of affairs, let us look at the data nevertheless. I restrict the set of verbs to those relevant for OF, that is, to those which could occur with an accusative associate pronoun. All the grammaticality judgments in (67) are from Kálmán (2001).

(67) a **Ki mondtad, hogy jön a partira?*
who said.2SG that(C) comes the party.to
'Who did you say that comes to the party?'

b *Mit mondtál/ gondolsz/ állítasz/ álmodtál/ ?sejtesz/*
what.ACC said.2SG think.2SG claim.2SG dreamed.2SG suspect.2SG
**fontolgatsz/ *ígértél, hogy ki jön a partira?*
contemplate.2SG promised.2SG that(C) who comes the party.to
'What did you say/think/claim/dream/suspect/contemplate/promise, who comes to the party?'

c *Mit szeretnél/ akarsz/ javasolsz/ tanácsolsz/*
what.ACC would.like.to.2SG want.2SG suggest.2SG advise.2SG
parancsolsz, hogy ki jöjjön a partira?
order.2SG that(C) who come.SBJV.3SG the party.to
'What do you like/want/suggest/advise/order, who should come to the party?'

So according to Kálmán (2001), the set of bridge verbs in Hungarian includes: *mond* ('say'), *gondol* ('think'), *állít* ('claim'), *álmodik* ('dream'), *sejt* ('suspect'), *szeretne* ('would like to'), *akar* ('want'), *javasol* ('suggest), *tanácsol* ('advise') and *parancsol* ('command').

As a first reaction, there are some unexpected items on this list. While *mond*, *gondol* and *állít* are core members of the set of bridge-verbs cross-linguistically, *dream*, *advise* and *command* are definitely not.

[70] I do not venture into the realm of syntactically and semantically analyzing preverbs in Hungarian, so I restrict the discussion to bare verbs.

(68) a **Who did you dream that you gave a flower to?*

b **Who did you advise/command that we should invite?*

Also, I would like to express doubts about Kálmán's (2001) locality condition: it seems to me that (67a) is a grammatical OF-sentence. The contrast is particularly stark if it is compared to (69):

(69) **Ki álmodtad, hogy jön a partira?*
who dreamed.2SG that(C) comes the party.to
'Who did you dream that comes to the party?'

As for the other judgments in (67b-c), I also express doubt. For me all of the verbs seem to be grammatical in the sentences. I think the reason for this is that the "fake question-word question" is an alternative strategy which circumvents extraction, so it is possible even with nonbridge verbs. Since nothing is extracted in a "fake question-word question", it is hard to see what the relevance of the bridge quality could be in those cases. In such sentences, the OBJ argument of the predicate is the question word *mit* ('what.ACC').

So what happens if we investigate the verbs along the lines of proper extraction? It turns out that the already introduced distinction between elements bearing their original case and "case-switched" elements (illustrated with subjects) surfaces again. When we look at nominative fronted subjects, we get the cross-linguistically standard set of bridge verbs (70a-b). However, accusative-marked fronted subjects occur with a much wider range of verbs (70c-d).

(70) a *Ki mondtad/ gondoltad/ állítottad/ remélted/ *álmodtad/*
who said.2SG thought.2SG claim.2SG hope.2SG dreamed.2SG
**fontolgattad/ *furcsállottad/ *kételted/ *sérelmezted, hogy*
contemplated.2SG found.strange.2SG doubted.2SG resented.2SG that(C)
jön a partira?
comes the party.to
'Who did you say/think… that comes to the party?'

b *Ki akarod/ szeretnéd/ javaslod/ *parancsolod/ *tanácsolod,*
who want.2SG would.like.2SG suggest.2SG order.2SG advise.2SG
hogy jöjjön a partira?
that(C) come..SBJV.3SG the party.to
'Who do you want/would you like… that (s)he should come to the party?'

c *Kit mondtál/ gondoltad/ állítottad/ reméltél/ *álmodtál/*
who.ACC said.2SG thought.2SG claim.2SG hope.2SG dreamed.2SG
fontolgattál/ furcsállottál/ ?kételtél/ ?sérelmeztél, hogy
contemplated.2SG found.strange.2SG doubted.2SG resented.2SG that(C)
jön a partira?
comes the party.to
'Who did you say/think… that comes to the party?'

d *Kit akarsz/ szeretnél/ javasolsz/ *parancsolsz/*
who.ACC want.2SG would.like.2SG suggest.2SG order.2SG
?tanácsolsz, hogy jöjjön a partira?
advise.2SG that(C) come. SBJV.3SG the party.to

A similar distribution may be obtained if we use indicative OF examples. That is, OF with nominative-marked fronted subjects (original case) is grammatical with the set of cross-linguistically attested bridge verbs (71a-b), the accusative version (switched case) shows greater flexibility (71c-d).

(71) a *János mondtad/ gondoltad/ ??állítottad/ remélted/ *álmodtad/*
John said.2SG thought.2SG claim.2SG hope.2SG dreamed.2SG
**utáltad *fontolgattad/ ?furcsállottad/ *kételted/ *sérelmezted,*
hated.2SG contemplated.2SG found.strange.2SG doubted.2SG resented.2SG
hogy jön.
that(C) comes
'(Of) John you said/thought/… that he will come.'

b *János akarod/ szeretnéd/ javaslod/ *parancsolod/ *tanácsolod,*
John want.2SG would.like.2SG suggest.2SG order.2SG advise.2SG
hogy jöjjön.
that(C) come. SBJV.3SG
'(Of) John you want/would like/… that he should come.'

c *Jánost mondtad/ gondoltad/ ?állítottad/ remélted/ *álmodtad/*
John.ACC said.2SG thought.2SG claim.2SG hope.2SG dreamed.2SG
??utáltad fontolgattad/ furcsállottad/ ?kételted/ ?sérelmezted,
hated.2SG contemplated.2SG found.strange.2SG doubted.2SG resented.2SG
hogy jön.
that(C) comes
'(Of) John you said/contemplated/found it strange/doubted/resented that he comes.'

d *Jánost akarod/ szeretnéd/ javaslod/ *parancsolod/ ??tanácsolod,*
John.ACC want.2SG would.like.2SG suggest.2SG order.2SG advise.2SG
hogy jöjjön.
that(C) come. SBJV.3SG
'(Of) John you want/would like/suggest/order/advise that he should come.'

This is not to say that the accusative version is completely free in its distribution. For instance, in the examples above *álmodik* (‘dream’) or *parancsol* (‘order’) rejects it even when the fronted element bears a switched case.

It seems then that the case-retaining version of OF (where the fronted element bears the case that it would also have in its original position (e.g. 71a-b, but also oblique elements in earlier examples like 49) pretty much patterns with the prototypical bridge verbs.

With case switching accusative OF (where the fronted element becomes the OBJ/OBL of the main clause, regardless of its original function), the state of affairs is more delicate. There seems to be a core set of verbs, with which case-switching OF is flawless. These are *mond* (‘say’), *gondol* (‘think’), *ígér* (‘promise’) *kérdez* (‘ask’), *remél* (‘hope’), also *beszél* (‘speak’) and *mesél* (‘tell’). Even this is larger than the set of bridge verbs, licensing the nominative version OF.

After these comes a big “grey zone”, verbs that are strange but acceptable to a considerable extent. Examples for this are *fontolgat* (‘contemplate’), *sejt* (‘surmise), *tud* (‘know’), *sajnál* (‘pity’), *tanácsol* (‘advise’), *gyanít* (‘suspect’), etc. Although there could be differences in acceptability (also between speakers), these are more or less acceptable in accusative OF.

Finally, some verbs which otherwise occur with a subordinate clause which are quite poor in OF. Examples for this group are *álmodik* (‘dream’), *hazudik* (‘lie), *válaszol* (‘answer’), *felel* (‘reply’), *terjeszt* (‘spread the news’), *utál* (‘hate’), *parancsol* (‘order’), *érez* (‘feel’), *nyilatkozik* (‘speak to the press’), as well as manner of speaking verbs, like *suttog* (‘whisper’) or *ordít* (‘shout’).

Instances where the case-switch happens with an inherent oblique case seem fairly unrestricted.

(72) a *Abban bízok, hogy János jön.*
that.in trust.1SG that(C) John comes
‘I have trust about John coming.’

b *Jánosban bízok, hogy jön.*
John.in trust.1SG that(C) comes
‘About John I have trust that he comes.’

(73) a *Attól tartok, hogy János jön.*
that.from be.afraid.1SG that(C) John comes
‘I am afraid that(C) John comes.’

b *Jánostól tartok, hogy jön.*
John.from be.afraid.1SG that(C) comes
Of John I’m afraid that he comes.’

(74) a *Gratulálok ahhoz, hogy megszületettt a gyereked.*
congratulate.1SG that.to that(C) was.born.3SG the child.POSS.2SG
'I congratulate you on the birth of your child.'

b *Gratulálok a gyerekedhez, hogy megszületett.*
congratulate.1SG the child.POSS.2SG.to that(C) was.born.3SG
I congratulate you on your child, that s/he was born.

(75) a *Azzal számolok, hogy János győz.*
that.with count.1SG that(C) John wins
'I expect that John wins.'

b *Jánossal számolok, hogy győz.*
John.with count.1sg that(C) wins
'Of John I expect that he wins.'

What I would like to conclude from this discussion is that the literature's original claim, that OF is restricted to bridge verbs, cannot be maintained. There is a much wider range of verbs which participate, although the distribution is not completely free, especially with regards to switching to accusative case.

A final syntactic point to be discussed is the nature of the complementizer in OF. Remember from section 3.2.1 that in standard subordinate clauses, if the focus of the sentence is a pronoun, the *hogy* ('that(C)') may be absent, as in (76a). An interrogative subordinate clause makes the complementizer optional even in the case of nonpronominal foci, as in (76b).

When OF takes place, the complementizer is always compulsory if the embedded clause is indicative (77a-b). When case-switching takes place, the complementizer is obligatory even if the subordinate clause is interrogative (77c), in contrast with (76b). Omission of the complementizer seems better if the fronted element retains its original case (77d-e).[71]

(76) a *Azt mondtad, Párizsba mész.*
that said.2SG Paris.into go.2SG
'You said that you will go to Paris.'

b [*Jánostól*]$_{IF/CF}$ *kérdeztem, hova megy.*
John.from asked.1SG where goes.
'I asked from John where he is going.'

(77) a **Párizsba mondtad, mész.*
Paris.into said.2SG go.2SG
'To Paris you said that you will go.'

[71] I would like to thank Gábor Alberti for the observation.

b **János-(t) mondtad, jön.*
John-ACC said.2SG. comes
'John you said that he will come.'

c **Jánost kérdezted, jön-e.*
John.ACC asked.2SG. comes-QUESTION PRT
Intended: 'Of John you asked if he comes.'

d ?*János kérdezted, jön-e.*
John.ACC asked.2SG. comes-QUESTION PRT
Intended: 'Of John you asked if he comes.'

e ?*Párizsba kérdezted, megyek-e.*
Paris.into asked.2SG go.1SG-QUESTION PRT
Intended: 'Of Paris you asked me if I go there.'

However, remember that the fronted element may be associated with something embedded at a distance, not just with a grammatical function in an immediately subordinated clause (see 78). When that happens, it is only the first complementizer that is obligatory, the others are droppable along the general restrictions. This has gone unnoticed in the previous literature.

(78) a *János(t) mondtad, *(hogy) gondolod, (hogy) jön.*
John(.ACC) said.2SG that(C) think.2SG that(C) comes
'(Of) John you said that you think that he will come.'

b *János(t) mondtad, *(hogy) Maritól kérdezték, (hogy) jön-e.*
John(.ACC) said.2SG that(C) Mary.from asked.3PL that(C) comes-Q-PRT
'(Of) John you said that they asked Mary if he comes.'

As regards information-structure, it seems clear that the fronted elements bear some kind of discourse prominence, regardless of the potential case-switch. This translates as them being interpreted as either Information focus (IF), Contrastive focus (CF) or Contrastive topic (CT). One can rightly observe that these are exactly the discourse functions that the pronouns associated with subordinate clauses bear. To make sure that the intuition that OF can be either kind of focus, I presented 40 native speakers with sentences containing different versions of OF. The data were dialogues which elicited the Information focus (79) or Contrastive focus (80) interpretation of a fronted constituent, in different cases. Capital letters were added as an indication of sentence stress, and this was explicitly stated in the instructions. This was to further reinforce the CF-reading. The distribution of test-data was the following:

- 4 sentences with a fronted nominative element, IF discourse function (79a)
- 4 sentences with a fronted accusative element, IF discourse function (79b)

- 4 sentences with a fronted oblique element, IF discourse function (79c)
- 4 sentences with a fronted nominative element, CF discourse function (80a)
- 4 sentences with a fronted accusative element, CF discourse function 80b)
- 4 sentences with a fronted oblique element, CF discourse function (80c)
- 24 filler sentences

(79) a A: *Szerinted ki fog nyerni?*
in.your.opinion who AUX win.INF
'In your opinion, who will win?'
B: *Én János gondolom, hogy nyerni fog.*
I John.NOM think.1SG that(C) win.INF AUX.
'(Of)John I think that(C) he will win.'

b A: *Szerinted ki fog nyerni?*
B: *Én Jánost gondolom, hogy nyerni fog.*
I John.ACC think.1SG that(C) win.INF AUX.

c A: *Szerinted hova megy Gábor nyaralni?*
in.your.opinion where goes Gábor holiday.INF
'In your opinion, where will Gábor go on holiday?'
B: *Én Londonba gondolom, hogy megy.*
I London.TO think.1SG that(C) goes.
In your opinion, where will Gábor go on holiday? To London I think he goes.

(80) a A: *Ugye szerinted Péter fog nyerni?*
So in.your.opinion Peter AUX win.INF
'So you think Peter will win?'
B: *Nem, én JÁNOS gondolom, hogy nyerni fog.*
'No, (of) John I think that he will win.'

b A: *Ugye szerinted Péter fog nyerni?*
B: *Nem, én JÁNOST gondolom, hogy nyerni fog.*

c A: *Szerinted Párizsba megy Gábor nyaralni?*
in.your.opinion Paris.to goes Gábor holiday.INF
'Do you think Gábor will go on holiday to Paris?'
B: *Nem, én LONDONBA gondolom, hogy megy.*
'No, To London I think he goes.'

The sentences were presented to the speakers in a pseudo-randomized fashion using an online questionnaire and they had to rate the acceptability of the responses on a 1-to-5 scale in the given context (1: fully unacceptable, 5: fully acceptable). The results are the following.

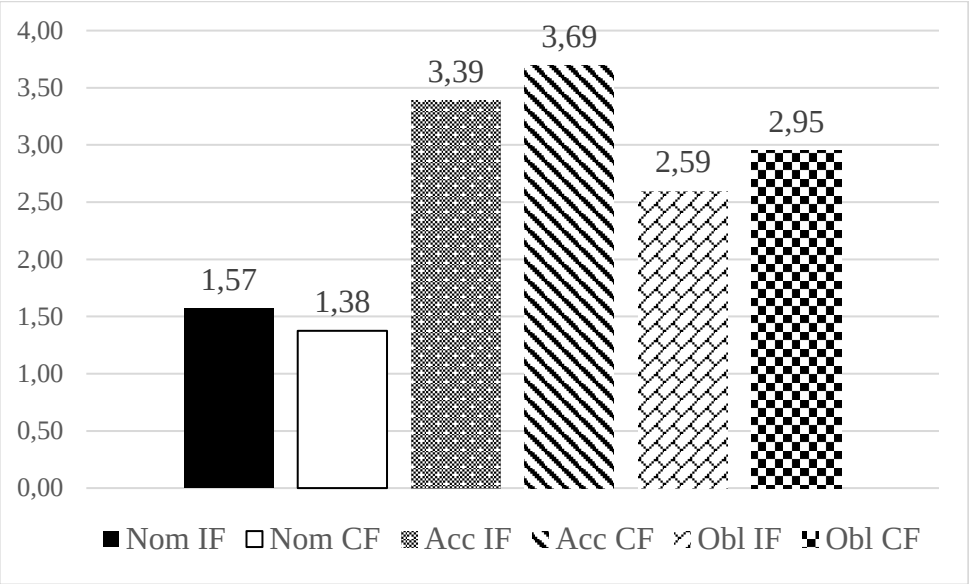

Figure 1.
Focus-types in Operator fronting.

As we can see, accusative-marked (like 79b and 80b) and oblique-marked (like 79c and 80c) elements are generally acceptable, although contrastiveness seems to further enhance their rating. As for nominative-marked elements, they were generally rejected in this experiment, regardless of their contrastivity. This might lead us to believe that this version of OF is nonexistent. This would be in line with some of the previous accounts (Kenesei 1994, Lipták 1998). However, later research (Gervain 2002, Jánosi 2013) established that case-retaining OF of embedded subjects does exist, so I reject that conclusion and seek other ways to explain this phenomenon.

The way that I pursue is to suppose that nominative OF is generally dispreferred if the fronted element is associated with a focus discourse-function, but it is much more acceptable as a contrastive topic.

To support this, two experiments were conducted. 10 male native speakers were invited to a studio and they had to read out sentences. There were 78 sentences, out of which 30 contained a fronted subject, 15 in the accusative and 15 in the nominative case. 30 of the remaining sentences contained fronted objects (which always bear accusative case for obvious reasons) or contained no OF at all. The sentences appeared one by one in a PowerPoint presentation and participants were instructed to say the sentences in the way that is most natural to them. There was no indication in the text layout about the discourse function of the fronted element (e.g. capital letters or underlining). A minimal pair example from the test is shown in (81).

(81) a *Béla* *mondtad,* *hogy* *megy külföldre.*
Béla.NOM said.2SG that(C) goes abroad.to
‘(Of) Béla you said that he is going abroad.

b *Bélát* *mondtad,* *hogy* *megy külföldre.*
Béla.ACC said.2SG that(C) goes abroad.to
‘(Of) Béla you said that he is going abroad.

The pronunciations of the participants were recorded. Then I listened to the recordings and based on my judgment and help of a program called Speech Analyzer, I determined where the speakers put the primary stress, which I assumed was marking the focus of the sentences. There was a very pronounced difference between the intonation of the sentences containing accusative-marked and nominative-marked fronted embedded subjects. The results are shown in Figure 2a and 2b.

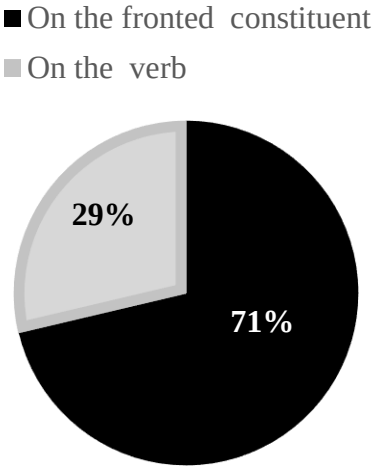

Figure 2a.
Primary stress in sentences with a fronted ACC-marked subject.

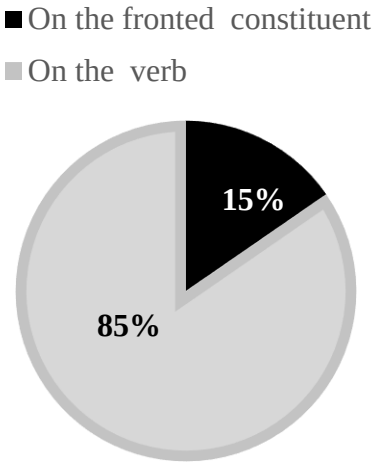

Figure 2b.
Primary stress in sentences with a fronted NOM-marked subject.

As can be seen, conforming to the hypothesis, when the fronted element was marked accusative, primary stress was put on it in 71% of the cases, creating a focus-interpretation. On the other hand, when the fronted phrase was marked nominative it only received primary stress in 15% of the cases. In this latter case, the stress primarily appeared on the verb itself, creating a contrastive topic reading (like “of John, I DID say that he is coming”).

To test the issue further, I investigated what happens if a preverb is put in between the fronted constituent and the verb. As was discussed in Chapter 3, Hungarian has immediately preverbal focus, so focus-interpretation of the fronted element is impossible in this case.

(82) *János(t) meg-mondtam, hogy jön a partira.*
John(.ACC) PV-said.1SG that(C) comes the party.to
'John, I did say that he will come to the party.'

If there is an association nominative-marking and contrastive topic interpretation, one would expect that sentences like (82) should be judged better if the marking is nominative, since focus interpretation is ruled out.

Following this line of reasoning, I conducted a small-scale survey. 15 native speakers were tested, using an online survey. There were 46 sentences, out of which 10 were distractors. The remaining 36 sentences included the following configurations. As indicated, sentence stress was shown with capital letters. The instructions were explicit about this.

(i) 6 accusative marked fronted subjects, focus marking on the fronted constituent (e.g. *JÁNOST mondtad, hogy jön*.) → "Acc F"
(ii) 4[72] accusative marked fronted subjects, focus marking on the verb (e.g. *Jánost MONDTAD, hogy jön*.) → "Acc V-foc"
(iii) 6 nominative marked fronted subjects, focus marking on the fronted constituent (e.g. *JÁNOS mondtad, hogy jön*.) → "Nom F"
(iv) 8 nominative marked fronted subjects, focus marking on the verb (e.g. *János MONDTAD, hogy jön*.) → "Nom V-foc"
(v) 6 accusative marked fronted subjects, focus marking on the fronted verb and preverb on the verb (e.g. *Jánost MEGMONDTAD, hogy jön*.) → "Acc Pv"
(vi) 6 nominative marked fronted subjects, focus marking on the fronted verb and preverb on the verb (e.g. *János MEGMONDTAD, hogy jön*.) → "Nom Pv"

The participants had to rate the sentences' acceptability from 1 (worst) to 5 (best). Here are the mean scores (Figure 3).

[72] The assymetry involving types (ii) and (iv) is due to a mistake in the preparation in the test sentences.

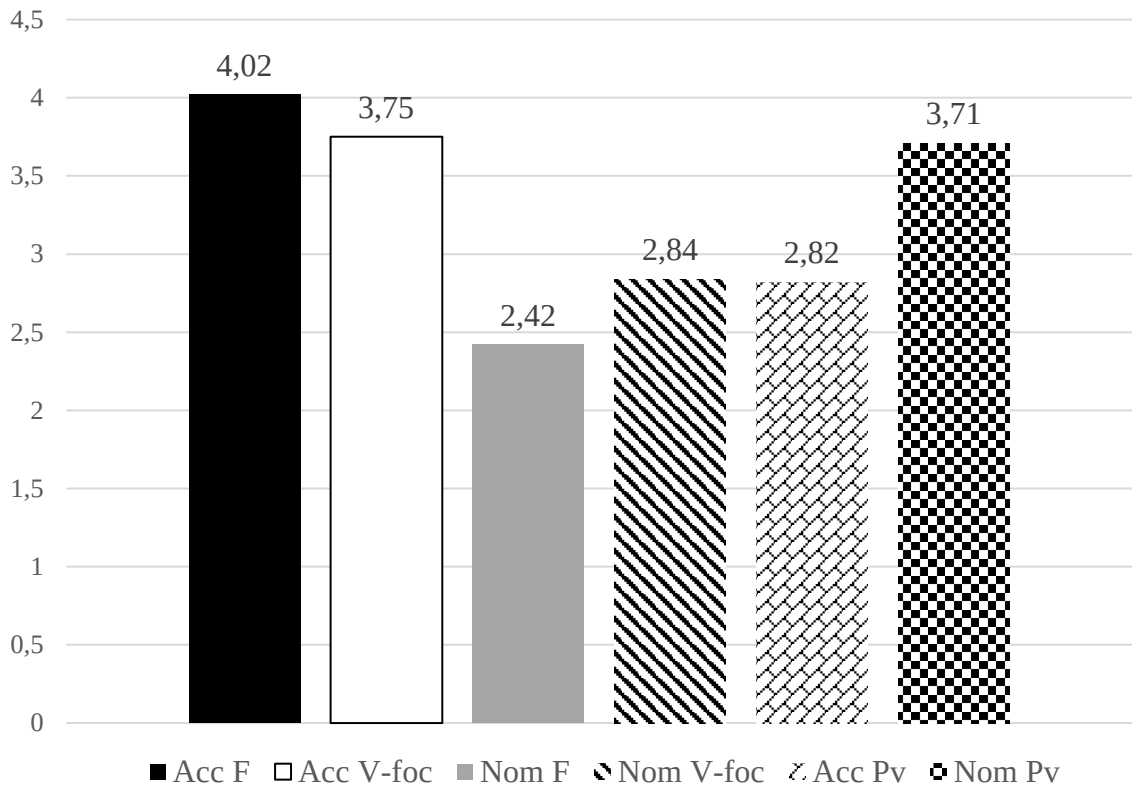

Figure 3.
Grammaticality judgments of fronted ACC/NOM subjects, with/without preverbs.

As we can see, fronted accusative elements were judged to be acceptable in either a focus or contrastive topic interpretation. Nominative fronted subjects are generally bad, but as expected they are slightly better if the verb is stressed, so the nominative element is not a focus. The last two columns are the most important from the perspective of the issue at hand. They show that when the preverb is present, so a topic-interpretation of the fronted element is forced, the sentences with ACC-marked fronted subjects are degraded. On the other hand, this is the configuration where nominative-marked fronted elements are most acceptable. That is, speakers do accept this version of OF, when the Contrastive topic interpretation is clearly accessible. So fronted subjects in their original, nominative case have a strong tendency toward being interpreted as contrastive topics, while ones in the accusative can more easily bear both focus and topic functions (though there is a preference towards the focus-readings). Although no testing has been performed on OBLs, OBJs, ADJUNCTs, my intuition is that they pattern with accusative-marked fronted subjects. In other words, I take them to lack the strong association with topic-interpretation of fronted nominative subjects. I would like to add that what we are talking about here are tendencies and the degree of these preferences may be subject to individual differences. What is certain is that the interpretation of OF is tied to D-STUCTURING functions except the neutral Topic category.

5.2.2 Previous accounts of Operator Fronting

In this section, I will provide a critical overview of the existing literature on Operator Fronting. Most of the work has been done in Chomskyan frameworks. Instead of going into the technical details of these analyses, I will focus on the general underlying analytical ideas and try to present them in a theory-neutral manner. Coppock (2003) is the only account of Hungarian OF couched in LFG (apart from my work), so her analysis is of particular importance for me. Overall, I will argue that the existing accounts are not satisfactory, they all fall short on either empirical or theoretical grounds.

Kenesei's (1992/1994) account (couched in a Government and Binding Theory) is based on his view of Hungarian subordinate clauses where the associate pronoun of Hungarian subordinate clauses is an expletive and it forms a chain with the subordinate clause. I have reviewed this extensively in section 3.2.2.

When OF takes place the fronted constituent originates in the subordinate clause, where it assumes its original case, then it moves to the main clause, to the position of the pronoun, also assuming the pronoun's case. This explains why the pronoun cannot be present in OF sentences. Furthermore, a case hierarchy is assumed (NOM<ACC<OBL), where stronger cases may suppress weaker ones. So if the subject of an embedded clause (bearing nominative case) moves to the position of an accusative main clause pronoun, the subject ultimately gets accusative case (83a). However, if the oblique argument is raised to the main clause, it retains its original case (83b). Nevertheless, in this approach it still lands in the position of the accusative pronoun, barring its appearance.

(83) a *(*Az-t)* *János-t* *mondtad,* *hogy* *megy* *Párizsba.*
that-ACC John-ACC said.2SG that(C) goes Paris.to
'John you said that he will go to Paris.'

b *(*Az-t)* *Párizsba* *mondtad,* *hogy* *megy* *János.*
that-ACC Paris.to said.2SG that(C) goes John.
'To Paris you said that John goes.'

It follows from this account that nominative case should never occur on the fronted constituent, since it is weaker than the accusative given by the main verb. However, later research (Gervain 2002, Jánosi 2013) proved beyond doubt that such a configuration is possible.

The account also leaves unexplained why idiom chunks cannot participate in case-switched OF. In comparable cases of English raising structures, idiom chunks are possible, since prior to movement, they form one constituent with the rest of the idiom.

(84) *I believe the cat to be out of the bag.* (“I believe that the secret is revealed.”)

One might argue that it is only because the fronted element gets a focus discourse function that idiom chunks are ungrammatical. However, if the fronted subject is postverbal on standard assumptions about Hungarian, the fronted constituent does not get a topic or focus discourse-function. Still, case-switched OF is bad.

(85) *#(De hiszen) te mondtad a gépszíjat, hogy elkapta*
but you said.2SG the driving belt.ACC that(C) caught.3SG
Jánost.
John.ACC
‘But you said about the driving belt that it has caught John. / *But you said that John has to work a lot lately.’

Also, it must be noted that Kenesei’s (1992/1994) account hinges on the idea that the associate pronoun of subordinate clauses in Hungarian is a non-thematic element, an expletive. However, as I have demonstrated in 3.2.2, a number of factors argue against this view.

Lipták (1998) builds on Kenesei’s ideas and develops it further in a Minimalist framework. She restricts her account in two ways. First, she only discusses fronted elements with a focus interpretation, because she regards those elements in this construction that have a topic discourse function as an entirely different construction. Second, she excludes fronted nominative subjects even if they have a focus interpretation. She does so because she claims that such sentences are not grammatical.

Her analysis distinguishes between two kinds of OF: one with a case-switch (“reagreement”, in Lipták’s terminology) and one without. The first includes fronted subjects bearing accusative case, (86a). The second includes fronted obliques like (86b).

(86) a *János-t mondtad, hogy jön a partira.*
John-ACC said.2SG that(C) comes the party.to
‘(Of) John you said that he will come to the party’

b *Párizsba mondtad, hogy mész.*
Paris.into said.2SG that(C) go.2SG
‘To Paris you said that you will go.’

Lipták (1998) assumes that the head of the complement clause (the C head) has case and phi-features. It also bears a +D/N feature, which likens it to nominal projections. These features have to be checked in some way. In the case of (86a), the fronted element is base-generated as the subject of the embedded clause, so it bears a nominative case feature. It also bears an

accusative case, which is going to be important later in the derivation. The nominative is checked in the embedded clause. It also bears a +D/N feature, so when it moves through Spec/CP, it can check this feature on the C head. Then it moves to the object position of the matrix clause (AgrOP), where it checks the accusative. It finally moves to the focus position of the main clause (Spec, FP).

If an oblique or an adjunct is fronted (86b), it is assumed to have a –D/N feature. The oblique in (86b) is base-generated in the embedded clause and checks its inherent case there. Then it moves through Spec/CP. Since it is –D/N, it does not check this case on the C head. Then the fronted element moves to Spec, FP to get focus interpretation. The +D/N feature of the C head is checked by the whole embedded clause itself moving to the matrix AgrOP at LF, and this ensures definite agreement on the matrix verb. In both versions of OF, the fronted element is base-generated in the embedded clause and during the derivation it passes through Spec/CP, which explains the obligatoriness of the complementizer.

The problems with this approach are similar to the ones with Kenesei's (1992/1994) account. The nominative version of OF is not explained, nor is the possible plural downstairs agreement of fronted quantified constituents. Also, the impossibility of postverbal idiom-chunks in the construction remains a mystery.

There are also two theory-internal problems with Lipták's (1998) account. Firstly, positing a D/N feature for the fronted element and for the C head happens in an arbitrary way, motivated only by the goal of describing the phenomena. Without further, external evidence the presence of these features is only restating the problem, using an artificial device to express the different behavior of fronted constituents with structural and inherent case. Secondly, although there have been proposals in this direction (Bejar & Massam 1999), an element checking multiple cases in a derivation is still unusual, as Lipták herself notes it too.

Since both Kenesei's (1992/1994) and Lipták's (1998) accounts are based on the movement some element from the subordinate clause to a main clause case-position, the number variation cannot be accommodated in these systems, as the embedded verb should show strict syntactic agreement with its (subsequently fronted) element. Some other mechanism must also be posited to be possible in OF, and this is one of the main contributions of Gervain (2002).[73]

She builds on and also expands upon the previous approaches. There are several empirical developments in her work. As was discussed earlier, she discovered the possible number-variation on the embedded verb. She also investigated the relationship of this variation with case. As was mentioned in the previous section, for Gervain, there are two groups of

[73] The account is further elaborated in Gervain (2004) and (2009).

speakers. One group accepts both plural and singular agreement on the embedded verb, but only accusative case on the fronted element. The other group accepts both nominative and accusative case, but only with singular agreement.

In Gervain's (2002) account these groups are the results of two derivational strategies for fronting embedded subjects. The first group uses a strategy where the fronted element is base-generated in the matrix clause and is co-indexed with a resumptive pronoun in the embedded clause. So in this derivation, no movement from the embedded clause to the matrix clause occurs. In this account, the number variation shown earlier (repeated here as 87) is given two possible explanations (categorial labels are from those of Gervain's).

(87) *Az [$_{NumP}$ összes [$_{NP}$ lányt]] mondtad, hogy jön/ jönnek.*
the all girl.ACC said.2SG that(C) come.3SG come.3PL
'(Of) every the girls you said that they will come.'

One is that the dependency may be either syntactic or semantic in nature. In the first case, the head and tail of the dependency share their φ-features, resulting in *jön* ('come.3SG'). In the second case, notional agreement occurs between the syntactically singular but semantically plural antecedent and the resumptive pronoun, like in a standard anaphoric dependency.

An alternative is that the resumptive pronoun may be co-indexed with either the entire NumP, yielding plural agreement or just the NP itself, resulting in singular agreement. To me the first one seems to be a better explanation, as quantifiers in Hungarian generally do not trigger plural agreement in Hungarian, so it would be mysterious why they should do so in OF.

(88) *az összes/ két lány/ *lányok*
the every two girl girls
'every girl/ two girls'

The second speaker group uses a movement-strategy, basically along the lines of Lipták (1998). Since the fronted subject is base-generated in the embedded clause, no number variation can surface. The accusative-nominative variation is accounted for by positing a choice between multiple case-checking (resulting in accusative case) or leaving the accusative of the matrix clause unchecked.

According to Gervain (2002), the ultimate reason behind the existence of the two groups is that for the ones who use a resumptive strategy, the complementizer *hogy (*'that(C)') is lexically specified as being "opaque", preventing proper government into the embedded clause.

Since the movement operation would leave a trace/copy, requiring proper government, movement is ruled out and resumption is used as an alternative mechanism.

Gervain (2002) is a significant improvement upon previous approaches as it manages to account for the number and case variation as well. However, it is not without problems. First, as was noted earlier, it is not straightforward that consistent speaker groups can be determined (Jánosi 2013).

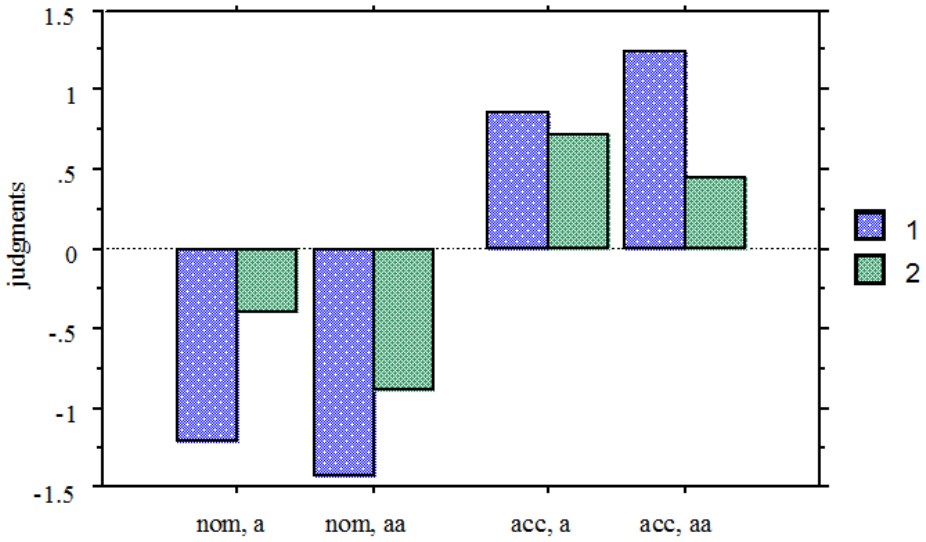

Figure 4.
Sentence-acceptability in Gervain (2002).

An empirical problem is that according to Gervain's (2002) empirical data, although plural downstairs agreement (with a fronted accusative quantified subject) is worse than singular agreement for movement-speakers, it is actually still better than a nominative-marked subject with singular agreement. See, Figure 4, which is a diagram of Gervain's empirical survey ("aa" stands for "anti-agreement", Gervain's term form downstairs plural agreement in case of quantified accusative fronted subjects, "a" is normal agreement, 1 is the resumption group, 2 is the movement group). This is a strange pattern, as movement-speakers should rule plural agreement out entirely.

There are also some problems caused by the theoretical machinery posited by Gervain (2002). The accusative case of the main verb which remains unchecked if the fronted element bears accusative case should cause the derivation to crash. However, in Gervain's analysis it just worsens the acceptability.

The other theory-internal problem is that if for the resumption-group, the complementizer is "opaque", there remains no room to derive sentences where an oblique or an adjunct is fronted, with their original case. That is, this group should derive every type of OF with resumption, but there is no sign of embedded resumptives in the case of obliques and

adjuncts, so these must be derived via movement (assuming a Minimalist framework). Since this is ruled out by the opacity of the complementizer, this explanation is too restrictive, barring existing structures.

Finally, let us take a look at Coppock (2003). This is an adaptation of Gervain's (2002) account to LFG. The primary concerns of her account are fronted subjects. She subscribes to the expletive-associate-chain analysis of subordinate clauses (Kenesei 1992/1994) and consequently she analyzes fronted accusative-marked subjects as athematic arguments of the main verb, just like an expletive is athematic. So she claims that the lexical entry of a verb that participates in OF looks like (89), where the object function is located outside the angle brackets, indicating it being a non-thematic argument.

(89) *mond* <(SUBJ)(COMP)> (OBJ)

The two speaker groups identified by Gervain (2002) (movement- and resumption-speakers) are modelled by positing that the groups use different mechanisms to unify the athematic matrix object with the appropriate thematic function in the embedded clause: the movement-group uses functional identification, while the resumption-group utilizes anaphoric identification. Recall from section 2.3 that the essential difference between the two mechanisms is that functional identification means strict f-structural identity, so a single element is shared by two sub-f-structures, while anaphoric identification is a looser, semantic kind of relationship, like a coreference of two distinct elements. Functional identification requires fully identical grammatical features (so no mismatching agreement is allowed downstairs), while anaphoric identification may allow for a number mismatch. The two approaches are schematically illustrated in Figure 5a and 5b (for simplicity, the discourse functions are omitted).

(90) a *Az összes lányt mondtad, hogy jön/ jönnek.*
the every girl.ACC said.2SG that(C) come.3SG come.3PL
'Every girl you said that they will come.'

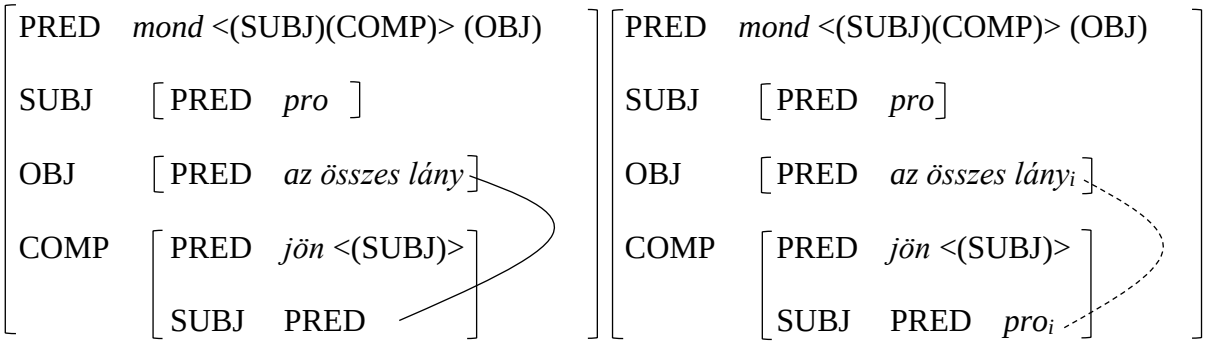

Figure 5a. **Figure 5b.**

OF with fuctional (5a) and anaphoric (5b) identification in Coppock (2003)

Notice that Figure 5b violates the Semantic Coherence condition (discussed in section 2.3), as the matrix OBJ is a meaningful element (*every girl*), but it is not linked to any thematic grammatical function (anaphoric identification is not sufficient). To address this issue, Coppock (2003) theorizes that the condition is just an Optimality Theory-style constraint, which may be violated if it conflicts with some higher ranked constraint. This constraint is identified as a ban on subject extraction, in effect a ban on fronting nominative marked subjects. So there are two constraints "Semantic Coherence" and "*Subject Extraction", and the two speaker-groups have different rankings. The "movement"-group ranks Semantic Coherence higher, so they cannot allow for anaphoric identification, thus they can only have the f-structure in Figure 5a. In return they tolerate an extracted subject. The resumption-group on the other hand has the reverse constraint-ordering. They tolerate the violation of Semantic Coherence (so they allow the f-structure in Figure 5b), but they do not permit an extracted subject. (With the extraction-constraint restricted to subjects, she avoids the problem of barring existing structures.)

As for c-structure and ID-rules, Coppock (2003) argues that the fronted element sits in the specifier of a focus projection (FP), having the grammaticalized discourse function FOCUS. Also, this position is specified as being functionally identified with an embedded grammatical function. Furthermore, it is specified that any element in Spec/FP, bearing accusative case is interpreted as the object of the main clause. Coppock's annotated ID rule is given in (91).

(91) FP → NP F'
(↑FOCUS)=↓ ↑=↓
(↑COMP* GF)=↓
(↓CASE=ACC) → (↑OBJ=↓)

Like in the previous accounts, the impossibility of postverbal idiom chunks is still unexplained. Furthermore, to accommodate the fact that not just embedded subjects may participate in case-

switching OF but objects, obliques and adjuncts too, Coppock would actually have to modify the extraction-ban to include all GFs, but then this would block all case-retaining OFs as well. But then this would lead to over-constraining like in the case of Gervain (2002).

However, the fundamental objection that may be raised against Coppock (2003) is the downgrading of Semantic Coherence. The well-formedness conditions are fundamental in LFG, so modifying them to be violable seems a pretty strong move, especially considering that this is done on the basis of a single construction, as no other phenomena has been shown to violate the (Semantic) Coherence condition. If an alternative solution exists, it must be preferred. The alternative is regarding the object as thematic in Figure 5, and this is exactly the analytical idea that is going to be taken in this dissertation. This is to be presented in the next section. [74]

5.2.3 The analysis of Operator Fronting in Hungarian

Let us repeat the basic examples for OF, for which an analysis is to be presented in this section.

(92) a *János(-t) mondtad, hogy jön.*
John(-ACC) said.2SG that(C) comes
'(Of) John you said that he will come.'

b *János-ban bízok, hogy jön.*
John.in trust .1SG that(C) comes
'In John trust that he comes.'

c *Párizsba mondtad, hogy mész.*
Paris.into said.2SG that(C) go.2SG
'To Paris you said that you will go.'

We can divide the analysis of OF into two main parts. The first is the cases where no case-switching happens, like the nominative version of (92a), where the subject of the embedded clause is fronted or (92c), where an oblique argument occupies the fronted position. On the other hand, there are cases where the fronted element becomes the argument of the main verb, like the accusative version of (92a). Sentences when the fronted element is assigned some oblique case like in (92b) also belong to this pattern. This mostly happens when the fronted argument corresponds to the embedded subject, but other grammatical functions can also be affected, see examples (55c-d, 56).

[74] There is also a c-structure-related problem with Coppock (2003). Although this is not directly related to the analysis of OF, it might be worth mentioning that Coppock (2003) posits an FP (Focus Phrase) functional category. But this is not justified in LFG for Hungarian, as there is no lexical item invariably associated with the F head, so the Economy of Expression is violated. As was shown in 3.1, a satisfactory account can be achieved without such projections if focused elements are assumed to be in Spec/VP, following (Laczkó) 2014.

Of these two cases, the first one has a more straightforward analysis. These cases are "fronting" constructions in a more traditional sense, so *János* ('John') and *Párizsba* ('to Paris) in (92) functionally and thematically belong to the embedded clause, the only trait that ties them to the main clause is their position. Such configurations are sensitive to factors that are linked to movement-like dependencies in Chomskyan frameworks: they are ungrammatical if the embedded function associated with the fronted element is in an island, they are also bad in crossover, they show reconstruction effects and the main verb must be one that has the bridge quality. Although there is no such thing as movement in LFG, the framework of this dissertation, functional identification is a tool that is a close nonderivational parallel to it, as already seen with TOP-Eng.

The case-switching scenario, where some element associated with an embedded function actually becomes the object or an oblique of the main verb, is more complex from a theoretical point of view. In many respects its properties are the direct opposites of the case-retaining version: it is unaffected by islands, crossover and by the non-bridge quality of the main verb and it shows no reconstruction effects. In Chomskyan frameworks, such properties have warranted a base-generation analysis (as in Gervain 2002). Here in an LFG framework I argue that this version of OF should be viewed as "prolepsis", a type of dependency posited for a number of similar phenomena in the literature. I will give a typological and a theoretical overview of proleptic and related structures in Chapter 6, here let us just state the basic definition, as worded by Salzmann (to appear:1): prolepsis is a construction where a structural complement of the matrix verb is semantically related to the predicate of a finite embedded clause. This definition leaves the details of the analysis unspecified, and rightly does so, since beyond crucial similarities there are also important differences in the cross-linguistic realizations of prolepsis. In the Hungarian case, my position is that the proleptic constituent should be analyzed as a thematic argument of the main verb, which is associated with the embedded function via obligatory anaphoric identification.

Let us now go into the analysis of OF. As discussed in section 5.2.1, fronted elements in OF may bear a range of discourse functions: information/contrastive focus, contrastive topic, although there are preferences for certain discourse functions depending on the function of the fronted constituent. Nominative-marked phrases tend to be contrastive topics, while accusative and oblique-marked elements prefer a focus interpretation. What this means is that in OF, the fronted elements may occur under the S node in the topic field or in Spec/VP.

As claimed in the previous section, in cases where no "case-switch" takes place, the dependency between the fronted element and the function associated with it in the embedded clause should be analyzed as functional identification. These are the cases where a subject, an

object or an oblique argument is fronted and the original case is retained. To model this, we must add the following optional specifications to the relevant nodes:

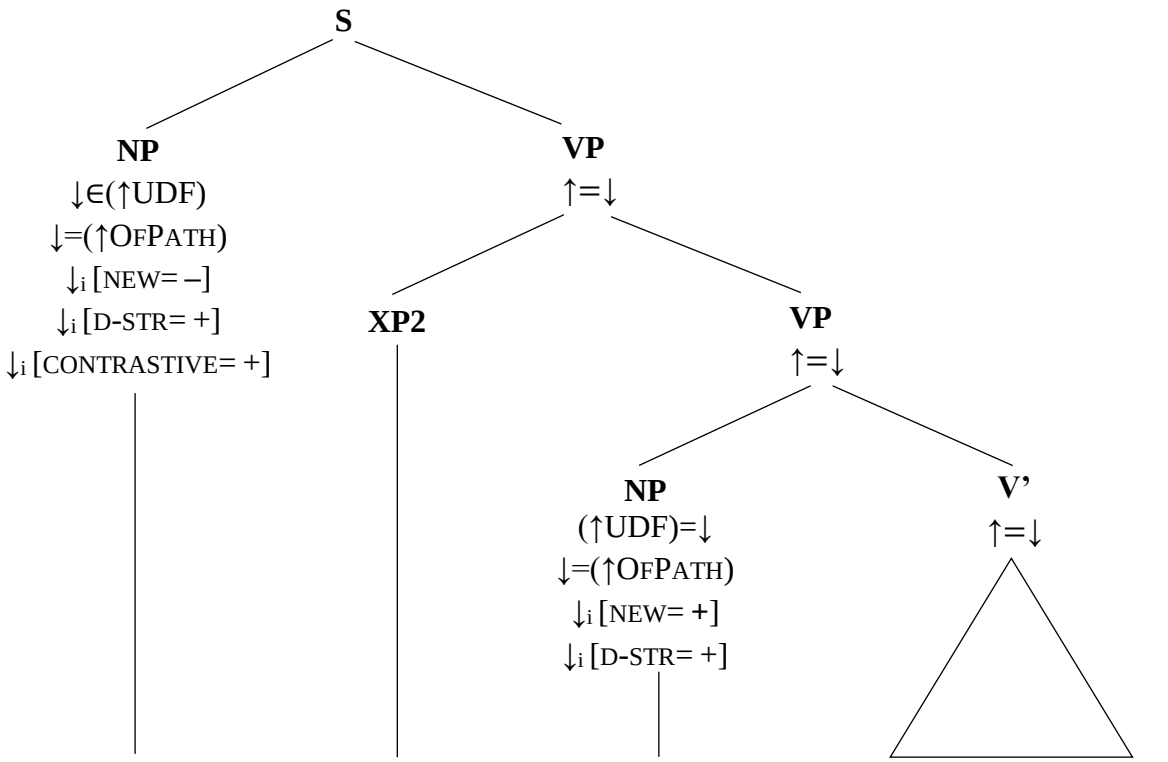

Figure 3.
C-structure for OF in case-retaining scenario.

These annotations encode that a constituent occupying these positions can be identical with some grammatical function in an embedded clause. Following the convention adopted in 4.1, I label the path to the embedded function as OFPATH and its details are shown in (94).

(94) OFPATH ≡ {OBJ | SUBJ}$^{+}$ GF
(→TENSE)
(→LDD ≠ –)

(94) indicates that the fronted element may have any grammatical function in an arbitrarily deeply embedded complement clause. Recall from 3.2 that complement clauses are subject or object arguments of the matrix predicates if there is no associate pronoun. The $^{+}$ notation is called a "Kleene-plus" and it means "at least one", so it ensures that embedding does take place. These objects will have tense features (as they are finite clauses), as opposed to scenarios where the SUBJ or OBJ is a lexical noun or pronoun associated with an embedded clause (where the

clause itself is an ADJ). In such cases OF will not take place, so CNPC- and Adjunct-islands will be ruled out. The annotation also requires every predicate along the path to have a positively specified LDD feature (that is, the predicates must have bridge-quality).

In accordance with the interpretation of OF, the appropriate features are contributed to these elements at information structure. This ensures that depending on their position, the Operator fronted constituents are interpreted as IF/CF/CT.

As discussed, speakers may vary according to their preferences of these positions. The general pattern is that nominative fronted embedded subjects are strongly preferred as contrastive topics, while nonsubject functions may be any D-STRUCTURING category except for neutral Topics. I think that rather than positing syntactic restrictions, invoking general mapping principles between grammatical function and IS-function could give us a simpler reason for this pattern. There is a strong cross-linguistic tendency for the association of subjects and topicality, since subjects are default Topics. As such, it is no surprise that even when fronted they prefer configurations in which they are interpreted as topic-like entities. Since the neutral topic category is unavailable for fronted elements, the contrastive topic discourse function is the closest match. Other grammatical functions do not possess this association, so they can easily accommodate either discourse function. I offer no formal treatment of the preferences here, but combining LFG with Optimality Theory could be a fruitful research avenue for this topic. Such a treatment would allow for the specification of constraints expressing the preference patterns, so a mapping like SUBJ-CT would be more optimal than SUBJ-IF/CF. Taking the focus-interpretation as an example, Figure 4 shows the f-structure corresponding to (93).

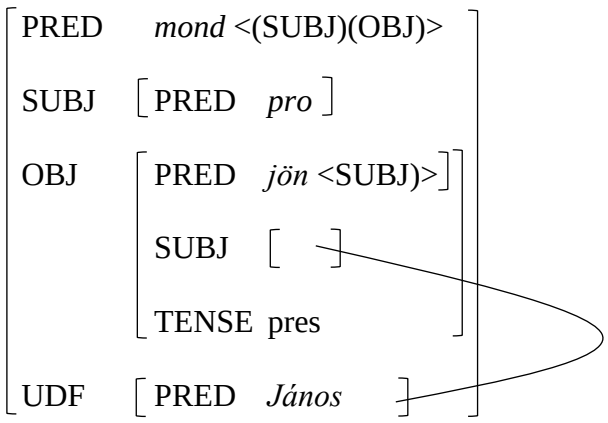

Figure 4.
F-structure corresponding to (93).

In accordance with our discussion in 3.2, I take *mond* ('say') and similar verbs to subcategorize for a SUBJ and an OBJ grammatical function. The OBJ can be a lexical DP/NP, a pronoun, a bare S or a CP-clause as is the case is in this particular example.

However, if the object argument is expressed as a pronoun, the clause itself is an adjunct. As such, it is an island, which means that functional identification into it is not licensed by (94). So a pronoun cannot surface here, as (95) shows.

(95) **János mondtad azt, hogy jön.*
John said.2SG that.ACC that(C) comes

Similarly, non-bridge verbs, having a negatively specified LDD-feature, block functional identification, as in (96).

(96) **János álmodtam, hogy jön.*
John dreamed.1SG that(C) comes

Since the fronted element is via functional identification fully present in both the matrix and the embedded f-structure, reconstruction effects are explained. Also, since functional identification means complete syntactic identity, the lack of potential number-variation in the embedded clause is also expected. An example is repeated here for convenience.

(97) *Az összes lány mondtam, hogy jön/ *jönnek.*
the every girl said.1SG that(C) come.3SG come.3PL
'Every girl I said that they will come.'

The other scenario, in which the fronted element actually becomes the argument of the main predicate, requires more consideration, especially from a thematic, argument-structural point of view. This is because, in this case, the lexical entry for the relevant main predicates is augmented. Consider (98).

(98) a *Én Jánost mondtam, *(hogy jön).*
I John.ACC said.1SG that(C) comes
Intended: 'Of John I said that he comes.'

b *Én Jánosban bízok, *(hogy jön).*[75]*
I John.in trust.1SG that(C) comes
Intended: 'About John I believe that he comes.'

Without context, the *that*-clause is not omissible. However, so far I have treated *mond* ('say') and similar verbs as subcategorizing for two grammatical functions, a SUBJ and an OBJ. If this is the case, it seems strange that the *that*-clause is compulsory, given that *I* and *John* satisfy these requirements already. In order to account for this, I posit that via an argument-structural operation, the lexical entry of certain predicates (those ones that participate in this pattern) is altered, resulting in the modification of the aforementioned <(SUBJ)(OBJ)> frame to a frame like the one in (99).

(99) a *mond* ('say'), *gondol* ('think), *hisz* ('believe'), etc. <(SUBJ)(OBJ)(COMP)>
b *bízik* ('trust'), *tart* ('be.afraid'), *számol* ('expect'), etc. <(SUBJ)(OBL)(COMP)>

In order to be able to analyze the modification of the argument structure of these predicates, we first have to make some notes on the initial mapping. I have adopted Kibort's (2007) version of Lexical Mapping Theory (LMT) in section 2.4, so that is going to be the frame of the discussion here. For convenience, I repeat the featural decomposition of grammatical functions, the universal valency frame and the mapping principles here.

	–o	+o
–r	SUBJ	OBJ
+r	OBL	OBJΘ

Table 1.
Grammatical functions in LMT.

(100) Universal valency frame: –o/–r –r +o –o … –o
arg1 arg2 arg3 arg4 arg_n

[75] Since *bízik* ('trust') may easily take a bare OBL argument (98b) is grammatical without the *that*-clause, but not in the relevant sense. Without the *that*-clause (98b) means that "I trust in John" (he is a trustworthy person), while the intended reading is that I have a trust in connection with him, that he will come. The distinction is easier to see if the OBL is replaced with something that is not trustable in the first sense, like an inanimate object. In that case, the irrelevant reading is semantically anomalous, while the intended one is not.

(i) *Én az új virágokban bízok, #(hogy megszépítik a kertet).*
I the new flowers.in trust.1SG that(C) make.nice.3SG the garden.ACC
'I have a trust regarding the new flowers, that they will make the garden nice.'

(101) **Mapping Principle**: The ordered arguments are mapped onto the highest (i.e. least marked) compatible function on the markedness hierarchy. (markedness: having positive feature-specification)
Markedness hierarchy: SUBJ (–o/–r) > OBJ (–r/+o), OBLθ (–o/+r) > OBJθ (+o/+r)

Let us take a look at predicates requiring an object complement (98a). These verbs semantically subcategorize for an agent and theme/propositional argument (represented in 102a-a'). In principle, the <(SUBJ)(OBJ)> subcategorization can follow from two mappings of these arguments onto the valency template (102b-b'):

(102)	a	predicate	ag	prop		a'	predicate	ag	prop
	b		–r	+o		b'		–o	–r
			arg1	arg3				arg1	arg2
	c		SUBJ	OBJ		c'		SUBJ	OBJ

Either of these configurations is in harmony with the mapping principles and yields the desired outcome. However, there is evidence from nominalization that the actual mapping is the second one. The +o specification in (102a) would make it impossible to nominalize the object, as the resulting nominal would have the POSS function, which is –o, according to Laczkó 1995. However, this prediction is not borne out, as (103)-(105) shows.

(103) a *Azt mondtad, hogy tökéletes vagy.*
that.ACC said.2SG that(C) perfect be.2SG
'You said that you are perfect.'

b *Annak az állandó mondása, hogy tökéletes vagy, nem szép.*
that.POS the constant say.NOUN that(C) perfect be.2SG not nice
approx. 'The constant saying of you being perfect is not nice.'

(104) a *Azt bizonyították, hogy bűntény történt.*
that.ACC proved.3PL that(C) crime happened
'They proved that a crime had happened.'

b *Annak a bizonyítása, hogy bűntény történt, könnyű volt.*
that.POS the prove.NOUN that(C) crime happened.3SG easy was
'Proving that a crime happened was easy.

(105) a *Azt firtatták, hogy milyen a magánéletem.*
that.ACC asked.3SG that(C) how the personal.life.POSS.1SG
They asked me about my personal life.'

b *Annak a firtatása, hogy milyen a magánéletem, bosszant.*
that.POSS the ask.NOUN that(C) how the personal.life.POSS.1SG annoys
'Asking about my personal life annoys me.'

So what happens that results in the modification of a subcategorization frame like (102) to something like (99)? The fronted element becomes the OBJ of the main verb, but its behavior is different from the OBJ of the basic template. Unlike that, it cannot be nominalized, compare (103-105) with (106-108).

(106) a *Jánost mondtad, hogy jön.*
John.ACC said.2SG that(C) comes
'John you said that he will come.'

b **János állandó mondása, hogy jön nem szép dolog.*
John constant say.NOUN that(C) comes not nice thing
intended, approx.: 'The constant saying of John that he will come is not a nice thing.

(107) a *?Jánost bizonyították, hogy jön.*
John.ACC proved.3PL that(C) comes
'John they proved that he will come'

b **János bizonyítása, hogy jön, könnyű volt.*
John prove.NOUN that(C) comes easy was
intended, approx.: 'Proving John that he will come was easy.'

(108) a *Jánost firtatták, hogy jön-e.*
John.ACC asked.3PL that(C) comes-QUESTION PRT
'(Of) John they asked if he comes.'

b **János firtatása, hogy jön-e bosszant engem.*
John ask.NOUN that(C) comes-QUESTION PRT annoys me
intended, approx.: 'The asking of John if he comes annoys me.'

This suggests that the newly added object is specified as +o, which means that it cannot map to the POSS function. Now if the extra argument (tentatively labelled as having subject matter thematic role) is simply added to a base template as +o argument, in accordance with the Mapping Principle, the following would be the outcome.

(109)	a	predicate	ag	prop	subject matter
	b		–o	–r	+o
			arg1	arg2	arg3
	c		SUBJ	OBJ	OBJΘ

OBJΘ as a subcategorized grammatical function has not been found to be present in Hungarian, so (109) would present us with an interesting research outcome. However, (109) cannot be right, as it would predict the possibility of the simultaneous occurrence of a normal object and this

added object. But this is not the way OF works. The newly added object can only co-occur with a clause, never with a pronoun or other nominal object.

(110)	**Jánost*	*mondtad*	*[a*	*hírt]/*	*[azt,*	*hogy*	*jön].*
	John.ACC	said.2SG	the	news.ACC	that.ACC	that(C)	comes

It follows that the added object is the sole OBJ argument of the predicate. But then the propositional argument cannot be an OBJ anymore. This is ensured if we posit that when the new argument is added, the proposition is not mapped to the second position of the universal valency frame, but it moves to the fourth, -o position. This is possible in the Kibort (2007)-version of LMT, as morphosemantic alternations are modelled as semantic arguments realigning along the universal valency frame (see the discussion about the dative shift in section 2.4). The result is (111).

(111)	a	predicate	ag	subject matter	prop
	b		–o	+o	–o
			arg1	arg3	arg4
	c		SUBJ	OBJ	COMP

It is interesting to note that Kálmán (2001:115) mentions some rare cases where such objects appear with standard oblique dependents.

(112)	a	*Merre*	*mondták*	*a*	*kocsmát?*
		what.direction	said.3PL	the	pub.ACC
		approx.: ‘What did they say of the pub, where it is?’			
	b	*Hova*	*gondolod*	*a*	*szekrényt?*
		to.where	think.2SG	the	wardrobe.ACC
		approx.: ‘What do you think of the wardrobe, where should it be put?’			

This might lend support for COMPs being analyzed as kinds of OBLs. Case-switching Operator fronting standardly realizes this OBL function as a *that*-clause, but as (112) shows, standard oblique dependents are also possible sometimes.

Now let us take a look at verbs with oblique complements (99b). The initial mapping is shown in (113). Then a similar argument structural process takes place (114). A subject matter argument is newly added to the argument position of the original oblique complement, and simultaneously, the propositional argument moves to the fourth, –o argument position.

(113)	a	predicate	ag	proposition	
	b		–o	+r	
			arg1	arg2	
	c		SUBJ	OBL	

(114)	a	predicate	ag	subject matter	proposition
	b		–o	+r	–o
			arg1	arg2	arg4
	c		SUBJ	OBL	COMP

Another change that happens along with the addition of the new argument is that the lexical entry specifies these entities as obligatorily co-referent with some element in the embedded clause. So taking *mond* ('say') a prototypical verb of this class, its lexical entry is the following.

(115) *mond* <(SUBJ)(OBJ)(COMP)>
OBJ INDEX={COMP^{+} GF*} GF INDEX

As (115) indicates, the coreferent element may have any GF in the immediately following COMP, or it may be embedded somewhere at depth. The * is a "Kleene-star", meaning "zero or more".

What has not been addressed so far is the nature of the extra argument. I propose that the phenomenon that has just been described is an instance of "prolepsis": a construction where "a structural complement of the matrix verb is semantically related to the predicate of a finite embedded clause" (Salzmann (to appear:1). The extra argument is referred to as a "proleptic" one. Theoretical and cross-linguistic perspective on it will be provided in the next chapter. Generally, it should be viewed as a lexical augmentation process, whereby the adicity of certain predicates is expanded with a new argument, bearing a subject matter theta-role. I agree with Kotzoglou and Papangeli (2007), who, building on Pesetsky (1995) and Reinhart (2002), suggest that the proleptic argument is properly characterized as bearing the "subject matter" theta-role. (Kotzoglou and Panangeli 2007 concern themselves with a Greek proleptic construction. I will introduce that and other proleptic structures in the next chapter).

As discussed in section 5.2.1, predicates can accept this extra proleptic argument to different degrees. Verbs subcategorizing for OBLs are generally free. This is probably because obliques are generally semantically closely related to the governing predicates. A reflex of this is their "+restricted" specification in LMT. The semantic connection with objects is weaker, so their occurrence in proleptic OF is more restricted. As a reminder, here is the basic picture:

(i) Verbs that are entirely acceptable in prolepsis: *mond* (‘say’), *gondol* (‘think’), *ígér* (‘promise’) *kérdez* (‘ask’), *remél* (‘hope’), *beszél* (‘speak’) and *mesél* (‘tell’).

(ii) Verbs that are moderately acceptable in prolepsis: *fontolgat* (‘contemplate’), *sejt* (‘surmise), *tud* (‘know’), *sajnál* (‘pity’), *tanácsol* (‘advise’), *gyanít* (‘suspect’).

(iii) Verbs that are ungrammatical in prolepsis: *álmodik* (‘dream’), *hazudik* (‘lie), *válaszol* (‘answer’), *felel* (‘reply’), *terjeszt* (‘spread the news’), *parancsol* (‘order’), *érez* (‘feel’), *suttog* (‘whisper’) or *ordít* (‘shout’).

Can we find some explanation or at least motivation for this pattern? Although I think it ultimately boils down to idiosyncratic lexical properties and individual differences, there are some factors that can be identified as affecting the possibility of proleptic OF.

As we will see in the next chapter in languages with proleptic structures, the proleptic element is often not an argument, but a thematic adjunct, bearing some oblique case-marking. This pattern may be found in Hungarian, too. Verbs that occur with prolepsis almost always also occur with a delative dependent, which intuitively has the same subject matter thematic relationship to the main verb that the proleptic argument has. Verbs that reject OF often do not occur with such an oblique.

(116) a *Azt mondtam/ gondoltam/ ígértem/ meséltem/ gyanítottam*
that.ACC said.1SG thought.1SG promised.1SG told.1SG suspected.1SG
János-ról hogy jön.
John-DEL that(C) comes
‘I said/thought/promised/told/suspect about John that he will come.’

b *Azt éreztem/ utáltam/ tanácsoltam/ parancsoltam/ (*János-ról),*
that.ACC felt.1SG hated.1SG advised.1SG ordered.1SG John-DEL
hogy jön/ jöjjön.
that(C) comes/ come.SBJV.3SG
‘I felt/hated/advised/ordered (*of John) that he (should) come.’

However, the correspondence is far from perfect. Although it seems a solid generalization that verbs that occur with direct object prolepsis can alternatively take a delative dependent, the reverse is not true: there are predicates that can take a delative dependent but are dispreferred in proleptic object OF. The delative dependent is completely optional, so it is in all probability a thematic adjunct of these predicates.

(117) a *Azt terjesztettem/ nyilatkoztam/ válaszoltam/ ?suttogtam*
that.ACC spread.1SG spoke.to.the.press.1SG answered.1SG whispered.1SG
Jánosról, hogy jön.
John.DEL t hat(C) comes
approx.: 'I spread (the news)/ spoke to the press/ answered about John that he will come.

b *???Jánost terjesztettem/ nyilatkoztam/ válaszoltam/ suttogtam,*
John.ACC spread.1SG spoke.to.the.PRESS.1SG answered.1SG whispered.1SG
hogy jön.
that(C) comes
Intended: 'I spread (the news)/spoke to the press/ replied/ whispered of John that he will come.'

So the possible oblique dependent can be seen as a prerequisite for proleptic OF, but other factors are also at play. Another such factor seems to be that the verb should be obligatorily associated with a proposition. Verbs that reject the direct object OF are often such that they have an intransitive counterpart. Verbs that are good with OF tend not to have intransitive versions. For these verbs, the proposition itself seems to be an additional argument, not required by the basic semantics of the verb. Prolepsis is obligatorily associated with a proposition, so it is expected that these predicates do not constitute a proper basis for the argument-structure of prolepsis.

(118) a **János mondott/ gondolt/ ígért/ gyanított.*
John said.3SG thought.3SG promised.3SG suspected.3SG

b *János hazudott/ nyilatkozott/ válaszolt/ suttogott.*
John lied.3SG spoke.to.the.press.3SG answered.3SG whispered.3SG
'John lied/spoke to the press/answered/ whispered.'

A final factor that I would like to mention is that it seems that the verb's flexibility in realizing its propositional argument seems to correlate with its ability to participate in proleptic OF. It was discussed in 3.2 that in Hungarian subordinate clauses, by default, neither the object pronoun nor the complementizer is obligatory.

(119) *(Azt) mondom/ gondolom/ ígérem, (hogy) jövök.*
that.ACC say.1SG think.1SG promise.1SG that(C) come.1SG
'I say/think/promise (that) I come.'

This means that the OBJ argument may be either a lexical noun, a pronoun, a *that*-clause or a *that*-less, bare clause. However, not every verb allows this flexibility. Although almost every verb at issue allows the complementizer to be deleted (with the restrictions discussed earlier), not all of them allow the pronoun to be absent (120a), if there is a propositional CP. Furthermore, only a small subset of this latter group allows the deletion of both the complementizer and the pronoun (120b).[76] As it happens, this last group (*mond* 'say', *gondol* 'think', ígér 'promise', *remél* 'hope', *gyanít* 'suspect') coincides with the best direct object proleptic verbs. (The tense varies in (120) to create pragmatically more plausible sentences, e.g. one cannot dream while uttering a sentence).

(120) a *Mondom/ gondolom/ sérelmezem/ sejtem/ *feleltem/ *képzelem/*
say.1SG think.1SG resent.1SG surmise.1SG replied.1SG imagine.1SG
**álmodtam, hogy János jön.*
dreamed.1SG that(C) John comes.
'I say/think/ resent/ surmis/replied/imagine/dreamed that John will come.

b *Mondom/ gondolom/ ígérem/ remélem/ gyanítom/ *sérelmezem/*
say.1SG think.1SG promise.1SG hope.1SG suspect.1SG resent.1SG
**sejtem János jön.*
surmise.1SG John comes.

Ultimately, I claim it is lexical variation that decides whether the proleptic argument can be added to the lexical entry of a predicate. At least three factors contribute to this potential: the ability to occur with a delative dependent, the obligatory nature of the propositional argument and the categorial flexibility with which the proposition is realized. The more of these factors a verb is positive about, the more likely it is that it will accept the addition of an extra subject matter argument to its lexical frame, which results in proleptic, direct object Operator fronting. The process itself is not without any theoretical background: in 2.4, I mentioned that there seem to exist lexical processes that turn thematic adjuncts into thematic arguments. From this perspective, the proleptic object is a "derived argument" in Toivonen's (2013) sense: an argument that is not part of the initial argument list, but added through an argument structure operation.

This time no additional phrase-structural rules are necessary, since the positions needed independently exist in Hungarian grammar. The proleptic element is the object/oblique

[76] So while the complementizer can be freely deleted if the pronoun is present, once it is absent, the complementizer becomes obligatory for many verbs. The reasons for this are unknown to me. It is possible that investigating this along the line of É. Kiss's (2005) theory of event structure would provide some insights.

argument of the predicate (*mond* in this case), and it can sit in the positions that may be occupied by the respective arguments of these verbs.

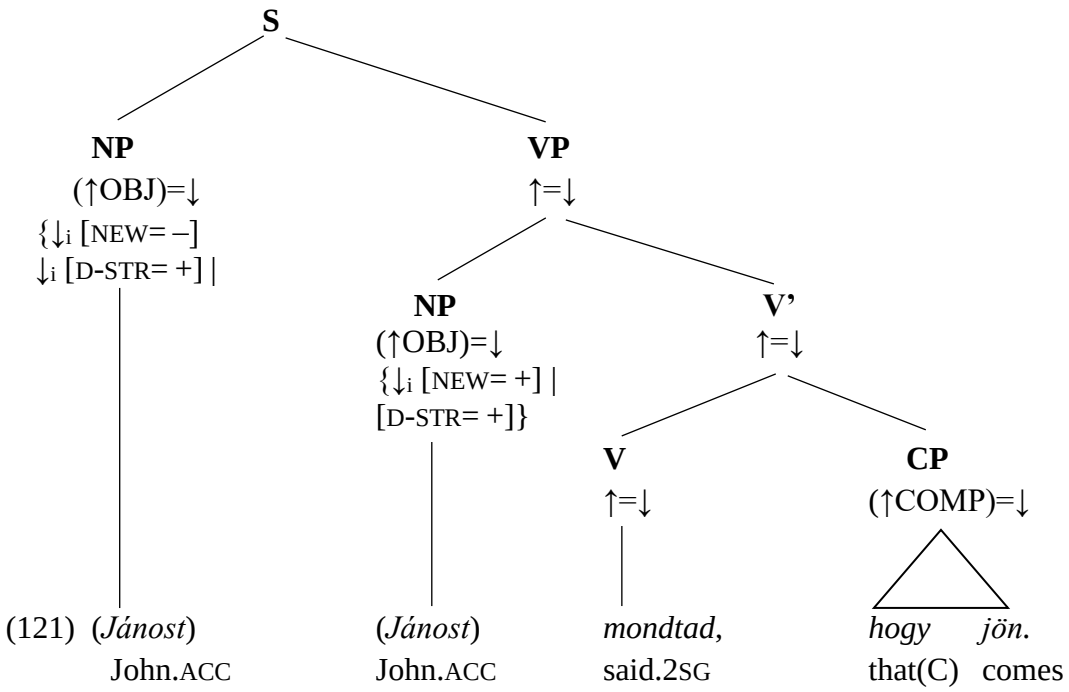

Figure 5.
C-structure for OF in the proleptic scenario.

The corresponding f-structure is shown in Figure 6.

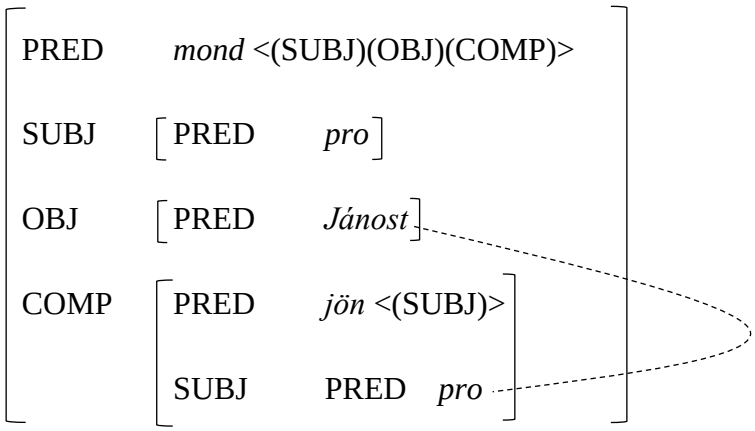

Figure 6.
F-structure for (121)

So crucially, the proleptic object is the thematic argument of the verb, *mond* ('say') in (112). Let us summarize the arguments for this. First, recall from section 5.2.2 that Coppock's (2003) analysis posited that the proleptic element is non-thematic, but this leads to a violation of Semantic Coherence, as a meaningful element (the proleptic direct object) would not be

associated with a thematic argument slot of any predicate. An alternative analysis, if possible, is preferable to the suspension of a basic LFG principle.

Second, the proleptic element is in complementary distribution with the associate pronoun of demonstrative clause and I have argued extensively in section 2.2.2 that the pronoun is also a thematic argument and not an expletive. While in principle, a proleptic analysis could be posited even under the expletive-analysis, the plausibility is greater if the thematic status of the OBJ-argument is uniform in the argument-structural variations of the predicates. Also, this makes the analysis of object and oblique proleptic elements uniform, as the latter are to be analyzed as arguments anyway.

Third, OF is incompatible with idiom chunks, even when they are postverbal, see (122).

(122) a *#(De hiszen) te mondtad a gépszíjat, hogy elkapta Jánost.*
but you said.2SG the driving belt.ACC that(C) caught John.ACC
'But you said about the driving belt that it has caught John. / *But you said that John has to work a lot lately.'

Finally, in certain contexts, the complement clause may be elided (123a). This is comparable to the English examples in (123b-c), based on Bresnan (1982). *Persuade* takes a thematic object, while *believe* is a raising verb, taking an nonthematic object. So in (123b) *John* is thematically independently licensed in the main clause, while in (123c) it is left without a thematic role, causing ungrammaticality. That OF patterns with *persuade* lends support for the thematic object analysis of the proleptic element.

(123) a A: *Végül Péter jött.*
finally Peter came.3SG
'Finally, Peter came.'
B: *De hiszen te Jánost mondtad!*
but you John.ACC said.2SG
Intended: 'But you said (of) John (that he would come).'

b *Someone had to wash my car. I persuaded John (to wash my car).*

c *Someone stole my car. I believed John *(to have stolen my car).*

Since the proleptic element is linked to the embedded grammatical function via an anaphoric link and not functional identity, a compulsory matching of grammatical features is not expected, so there can be space for the number variation of the kind shown earlier. We will see in section 6.3 that anaphoric identification often leads to number mismatches in similar setups.

(124) *Az összes lányt mondtad, hogy jön/ jönnek.*
the all girl.ACC said.2SG that(C) come.3SG come.3PL
'All the girls you said that they are coming.'

Since the proleptic element is a genuine main clause entity, only semantically plugged into the embedded clause, the lack of reconstruction and island effects follow from the analysis.

Finally, some notes have to be made about the complementizer in OF. The basic pattern is the one observed in standard subordinate clauses: the complementizer is obligatory if the main clause contains a Focus or Contrastive topic apart from the associate demonstrative pronoun. All instances of OF are tied to such discourse functions, so *hogy* ('that(C)') is expected to be necessary.

(125) a *János(t) mondtad, *(hogy) jön.*
John(ACC) said.2SG that(C) comes
'Of John you said that he will come.'

A difference appears if interrogative subordinate clauses are considered. In standard subordinate sentences, this enables complementizer-drop, even in the cases of focussed main-clause elements. This is not the case in proleptic OF. (126) may only be interpreted as "you asked from John, whether he would come", and not as 'of John you asked whether he would come. In other words, the proleptic interpretation ("asked of John") is not available.

(126) **Jánost kérdezted, jön-e.*
John.ACC asked.2SG comes-Q-PRT

This may be explained if one considers the proleptic lexical entry of the relevant predicates (126). The grammatical function of the embedded clause is not an OBJ, it is a COMP. The most canonical realization of the COMP function is a CP, and the obligatory nature of the complementizer could be a reflex of this. Also, the categorial restriction may be a by-product of the argument-structural process creating the proleptic entry.

If the fronted element retains its original case, the original lexical entry is involved, so complementizer-drop should be an option. This is indeed the case, at least to a certain extent.

(127) *?János kérdezted, jön-e.*
John.ACC said.2SG comes-Q-PRT
'(Of) John you asked if he comes.'

I have no explanation why the degradation occurs. More work in the interaction of long-distance dependencies with complementizer-drop should be done in the future.

To conclude this section, let us summarize the analysis of Operator fronting. There are two distinct ways in which OF may happen. In the first case, some element of the subordinate clause appears in the preverbal field of the matrix clause and gets some information-structural specification. There is a strict syntactic dependency, functional identification between the fronted element and its embedded position. In the second case, via an idiosyncratic lexical process the basic entry of certain predicates is altered and an extra argument, bearing a "subject matter" theta role is added. This is called prolepsis. Hence the originally two-place predicate becomes a three-place one, subcategorizing for a subject, an object and complement clause. The extra argument is realized as an oblique or an object (depending on the original subcategorization of the base predicate) and is obligatorily co-indexed with an embedded grammatical function.

5.2.4 On split Operator fronting

Recently, in Jánosi (2013) a variation of OF received some attention. What she calls "long split focalization" involves some nominal that is modified by an adjective. It is possible to front the adjective-noun complex in a way that the adjective remains in the embedded clause, as illustrated by (128).

(128) *Autó* *mondtad,* *hogy* *új* *áll* *a* *garázsban.*
car said.DEF.2SG that(C) new stand.3SG the garage.in
'A car you said that a new one is standing in the garage.'

In (128) the fronted element is *autó* ('car') and there is the adjective *új* ('new') associated with it in the subordinate clause. The main verb shows definite conjugation, which indicates that it agrees with the subordinate clause in this respect (subordinate clauses count as definite objects). This construction parallels the "movement" version of OF, where the fronted element is functionally identified with its embedded canonical position and it is not thematically related to the main verb.

Split fronting can also behave like the proleptic version of OF. An example for this is shown in (129), where the fronted element bears accusative case, and the main verb is in indefinite conjugation, matching the fronted phrase.

(129) *Autót mondtál, hogy új áll a garázsban.*
car.ACC said.INDEF.2SG that(C) new stands the garage.in
'A car you said that a new one is standing in the garage.'

The split construction also parallels the unsplit version in other respects: as opposed to the functionally identified version, the proleptic pattern is insensitive to islands (130a) and can show number-mismatch (130b). Also, just like in unsplit OF, other grammatical functions can also be fronted (130c).

(130) a *Autót mondtál/*mondtad, hogy hallottad a hírt, hogy új áll a garázsban.*
car.ACC said.INDEF.2SG/ DEF that(C) heard.2SG the news.ACC that(C) new stands the garage.in
Intended: 'A car you said that you have heard the news that a new one is standing in the garage.'

b *Autót mondtál/*mondtad, hogy újak állnak a garázsban.*
car.ACC said.INDEF.2SG/ DEF that(C) new.PL stand the garage.in
Intended: 'A car you said that new ones is standing in the garage.'

c *Autót mondtál/mondtad, hogy újat vettél.*
car.ACC said.INDEF.2SG/ DEF that(C) new.ACC bought.2SG
'Of a car you said that you had bought a new one.'

What this suggests is that the underlying mechanisms of the split and unsplit versions of OF are the same. Split OF is possible as there is a general process by which adjectives and nouns may be separated in Hungarian (for details, see Jánosi 2014, chapter 2). When this happens, the adjective receives the case associated with the nominal (131b).

(131) a *Péter új autót vett.*
Peter new car.ACC bought.3SG
'Peter bought a new car.'

b *Autót vett Péter újat.*
car.ACC bought.3SG Peter new.ACC
'Peter bought a new car.'

Given the independent motivation for the possibility of this noun phrase split in (131b) and the analytical parallel of split and unsplit OF, split OF does not constitute an independent issue for the purposes of this dissertation. What seems worth noting is that in the proleptic version, the role of the resumptive element (the co-indexed element in the COMP) is fulfilled by the case-marked adjective.

I think if future research explores the LFG-analysis of split adjectival constructions in general, the details of the analysis should follow from the general approach adopted in this chapter.

5.2.5 On Long topicalization

There is a fronting construction in Hungarian that has been mentioned in the literature and it resembles OF in some respects but has not been discussed so far. É. Kiss (2002) calls it "Long topicalization" (LT) and I adopt this terminology. LT looks like OF in that there is some fronted element which is related to some embedded grammatical function. However, there are several crucial differences. First of all, the discourse function of the fronted element is a neutral Topic. In (132) there is no sign of contrastivity or other discourse-prominence associated with *János* ('John').

(132) *János azt gondolom, hogy jön.*
John that.ACC think.1SG that(c) comes.
'John, I believe that he comes.'

Also observe in (132) that the demonstrative pronoun is present, in contrast with OF structures. This means that the CP is an adjunct, from which extraction should be impossible. Other seemingly island-constraint violating examples are also shown, from É. Kiss (2002:258). (133a) contains a temporal adjunct and (133b) a complex noun phrase.

(133) a *János már dél felé járt az idő, [amikor felébredt].*
John already noon towards went.3SG the time when awoke.3SG
'John, it was already about noon when (he) woke up.'

b *Jánost nincs [az az ember aki fel tudná bosszantani].*
John.ACC isn't that the man who up could make.angry.INF
'John, there is no man who could make (him) angry.'

Moreover, the complementizer is completely optional in such sentences.

(134) *János azt gondolom, jön.*
John that.ACC think.1SG comes
'I think that John comes.'

From this it is clear that LT is a distinct structure. I agree with É. Kiss (2002:259) in that such elements are left-peripheral hanging topics. From the perspective of this dissertation, the construction is akin to LD-Eng or CFLD-Hun: a "syntactic orphan", unintegrated into the host sentence. Thus it is expected that the extra-sententiality effects associated with these structures are to be observed with LT as well. They are indeed: non string-initial LTs are dispreferred in in (135) in the same way as LD-Eng a CFLD-Hun are. These effects are not observable with OF (136).

(135) a *??Szerintem János azt mondtad, hogy jön.*
in.my.opinion John that.ACC said.2SG that(C) comes

b *??Úgy hallottam, hogy János azt gondolod, hogy jön.*
so.DIST heard.1SG that(C) John that.ACC think.2SG that(C) comes

c **Te vagy az az ember, aki János azt gondolja, hogy jön?*
you are that the person who John that.ACC think.3SG that(C) comes
Intended: 'Are you the person who (of) John$_i$ thinks that he$_i$ comes?'

(136) a *Szerintem János(-t) mondtad, hogy jön.*
in.my.opinion John-(ACC) said.2SG that(C) comes
'In my opinion (of) John you said that he comes.'

b *Úgy hallottam, hogy János(-t) gondolod, hogy jön.*
so heard.1SG that(C) John-(ACC) think.2SG that(C) comes
'I've heard that (of John) you think that he comes.'

c *Te vagy az az ember, aki János-(t) gondolja, hogy jön?*
you are that the person who John(-ACC) think.3SG that(C) comes
'Are you the person who (of) John thinks that he comes?'

Thus LT is to be treated as an independent syntactic structure, set apart from OF. It could be viewed as a topic-marking counterpart of CFLD-Hun. Actually, such a state of affairs is not unexpected, as nothing prescribes that left-peripheral syntactic orphans in Hungarian are to be associated with a Focus discourse-function.

5.3 Conclusion of chapter 5

In this chapter I investigated Hungarian left-peripheral constructions. The two main areas of investigation were Left-dislocation (LD-Hun) and Operator fronting (OF). Both of them turns out to comprise of several distinct configurations.

In LD-Hun, some element in the topic field of the Hungarian sentence (a neutral or a Contrastive topic) is associated with a pronominal. There is anaphoric dependency between them and some syntactic feature-matching is also required.

In Operator fronting, some element that is associated with an embedded grammatical function appears in the matrix clause and bears some prominent discourse function (Contrastive topic, Contrastive/Information focus). This may happen in two ways: either along the lines of conventional "fronting" or as prolepsis. In the first case, the fronted element is related to its embedded position via functional identification. This means strict syntactic matching, including case, person and number features. In prolepsis, the fronted element becomes the argument of the main predicate and an anaphoric link is established with the embedded function. The dependency is weaker here, allowing for case and number mismatches and the circumvention of island-constraints.

Apart from these, I also discussed two configurations where a fronted element is "radically" left-peripheral. In Contrastive focus left dislocation and Long topicalization, the sentence-initial constituent is a "syntactic orphan", an element not integrated into the host sentence.

CHAPTER 6

CROSSLINGUISTIC AND THEORETICAL PERSPECTIVES

Throughout the dissertation we have discussed various left-peripheral ("fronted") discourse-related constructions in English and Hungarian. The aim of this chapter is to put them into perspective, by looking at similar structures in other languages, making comparisons and also to investigate how the proposed analyses fit into the existing theoretical space of fronting constructions.

The configurations discussed in the dissertation may be put into three distinct categories:

i. **"Fronting proper"**: a left-peripheral element is "extracted" from a sentence-internal position. Although the c-structural position of the fronted element is left-peripheral, functionally it is plugged into its canonical position. This manifests in strict agreement-properties and syntactic reconstruction effects. English Topicalization (TOP-Eng) and the case-retaining version of Hungarian Operator fronting belong here.
ii. **Left-dislocation**: some discourse-prominent element of the left-periphery is associated with a pronoun. The constructions here are LD-Eng and LD-Hun.
iii. **Prolepsis**: some internal argument of the main clause predicate is anaphorically identified with an embedded grammatical function. This is the analysis proposed for the case-switching version of Operator fronting.

Additionally, we discussed Clause-initial adjuncts in English. These may be viewed either as Fronting proper (in the case of event-internal adjuncts) or as Left-dislocation without a pronoun associate. These are not particularly remarkable from the perspective of this dissertations apart from the remarks already made, so no further discussion is going to be provided here about them. The interested reader is referred to Ernst (2002) and Lang, Maienborn & Fabricus-Hansen (2003).

Of the other three, proper fronting and left-dislocation constructions will be discussed together in the next section. Prolepsis is the construction that has been given the least attention in the literature, so here I will give it the most exposition.

6.1 Perspectives on Fronting proper and Left-dislocation

Let us begin with Fronting proper. Both TOP-Eng and case retaining OF involves the extraction of some argument from a sentence-internal position. In both structures, the fronted element maintains strong ties with its original position, which results in connectivity-effects: the case, the number and the person features of the fronted element must match those that it would receive in its original position.

The main syntactic difference between TOP-Eng and OF is that the former is licensed in a specialized left-peripheral IP-adjoined position, while OF is part of the standard Hungarian sentence-structure, as an element in the regular topic- or focus-field of the Hungarian sentences.

This difference translates into a difference in information-structure: while TOP-Eng is linked to contrastive i-structural categories (Contrastive topic and Contrastive focus), OF may also receive an Information focus interpretation. The contrastiveness of a fronted element in Hungarian is only preferred in the case of fronted subjects. This, however, is probably not related to the structural properties of the construction, but has to do with the subject grammatical function itself. SUBJ is the most unmarked grammatical function, both in terms of argument-structure specification (in LMT it is –restricted, –objective) and case properties (almost always nominative/absolutive cross-linguistically). Also, note that the extraction of subjects is often restricted cross-linguistically (e.g. the "*that*-trace filter"). What I suggest is that since SUBJ is the most problematic GF for the establishment of long-distance dependencies, one needs the most motivation for creating one, even in languages where the link is otherwise licensed. Contrastiveness could provide that motivation which translates to CF and CT being the most preferred IS-categories for extracted subjects.

The TOP-Eng structure is not without cross-linguistic correlates. Obligatory contrastiveness has been noted with regards to a couple of constructions in other languages as well. For example, Vilkuna (1995) suggests that Finnish has a left-peripheral position which may be occupied either by a Contrastive focus or a Contrastive topic.[77] The following examples are from Molnár & Winkler (2010:1399).

(1) a A: *Pekka* *lensi* *Tukholmaan.*
Pekka flew.3SG to.Stockholm
'Pekka flew to Stockholm.'

[77] A similar construction in Dutch was also mentioned in section 2.6.1.3.

B: *Eihän, vaan [Reykjavikiin]$_{CF}$ Pekka lensi.*
no, but to.Reykjavik Pekka flew
'That's not true, to Reykjavik Pekka flew.'

b *[Tukholmaan]$_{CT}$ Pekka lensi Finnairilla. [Reykjavikiin]$_{CT}$*
to Stockholm Pekka flew.3SG by.Finnair to Reykjavik
Icelandairilla.
by.Icelandair
'To Stockholm, Pekka flew by Finnair. To Reykjavik, by Icelandair.'

As Contrastive topics need not be referential, even a verb may occupy this position, as in (2), from Vilkuna (1995:252).

(2) a *Mikko on hyvä ruoanlaitossa mutta [leipoa]$_{CT}$ hän ei osaa.*
Mikko is good cooking.in but bake.INF he not.3SG can
'Mikko is good at cooking, but bake, he can't.'

German offers another example. Frey (2010) calls attention to the fact that certain sentence-initial elements in German must receive heavy stress and should be interpreted contrastively. For instance, see the pair in (3).

(3) a *Einen Kollegen hat heute ein Polizist verhört.*
a.ACC colleague has today a policeman interrogated.3SG
'A policeman interrogated a colleague today.'

b *GRÜN will Maria die Tür streichen.*
green wants Mary the.ACC door paint.INF
'Mary wants to paint the door GREEN (and not some other color)', 'GREEN Mary wants to paint the door.'

According to Frey (2010:1418), it is only (3a) that is "acceptable even if the prefield elements are not stressed beyond the word accents". That is, while *einen Kollegen* is acceptable as a neutral topic or Information focus, *grün* ('green') must be contrastive. Although Frey (2010) is not explicit about this, I suspect that both CT and CF are acceptable interpretations.

Frey's (2010) explanation is that the fronting of the two constructions is executed by distinct mechanisms. In (3a) we see "formal movement": "the highest maximal phrase in the middle field is moved to the adjacent prefield (…) without any intonational or interpretative effects of its own". (3b) on the other hand involves A-bar movement, which provides an additional interpretative effect, namely contrastivity. Formal movement is clause-bound, which means that any element fronted from an embedded sentence may only undergo the second mechanism, as in (4).

(4) *Den CHEF/*chef meint Maria, dass Paul zur Party einladen sollte.*
the.ACC boss thinks Mary that(C) Paul to.the party invite.INF should.3SG
'The boss Mary thinks that Paul should invite to the party.'

So what Frey (2010) analyzes as A-bar movement is a close correlate of TOP-Eng, while "formal movement" is closer to OF. Note that neither A-bar movement nor TOP-Eng is obligatory in any context: contrast is dependent on the discourse, including the intention of the speaker. Frey (2010) demonstrates this with (5), where B1, with *einen Kuchen* ('a cookie') is a CF is preferred over B2. The latter with *einen Kuchen* as IF is nevertheless a possible answer.

(5) A: *What does Paul want? Ice-cream or cookie?*
B1: *Einen Kuchen möchte Paul.*
a.ACC cookie wants Paul
'An ice-cream Paul wants.'
B2: *Paul möchte einen Kuchen.*
Paul wants a.ACC cookie
'Paul wants a cookie.'

This is parallel to the observation made by Krifka's (2008) (mentioned in section 2.6.1) that the answer in (6) is not necessarily contrastive. That is, (6/B1) may optionally be expanded into a TOP-Eng structure if B wants to evoke contrast.

(6) A: *What do you want to drink, tea or coffee?*
B1: *Tea*.
B2: *Tea I want (and not coffee).*
B3: *I want tea.*

Also, while contrast is obligatory in Frey's (2010 A-bar movement construction and in TOP-Eng, it is not exclusively related to these configurations, as in-situ contrast is also possible.

(7) a *Paul möchte EINEN KUCHEN.*
Paul wants a.ACC cookie
'Paul wants a cookie.'
b *I want TEA, and not a coffee.*

That is, while certain configurational positions are necessarily contrastive, a contrastive element does not have to move into these slots. These considerations argue against the cartographic

view of contrast (e.g. Molnár & Winkler 2010), where "contrast" is a syntactic feature, triggering movement for checking in the specifier a designated position. I agree with Neeleman & van de Koot (2008) in that such fronting constructions should be viewed as interface optimizations, whereby the syntactic structure of a sentence is matched to an information-structural template (the designated contrastive positions).

Also, from these it seems that while certain left-peripheral constructions are contrastive, it cannot be claimed that all of them are: CIADJ-Engs, OF and "formally moved" German elements are not. So while contrast may be a prominent motivation for dislocating elements, it cannot be responsible for all frontings.

German also offers good grounds for the comparison of LD-constructions (where the left-peripheral element has a pronoun associate). Following Grohmann (2003), two main types are distinguished. The first one is a "hanging topic left-dislocation" (HTLD-Ger, 8a), which involves a loosely integrated left-peripheral element paired with a sentence internal demonstrative/personal pronoun, with a topical interpretation. The second is labelled Weak pronoun left-dislocation (WPLD-Ger, 8b) and invariably involves a demonstrative pronoun, just like CTLD-Hun and NCLD-Hun. I will use personal pronouns in the HTLD-examples to expose the difference.

(8) a *Den/ Der Hans, jeder mag ihn.*
the.ACC the.NOM Hans everyone likes him
'Hans, everyone likes him.'

b *Den Hans, den mag jeder.*
the.ACC Hans demonstrative likes everyone
'As for Hans, everyone likes him.'

HTLD-Ger may be viewed as an extrasentential syntactic orphan (Shaer & Frey 2004). As such it shares several properties with LD-Eng and CFLD-Hun. Prosodically, the fronted element is separated from the host sentence by a noticeable intonational break. Syntactically, as seen from (8a), there is no obligatory case-matching between the fronted element and its sentence-internal correlate. Like in LD-Eng and CFLD-Hun, reconstruction is weak, see (9) from Shaer & Frey (2004) (though as noted in connection with example 37 in chapter 4, there could be some not yet understood semantic reconstruction processes at play).

(9) a *?Sein$_i$ erster Artikel, ich glaube, dass [jeder Linguist]$_i$ ih als*
his first article I believe.1SG that(C) every linguist it as
Mißerfolg betrachten würde.
failure consider would.3SG
'His first article, I think every linguist would consider it a failure.'

b *Freunde von einander, Herforder erzählen ihnen selten Lügen.*
friends of each.other Herfordians tell.3PL them rarely lies
?'Friends of each other, Herfordians rarely tell them lies.'

Furthermore, HTLD-Ger seems to be marked if placed sentence-internally. (10) is from Grohmann (2003), where *einen Arschtritt* ('a kick in the ass') is a topicalized entity and *dieser Kandidat* (this candidate') is the constituent in HTLD-Ger.

(10) **Einen Arschtritt dieser Kandidat, sollte man ihm geben.*
a.ACC kick-in-the-ass this.NOM candidate should one him give
Intended: 'A kick in the ass, this candidate, one should give him.'

WPLD-Ger displays the opposite behavior regarding these phenomena. There are reconstruction effects and sentence-internal positions are grammatical.

(11) *Seine$_i$ Mutter, die verehrt [jeder Junge]$_i$.*
his mother DEMONSTRATIVE admires every boy
'His mother, every boy admires.'

(12) *Ich meine dass Hans, den jeder mag.*
I think.1SG that(C) Hans DEMONSTRATIVE everyone likes
'I think that Hans, everyone likes.'

Thus WPLP-Ger shows properties of TOP-Eng, CTLD-Hun and NCLD-Hun. All of these are syntactically integrated structures. Like the Hungarian structures, WPLD-Ger involves a demonstrative pronoun which featurally matches the fronted element. A difference is that WPLD-Ger and TOP-Eng always reconstruct while NCLD-Hun does not. This is probably a difference stemming from the reconstruction-properties of the respective languages, and not from the fronting constructions themselves.

An important i-structural addition is that WPLD-Ger is sometimes assumed to be always contrastive (e.g. Grohmann 2003) but actually it is not, as argued by Frey (2005). Consequently, LD-Ger resembles LD-Hun in the bifurcation into a syntactically non-integrated (HTLD-Ger, CFLD-Hun) and an integrated type (WPLD-Ger, CTLD/NCLD-Hun), where the integrated type may be, but not necessarily is contrastive.

6.2 Proleptic structures cross-linguistically

In my discussion of Operator fronting in Hungarian, I have argued for an analysis where the case-switched element in the main clause is a thematic argument of the main verb and it is co-indexed with an argument of the embedded clause. In this part of the dissertation I explore how the purported analysis fits into the cross-linguistic picture both in terms of descriptive data and from the perspective of theoretical syntax in general. I will show that similar accounts have been given for phenomena in various languages (English, German, Madurese, Greek, Korean, Japanese). Also, I will argue that prolepsis can be regarded as a kind of control, though this requires widening the concept of control to a certain extent. However, the parallels are clearly visible and we will see that from an LFG perspective it is actually not unexpected that such structures may exist. After locating prolepsis in the general typology of control(-like) relations I will also survey other occupants of the typology.

As I have already noted, this configuration is commonly known in the literature as "prolepsis". The term itself originates in rhetoric. There it is used to describe a figure of speech, where the speaker anticipates and answers possible objections to their argument. The element of anticipation or foreshadowing is carried over to linguistics, where an argument also related to an embedded verb already manifests in the matrix clause.

As a working defintion, Salzmann (to appear:1) characterizes prolepsis as "a construction where a structural complement of the matrix verb is semantically related to the predicate of a finite embedded clause".

It was first mentioned by Higgins (1981) for the characterization of sentences like (13), from Zacapoaxtla Nahuat (a Mexican language).

(13) *Nimickaki ke tiwa:la.*
hear.1SG.2SG that came.2SG.
'I hear that you came.'

Although it is not evident from the English translation, the object of the main verb is an incorporated 2nd person singular object ("you") and it is co-referent with the subject of the subordinate clause.

Some discussion of the structure and its possible English parallels takes place in Massam (1985). Then for a long time prolepsis went out of the theoretical limelight. The notion was resurrected by the work of Davies (e.g. 2001, 2005) and Salzmann (2006, to appear). In what follows, I survey the literature on proleptic constructions in different languages.

6.2.1 English

Reflecting on Higgins (1981) Massam (1985) discusses two types of sentences in English in connection which a proleptic analysis might be a possibility. All the examples in this section come from these works.

The first is a pattern with perception verbs:

(14) a *Catherine saw the nuns doing yoga.*
b *Barbara saw Geoffrey and Anvil singing.*

According to our characterization, (14) is not a good example for prolepsis even on the first glance. This is because the complement clause is non-finite. Disregarding this, the reason why Higgins (1981) and Massam (1985) entertain the possibility of a proleptic analysis is that *the nuns* and *Geoffrey and Anvil* seem to fulfil two roles in the sentences: they are the objects of perception and the doers of the action.

However, even in this sense the proleptic account cannot be maintained. This is because it can be shown that there is no thematic relationship between the main verbs and the objects. Although the first intuition about (14) may be that the objects act as targets of perception (that is, it is entailed that *Catherine* saw *the nuns* and *Barbara* saw *Geoffrey* and *Anvil*), this may be an illusion. Massam (1985) gives the following examples to dispel the misconception. In none of these do we have the reading that the object by itself is a target for perception. Rather, the entire event is heard/smelled/seen.

(15) a *I heard George start up a chainsaw.*
b *I could smell my neighbors cooking dinner.*
c *I saw him crush the Huey.*

Based on this, Massam (1985) argues for an analysis involving Exceptional Case Marking (ECM), where the matrix verb only subcategorizes for a clause and the expressions *George/my neighbors/him* are in the embedded small clause, receiving "exceptional case marking", similarly to the nonthematic objects of *believe, consider,* etc. (e.g. *I believe him to be happy*). Massam (1985:167) notes that the illusion of direct perception arises because "when the event determined by the clause is seen, it is normally the case (though not always and not necessarily) that the participants in the event are also seen". The nonthematic relationship between the main

verb and the object is also reinforced by the availability of idioms in the construction. This is impossible in genuine cases of prolepsis, like Hungarian, as discussed in the previous chapter.

(16) a *One could hear the shit hitting the fan in the next room.*
b *We used to see the fur fly when the phone bill came due.*

On the basis of these considerations, we can discard a proleptic analysis for perception verbs in English. The other construction that Higgins (1981) and Massam (1985) discuss is when a propositional verb takes a bare or a prepositional object and a clausal complement containing an element co-referent with the main clause object. The bare object version is not grammatical in contemporary English. In (17) we see some examples from "Biblical English" (in 17b-c, the object has undergone passivization).

(17) a *And God saw the light, that it was good.*
b *Then shall the prophet be known that the Lord hath truly sent him.*
c *Now if Christ be preached that he rose from the dead.*

Contemporary English uses the preposition *of* in the relevant structures.

(18) a *I know of Mrs. Dalloway that the light gave her headache.*
b *I read of Carrol that she was awfully shy.*
c *I've heard of him that he didn't realize he was oppressed.*

Apart from the preposition, this construction closely parallels the state of affairs discussed with regards to Hungarian Operator fronting. The (prepositional) objects bear the same kind of "aboutness"-relationship with the main verb that I have characterized as a subject matter theta role. (18a) also shows that the co-referent pronoun in the embedded clause may bear grammatical functions other than subject.

Although Higgins (1981) and Massam (1985) do not mention this fact, but idioms are excluded from the construction. (19) only has a literal interpretation.

(19) *#I believe of the shit that it hit the fan.*

According to Massam (1985), the thematic status of these *of*-complements is unclear. While (19) would suggest a thematic relationship, the optionality of the *of*-PP makes an argumental

analysis unlikely. Massam (1985) leaves the question open. In my analysis of Hungarian OF, I argued that the verbs licensing OF generally allow an oblique complement, which can be considered a thematic adjunct. The oblique (delative) complement (*-ról*/*-ről*) is quite similar in meaning to these *of*-PPs. Therefore, it seems plausible to argue for a thematic adjunct classification of these constituents, reconciling their ambivalent properties. While the lexical process that was argued to turn these thematic adjuncts into real arguments seems to be not operational in current English, the biblical examples indicate that it was part of the language. It should be noted at this point that a remnant of this lexical process might be that certain British English dialects allow such sentences as (20) (É. Kiss 2002:255, citing Chomsky 1981).

(20) *Whom do you suggest should be the president?*

The accusative case of the question word must have been assigned by the matrix verb. According to É. Kiss (2002:255), "Chomsky (1981:174) suggests that in such constructions the raised constituent receives accusative case from the intermediate verb when passing through the specifier of the lowest CP. It is only a nominative-marked constituent, bearing a phonologically null case marker, whose case can be superseded by an accusative assigned by the matrix verb". While this is a possible explanation (though one has to assume multiple case-assignment for it to work), an alternative could be an analysis where *whom* is base-generated in the matrix-clause as the argument of *suggest*. This would pretty much be like a proleptic analysis. Further data could reveal the behavior of idioms, the possibility of resumptive pronouns or the effect of island-constraints on this construction. As such data is not discussed by Chomsky (1981) or É. Kiss (2002) I must leave this question open.

6.2.2 German

Prolepsis in German is discussed in detail in Salzmann (2006, to appear). The examples in this section are from these works. In German, the proleptic construction involves the preposition *von*. There is no equivalent in German for the version where the proleptic object is the direct object or an oblique argument of the matrix verb. (Salzmann notes that prolepsis is much more common with combination with relative clauses (21a) but it is also possible to use it with *wh*-clauses (21b) and topicalizations (21c).

(21) a *ein Maler, von dem ich glaube, dass Maria ihn mag*
a painter of who.DAT I think.1SG that(C) Mary him likes
'a painter who I think that Mary likes'

b *Von welchem Maler glaubst du, dass Maria ihn mag?*
of which.DAT painter think.2SG you that(C) Mary him likes
'Which painter do you think that Mary likes?'

c *Von dem Maler glaube ich, dass Maria ihn mag.*
of the.DAT painter think.1SG I that(C) Mary him likes
'The painter, I think that Mary likes.'

Unlike Hungarian OF, which is lexically restricted to a considerable extent, the German prolepsis seems to work with basically every predicate that takes a complement clause. (22a) illustrates it with the German equivalent of *whisper*, a verb that would not work in Hungarian (22b).

(22) a *beim al-Qaida-Terrornetzwerk, von dem gleichzeitig sämtliche*
at.the al-Qaida-terror.network of which simultaneously all
Experten augenzwinkernd flüstern dass es überhaupt nicht
experts with.a.wink whisper.3PL that(C) it at.all not
mehr existiert
anymore exists
'with the Al-Qaida terror network that all experts simultaneously whisper with a wink that it does not exist anymore'

b **Jánost suttogtam, hogy jön a partira.*
John.ACC whispered.1SG that(C) comes the party.onto

Because of this free distribution, Salzmann (2006, to appear) discards an analysis where the proleptic constituent is an argument of the main verb. Just like in English, the optionality of the PP also argues against such an analysis. Based on various data involving binding, superiority and restrictions on extraction Salzmann posits that the proleptic constituent is base-generated in a main clause argument position (despite it not being an argument). For details see Salzmann (2006:193-208), here I present an example that shows that the construction is not sensitive to island-constraints. This directly parallels the state of affairs observed at OF. (23a) shows a CNPC-island, while (23b) is a *wh*-island.

(23) a *der Mann, von dem ich denke, dass Marie jedes Buch liest,*
the man of who.DAT I think.1SG that(C) Mary every book reads
das er schreibt
which he writes
'the man of whom I think Mary reads every book that he writes'

b *der Mann, von dem ich glaube, dass niemand weiß, wie er heißt*
the man of who.DAT I believe.1SG that(C) nobody knows how he is.called
'the man of whom I think nobody knows what he is called'

Salzmann (2006, to appear) also notes that an analysis where the proleptic element is a simple adjunct is also undesirable, as the *von*-constituent is licensed only if there is a co-referent element in the subordinate clause. As (24) shows a clause that is simply "about" the proleptic PP is not enough to license it.

(24) **Von Computern glaube ich, dass jeder einen PC kaufen sollte.*
of computers.DAT believe.1SG I that(C) everyone a.ACC PC buy.INF should.3SG
'I believe of computers that everyone should buy a PC.'

In the two languages surveyed so far, the proleptic entity is introduced by a prepositional phrase as a thematic adjunct, as opposed to the Hungarian situation, where it could also be a direct object or oblique argument. In the following sections, I show linguistic data that bears a closer resemblance to Hungarian OF in this respect.

6.2.3 Madurese

Madurese is a language related to Indonesian, spoken in Southeast-Asia. It is a close relative to Balinese, Javanese. Davies (2005) provides a detailed description of its proleptic construction. (25) is an example for Madurese prolepsis. This and other Madurese examples in this section are taken from Davies (2005).

(25) *Siti ngera Hasan bari' melle motor.*
Siti think Hasan yesterday buy car
'Yesterday Siti thought about Hasan$_i$ that he$_i$ bought a car.'

According to Davies (2005) there is no case- or tense-marking in Madurese. But the positioning of *bari'* ('yesterday') is a fair indication that *Hasan* is in the main clause.[78] The question is how

[78] Some details about Madurese morphosyntax are to be found in Davies (2005:646-649). These details are set aside here.

it gets there. Davies (2005) considers two options: the movement of *Hasan* from the embedded clause (a raising/movement analysis) or base-generation there (the proleptic analysis). Based on various data, to be briefly discussed below, he opts for the second possibility. These properties are essentially like the tones that also characterize the Hungarian OF, or the English and German cases discussed in the previous sections.

First, it is possible to pronounce a resumptive pronoun in the embedded clause. Although just like Hungarian, Madurese is a pro-drop language, so pronouns are generally dropped if independent factors (e.g. information structure) do not require their presence, (26) is possible.

(26) *?Siti ngera Hasan bari' aba'eng melle motor.*
Siti think Hasan yesterday he buy car
'Yesterday Siti thought about Hasan$_i$ that he$_i$ bought a car.'

Second, Davies (2005) observes, as the argument structure of the predicate and the projected syntax is altered, one can detect a shift in meaning compared to a version of the sentence where no prolepsis takes place. While (27a) describes a situation, (27b) puts the emphasis on an individual.

(27) a *Ita a-bukteagi ja' Hasan ngeco' sapedha.motor*
Ita prove that(C) Hasan stole motorcycle
'Ita proved that Hasan stole the motorcycle.'

b *Ita a-bukteagi Hasan ja ngeco' sapedha.motor*
Ita prove Hasan that(C) stole motorcycle
'Ita proved about Hasan that he stole a motorcycle.'

Third, it is possible for the proleptic object to refer to nonsubjects in the embedded clause. (28a) illustrates it with an embedded object, while (28b) shows an embedded possessor.

(28) a *Siti ngera Hasan ja dokter juwa mareksa aba'eng.*
Siti think Hasan that(C) doctor DEMONSTRATIVE examine he
'Siti thinks about Hasan that the doctor examined him.'

b *Marlena a-bala-agi Hasan ja embi'-eng ngekke Ali.*
Marlena say Hasan that(C) goat bite Ali
'Marlena said about Hasan$_i$ that his$_i$ goat bit Ali.'

Fourth, nonbridge verbs also participate in Madurese prolepsis, e.g. *kabarragi* (‘spread the news’), which is unexpected under a proper fronting approach.

(29) *Terdokter juwa ekabarragi Ina mon aba'engi ngobad ana'eng Marlena*
doctors DEM spread.the.news Ina if they cure child Marlena
‘Ina spread the news about the doctors that they cured Marlena’s child.’

In fact, Davies (2005) claims that the construction is possible with any verb capable of taking a clausal argument, which is like the situation in German.

Fifth, idiom chunks are not interpretable in Madurese prolepsis.

(30) a *Nase’la daddi tajjin.*
rice.already become porridge
‘It is too late to do anything about it.’ (lit.: ‘The rice has become porridge.’)

b *Siti ngera nase bari ja la daddi tajjin.*
Siti think rice yesterday that(C) already become porridge
Only lit.: ‘Siti thought about the rice that it became porridge.’

Sixth, island-constraints do not affect Madurese prolepsis, as the following sentence with CNPC shows.

(31) *Wati ng-enga’-e Atin careta-na ja aba’eng ng-angkep maleng.*
Wati remember Atin story that(C) she capture thief
‘Wati remembered about Atin the story that she captured the thief.’

Finally, Davies (2005) notes that alternatively, the proleptic construction may involve a PP. This also likens Madurese to the Hungarian, German and English structures.

(32) *Siti ngera parkara Hasan ja epareksa dokter juwa.*
Siti think about Hasan that(C) examine doctor DEMONSTRATIVE
‘Siti thinks about Hasan that that doctor examined him.’

The data shows that there is a massive parallelism between the properties of Hungarian OF and Madurese prolepsis as described by Davies. What seems to be a difference is the productivity. While in Hungarian only a subset of the verbs that occur with a thematic adjunct PP can participate in the direct object version of prolepsis, in Madurese, according to Davies’s (2005) assessment, all of them.

6.2.4 Greek

In Greek, the construction that most closely resembles prolepsis is referred to "quasi-ECM" by Kotzoglou & Papangeli (2007). In (33) it can be seen that the matrix verb takes a direct object and this direct object is co-referent with the (null) subject of the embedded clause. In (33a) the embedded clause is subjunctive, but in Ingria (1981) and Philippaki-Warburton (1987) we find examples with indicative clauses as well, (33b) is an example from Ingria (1981).

(33) a *I epivates perimenan ton kapetanio na ferthi me*
the passengers expected.3PL the captain.ACC SBJV behave.3SG with
aksioprepia
dignity
'The passengers expected of the captain that he should behave with dignity.'

b *Theoro ton jani pos ine eksipnos*
consider the John.ACC that he smart
'I consider John to be smart. / I believe of John that he is smart.'

Kotzoglou & Papangeli (2007) reports that the construction occurs with a couple of matrix verbs, including, apart from the ones in (33), *ipologhizo* ('estimate'), *pistevo* ('believe'), *ksero* ('know'), *thelo* ('want'). The words in this set express some mental state or propositional belief, and as such, they are comparable to the Hungarian data.

Although in (33a), the subject of the embedded clause is null, Kotzoglou and Papangeli (2007) show with various means that the embedded clause does contain an independent subject, just like (33b). From the evidence here only one piece is reproduced. Emphatic modifiers in Greek attach to and agree in case with a host. It is possible to put such an emphatic modifier in the embedded clause in the "quasi-ECM" construction, which indicates that there must be some host for it. Moreover, it bears nominative case, which suggests that the host is not the distant main clause direct object.

(34) *I epitheorites ithelan to jani na lisi monos tu/*
the inspectors wanted.3PL the John SBJV solve.3SG alone.NOM
**mono* tu to provlima
alone.ACC his the problem
'The inspectors wanted of John that he should solve the problem on his own.'

Like in proleptic constructions in the previously discussed languages, the Greek "quasi-ECM" is not compatible with idiom chunks.

(35) **Perimena psilus na tu bun st' aftia.*
expected.1SG fleas.ACC SBJV his get.3PL in.the ears
Only lit.: 'I expect of the fleas that they get into his ears.' (*idiomatic: I expect him to become suspicious.)

Another similarity with the previously discussed languages is the possibility of a paraphrase with PP proleptic element.

(36) *I epivates perimenan apo ton kapetanio na ferthi me*
the passengers expected.3PL from the captain SBJV behave.3SG with
aksioprepia.
dignity
'The passengers expected from the captain that he should behave with dignity.'

Although Kotzoglou and Papangeli (2007) seem to argue for an argumental analysis for both the PP and the direct object version of the structure, given the analyses discussed for the proleptic constructions so far it seems more likely that the PP should be assigned a thematic adjunct status, while the argumental analysis could be maintained for the direct object version.

All these properties liken the "quasi-ECM" constructions to the proleptic constructions I have been discussing. There are a number of differences, however. The first is that the construction is not unbounded, that is, the proleptic element must be co-indexed with an argument in the immediately embedded clause. Also, this embedded argument must be the subject of the embedded clause (Salzmann to appear:25).

These properties make the Greek construction bear closer resemblance to traditional object-control structures, like the English example in (37).

(37) *I persuaded John to leave.*

Indeed, this is the kind of analysis that is suggested by Kotzoglou & Papangeli (2007). They argue that the "quasi-ECM" construction is a case of "semantic control", the proleptic element gets a "weak theta role" from the main verb and then it gets co-indexed with the embedded subject. This assessment is actually not far from what I argue to be the case.

6.2.5 Korean and Japanese

Korean and Japanese possess the following constructions that could be regarded as cases of prolepsis. (38) is a Korean example from Yoon (2007), while (39) is a Japanese one, from Hoji (2005).

(38) *Cheli-nun Yenghi-lul yenglihaysstako mitnunta.*
Cheli-TOP Yenghi-ACC smart believed
'Cheli believed Yenghi to be smart.'

(39) *John-wa Mary-o Itariazin da to omotteita.*
John-TOP Mary-ACC Italian be that thought
'John believed Mary to be Italian.'

There has been a debate about the proper analysis of such sentences. Three analyses have been put forward. In the first one (proposed e.g. by Hong S. M. 2005 for Korean and Tanaka 2002 for Japanese) (38)-(39) is analyzed as "raising to object", where the accusative marked element starts out as the subject of the embedded clause and moves to the main clause. This is what the standard analysis for the English glosses of the sentences is. Another kind of raising analysis is offered by Yoon (2007), who proposed that what moves is the "major subject" (a topic constituent) of the embedded clause. So in this analysis the accusative-marked elements do not start out as the objects of the embedded clauses, but they are base-generated at the left periphery of the embedded clause and bind a zero pronominal variable in the object position. The third analysis is basically a prolepsis-analysis. In that, proposed for example by Hong K. S. (1997) for Korean and Hoji (2005) and Miura (2008) for Japanese, the accusative-marked constituent is base-generated as the direct object of the main predicate.

Yoon (2007) explains in detail that the first, object-raising analysis can be readily ruled out. To start with, the embedded clauses are finite and the construction can target nonsubject functions of the embedded clause. Such facts, while not entirely fatal, are unexpected under the standard raising analysis. (40a) is a Korean example involving a possessor, (40b) is from Japanese.

(40) a *Na-nun Cheli-lul meli-ka cohtako mitnunta.*
I-TOP Cheli-ACC head good believe
'I believe of Cheli that his head is good.'/ 'I believe Cheli to be smart.'

b *Taroo-wa Hanako-o atama-ga ii to shinjiteiru.*
Taroo-TOP Hanako-ACC head-NOM good COMP believes
'Taroo believes of Hanako that his head is good.'/ 'Taroo believes Hanako to be smart.'

Moreover, like the structures in the previous section, it is not affected by island-constraints. (41) is a Korean example.

(41) *Na-nun Yenghi-lul kunye-ka hanun ili mopemcekilako sayngkakhanta.*
I-TOP Yenghi-ACC she-NOM do work examplary think
'I think of Yenghi that the things she does are exemplary.'

Finally, a number of interpretative properties militate against the standard raising analysis. For example the construction fails on the familiar idiom-chunk test.

(42) *Hankwuksalam-un cakun kochwu-lul mayptako sayngkakhanta.*
Korean-TOP small pepper-ACC hot think
Only lit.: 'Koreans believe small pepper to be hot.' (*Idiomatic: 'Koreans believe that size does not matter.')

The distinction between the "major subject"-analysis and the prolepsis analysis is actually quite delicate. Since the "major subject" is a kind of topic that sits at an edge-position of the lower clause, the island and idiom facts follow. In (41) what is inside the island is the zero pronominal co-indexed with the major subject). Also, the major subject being a topic, incompatibility with idioms is expected.

What makes Yoon (2007) side with the "major subject"-analysis is that there are certain properties of the accusative element that suggest that it originates in the lower clause. For example, sometimes an oblique case assigned by the embedded predicate is retained and the accusative is stacked on the top of it.

(43) a *Cheli-hanthey-(man)-i mwuncey-ka issta.*
Cheli-DAT-(only)-NOM problem-NOM exist
'(Only) Cheli has problems.'

b *Na-nun Cheli-hanthey-(man)-ul mwuncey-ka isstako mitnunta.*
I-TOP Cheli-DAT-only-ACC problem-NOM exist think
'I believe only Cheli to have problems.'

Interestingly, although Yoon (2007) subscribes to the "major subject" analysis, he admits that in certain cases, an alternative parse is possible for such accusative-marked entities, as "many

of the matrix verbs that govern subject-to-object raising have a usage where they take the ACC-marked DP as an argument" (Yoon 2007:99), and these may be genuine proleptic structures. Furthermore, these argumental uses alternate with an adpositional version. A Korean example is shown in (44).

(44) *Na-nun Cheli-lul/ Cheli-eytayhay kunyesek-i tollassta-ko sayngkakhay.*
I-TOP Cheli-ACC/ Cheli-regarding that.guy-NOM crazy-COMP think.1SG
'I think of Cheli that the guy is crazy.'

Only when special care (like case-stacking in 43) is taken to rule out the proleptic parse can the "major subject"-analysis be firmly established.

It might be added, that according to Miura (2008), pronouncing the embedded subject is generally dispreferred in Japanese, but this effect can be ameliorated by adding a focus particle to the resumptive pronoun. This is comparable to the Hungarian situation, see example (55b) and footnote 68 in section 5.2.1.

(45) *Taroo-ga Ken-o kare-koso-ga tensaida to omotteita.*
Taroo-NOM Ken-ACC 3SG.MASC-FOC-NOM genius COMP thought.3SG
'Taroo thought of Ken_i that HE_i was a genius.'

In this dissertation I cannot do justice to the Korean and Japanese situation. While Yoon (2007) may be right that (39)-(43) is best analyzed not as prolepsis but as raising, (44) suggests that prolepsis may play a role in the syntax of these languages as well. Further research is needed for the exact differentiation between the two structures.

6.2.6 Summary of the crosslinguistic landscape

Having surveyed proleptic constructions in a number of languages, it is time to take stock. The following properties connect all these constructions:

i. There is a matrix verb that has 3 dependents: the subject, the proleptic element and a finite complement clause.
ii. The proleptic element bears a subject matter semantic relationship with the main verb.
iii. Because of this subject matter semantic relationship, the proleptic element must be a referential entity.

iv. There is a resumptive pronoun in the complement clause, co-indexed with the proleptic element.

The dimensions of divergence among the constructions are the following:

i. How productive is the proleptic pattern lexically? How many matrix verbs allow prolepsis? (fully productive, e.g. German, Madurese or more restricted, e.g. Hungarian, Greek)
ii. Functionally, is the proleptic element a (thematic) adjunct (as in German, English) bearing oblique case, or is it a direct object/oblique argument (as in Hungarian, Greek, Madurese)?
iii. Are there restrictions on the function and the distance of the embedded resumptive? (Greek: yes, other languages: no.)

It seems that every language surveyed has a version of prolepsis where the proleptic element is definitely not an argument of the main verb, but a thematic adjunct (elements marked with adpositions like *–ról*/*-ről* in Hungarian, *of* in English, *von* in German, *apo* in Greek, *eytayhay* in Korean). This is also suggested in den Dikken (2010:1) in a Minimalist framework. In this account the distinction between the delative and accusative version as prolepsis in Hungarian translates as the proleptic XP originating in either inside or outside the the matrix clause VP.

The version with thematic adjuncts (and the argumental oblique version in Hungarian) seems to be more productive than the direct object version of prolepsis. In Hungarian we observed that virtually all verbs allowing prolepsis with a direct object have an adpositional variant, while the reverse is not true, there are verbs that allow the proleptic element to be realized as an adpositional phrase, but they do not allow the accusative version. Also, in German and English the construction seems quite free, as Salzmann (to appear) also notes. In Madurese the two versions are equally free, at least it seems from Davies's (2005) account, though he is not very detailed about it. In Greek and Japanese, the accusative version is possible with a restricted set of verbs.

This suggests that the adpositional thematic adjunct version could be a base structure for the accusative version. If the predicate is able to take an adpositional dependent as a thematic adjunct, then an additional argument-structural mechanism of the kind proposed in section 5.2.3 can advance this relationship into a true argumental one, realizing the proleptic element as a direct object, making it a "derived argument", in Toivonen's (2013) sense. The productivity of the conversion seems to be a language-specific property. In Hungarian, Greek and Japanese it

is fairly restricted, but for example "in Madurese (…), there is a very productive applicative-type process that creates core (non-prepositional) arguments out of oblique arguments" (William D. Davies, p.c.). This seems to be on the right track, though the obliques should be thought of as thematic adjuncts, not arguments. The function of the proleptic element and its resumptive dependent in the embedded clause is quite free, apart from Greek, where it is restricted to the immediate subject.

6.3 Theoretical perspectives on prolepsis

Let us now change perspective and try to find a place for prolepsis in the general space of grammar. The Minimalist analysis offered by Salzmann (2006, to appear) is that prolepsis involves predication, with the aid of a null operator. In his approach, verbs taking a CP-complement are specified by a lexical redundancy rule that they can "take a CP whose head is specified for requiring a silent operator in it's spec" (Salzmann to appear:23). This silent operator turns the CP into a predicate, which then licenses the proleptic object as the subject of this predication. The predicate and the proleptic object together constitute a proposition that satisfies the semantic requirement of the main verb.

From an LFG-perspective, the null operator analysis is obviously problematic, since it involves a zero c-structural entity and such elements are not parts of the architecture. Apart from this theory-specific objection, Salzmann does provide a detailed and principled analysis for prolepsis in German which possibly extends to other languages as well. The compulsory nature of the co-referent entity constitutes a problem for an LFG account if it is accepted that the proleptic PP is an adjunct. It seems that distinguishing this adjunct from normal adjuncts by assuming it to be a thematic adjunct is motivated. However, it is not clear to me at this point how this translates to the obligatory presence of the co-referent entity. In Hungarian OF, this issue does not arise, since the lexical entry participating in OF was argued to be an added argument as a result of a lexical process, which may include the addition of such specifications. The proper modelling of predication-type dependencies of this type is a general challenge for LFG at this point, since the standard analysis involving XCOMPs and functional identification is not readily applicable to such cases as the complement clause has its own subject, ruling out functional identification. Thus, from an LFG perspective, another route has to be taken. I will argue that there is robust theoretical and empirical motivation for an alternative approach, where prolepsis is taken to be a subspecies of control-constructions.

6.3.1 Prolepsis & Control

To investigate prolepsis from the perspective of control, let us first present a broad theoretical overview of control. "Control" in the Chomskyian tradition is a term for constructions where the referent unpronounced subject of an embedded nonfinite clause is determined ("controlled") by a main clause subject or object. The unpronounced subject is called a PRO.

(46) a *John$_i$ tried [PRO$_{i/*j}$ to go]*.

b *I persuaded John$_i$ [PRO$_{i/*j}$ to go]*.

The overt one is called the controller, while the PRO is the controllee. The controllee is traditionally assumed to be restricted to the immediately embedded subject.

Control (46) is often contrasted with "raising" (47). The latter superficially mirrors the control-sentences in (46), but it is argued to have a distinct analysis, whereby the controller is originally an element of the embedded clause which subsequently moves to the main clause.

(47) a *John$_i$ seemed [~~John~~$_{i/*j}$ to go]*.

b *I believe John$_i$ [~~John~~$_{i/*j}$ to go]*.

It is a commonplace observation that *John* is semantically related to *try* and *persuade* but not to *seem* and *believe*. So while *John* is a "trier" in (46a) and a "persuadee" in (46b), he is not a "seemer" in (47a) or a "believee" in (47b).

Table 1 shows a collection of control and raising verbs (from Landau 2013:10, for more predicates see Davies and Dubinsky 2004:11-12).

	Subject	**Object**
Control	*try, condescend, promise, decide, plan, agree, hope, prefer, wonder, refrain*	*persuade, encourage, recommend, appeal, force, plead, order, urge, dissuade*
Raising	*seem, appear, turn out, happen, begin, continue, stop, likely, certain, sure*	*believe, consider, prove, show, take, expect*

Table 1.
Control and raising predicates.

Beneath the surface similarity, there is a wide range of properties that distinguishes the two constructions. Without the goal of completeness, we will survey here the most important ones.[79]

At least two features follow from the fact that the subject of *try* in (46a) is thematic, while the subject of *seem* in (47a) is non-thematic. First, expletive subjects can be associated with *seem* but not with *want*. (48b' shows that the problem is not simply that *want* cannot take a *that*-clause complement. (As discussed in section 3.2.2, the existence of object-expletives is contested so the *believe-persuade* contrast is not readily applicable here.)

(48) a *It seems that John is happy.*
b **It wants that John is happy.*
b' **It wants for John to be happy.*

Second, also because of their lack of individual meaning, idiom chunks can only appear in raising constructions, not in control ones.

(49) a *The cat seems to be out of the bag.* (idiomatic or literal)
b *The cat wants to be out of the bag.* (only literal)
(50) a *I believe the cat to be out of the bag.* (idiomatic or literal)
b *I persuaded the cat to be out of the bag.* (only literal)

The basic analytical distinction between equi and raising is that while in the case of control, the controller is entirely interpreted in the main clause (in GB/MP terms one may say that it is base-generated there), in raising the controller also sustains strong ties with the controllee-position.

In GB/MP, the analysis of raising is relatively uncontroversial. Putting the technical details aside, the standard analysis posits that the subsequent matrix subject is base-generated in the subject position of the embedded predicate and is moved ("raised") into the matrix clause subject/object nonthematic subject or object position.[80] For the analysis of control, there are two main approaches: a movement-based approach (Boeckx, Hornstein & Nunes 2010) and Landau's (2015) two-tiered account.

According the movement-theory of equi the analysis of control parallels that of raising. That is, the controller also starts out in the embedded clause and then moves into its final, main-clause position. The only difference is that in the case of control, both the embedded and the

[79] An extended survey is to be found in Landau (2013:8-28).
[80] Raising to subject is entirely uncontroversial, while Exceptional Case Marking (ECM) is sometimes offered as an alternative to raising to object. This debate largely stems from the theoretical layout of GB/MP, so I will not go into details here.

main clause positions are thematic (or A-positions, in Chomskyan terminology). This means that the constituent in question ends up with two theta-roles, so the original formulation of the Theta-criterion has to be loosened.

The other main approach is Landau's (2015) "two-tiered" model. In this model, the control may be established in two ways: either by predication (tier 1) or by variable binding (tier 2), where the second may be superimposed on the first one in attitudinal contexts. As we will see in the next section, Landau's view is actually quite close to the LFG-conception of control, where there are also two basic ways to establish a control relation.

In LFG, the term "control" is used more inclusively, for all constructions in which there is a lexically specified identification (functional/anaphoric) between two f-structure entities, so both raising and GB/MP control are covered by the term. What the GB/MP tradition calls "control" runs by the name "equi[81]" in LFG-literature. In other words, the Chomskyan tradition has a control-raising distinction and no general term for both of them, LFG has an equi-raising distinction and "control" as a general term for this range of phenomena. Since the chosen framework of this dissertation is LFG, I use the latter terminology.

LFG utilizes functional or anaphoric identification in the analysis of control constructions. The identification is specified in the lexical entries of the relevant predicates. Usually, only one of these entities is phonetically realized (in other words, only one of them has a c-structural exponent).

As mentioned in 2.3, functional identification means strict f-structural identity, one element provides value for two f-structural attributes. In the case of raising predicates, the subject or the object argument is nonthematic. The raised constituent is functionally identified with the embedded subject. This identification is specified in the lexical entry of *seem*, shown in (51).

(51) *seem* V <(XCOMP)> (SUBJ)
(SUBJ)=(XCOMP SUBJ)

The propositional argument is encoded as an XCOMP, the open complement, a grammatical function that allows its SUBJ to be provided from the outside. The f-structure for (47a) is provided in Figure 1.

[81] An abbreviation for "equivalent NP-deletion", a terminology from early Chomskyan analyses.

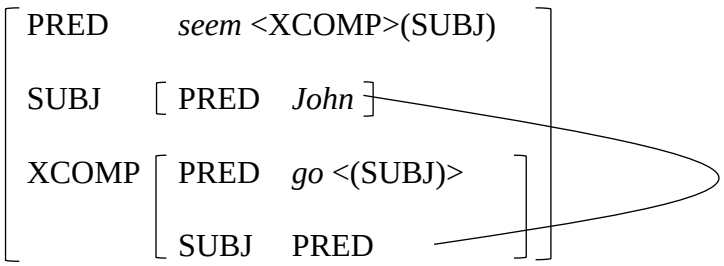

Figure 1.
F-structure of (47a) (*John seemed to go*.)

According to Falk (2001), equi-constructions can involve either functional or anaphoric identification, depending on the specification of the individual predicates. This is conceptually quite close to Landau's (2015) dichotomy of predicative/variable-binding control, as XCOMP is a predicative complement in LFG, while anaphoric identification actually involves binding a *pro* variable in the f-structure.

Falk (2001:137) proposes syntactic tests to determine which analysis is the correct one in individual cases. These are based on the difference between functional and anaphoric identification. Recall that functional identification is a strict syntactic identity, while anaphoric identification is more like a semantic connection.

> (…) in functional control, the controller must be present (in the f-structure) and it must be a core function. If it were absent, the subordinate clause would be incomplete because it would lack a SUBJ, and (...) only core functions can be specified by a control equation. On the other hand, neither of these is necessary for anaphoric control: pronouns need not have antecedents and there is no restriction on the grammatical function of an antecedent of a pronoun. On the other hand, anaphoric control should allow split controllers, because pronouns can take split antecedents; while a functional controller is the single element specified by the control equation. (Falk 2001:137)

Based on these, Falk (2001:138) shows that *try* involves functional identification, while *agree* utilizes an anaphoric one.

(52) a *It was tried (by the geneticist) to clone dinosaurs.*
b *It was agreed (by the geneticist) to clone dinosaurs.*

(53) a *The geneticist tried to clone dinosaurs.* (SUBJ of clone: the geneticist only)

b *The geneticist agreed to clone dinosaurs.* (SUBJ of clone: the geneticist + possibly other people)

What (53b) shows is a phenomenon called "partial control" (Landau 2000), where in addition to the matrix controller, other understood controllers are also possible. Similar predicates are e.g. *want, prefer, yearn, arrange, hope,* etc. Verbs that pattern with *try* are referred to as "exhaustive control" verbs. Other examples are *avoid, forget, fail, refrain, decline, neglect*, etc. The exact analysis of this phenomenon is currently under discussion in the linguistic literature, see Boeckx, Hornstein & Nunes (2010:182-190), Landau (2016), Piteroff, Alexiadou & Fischer (2015), also see Haug (2013) in an LFG framework. Without going into the details of the problem of partial control, it can be stated that any LFG analysis for it must be based on anaphoric identification, as the strict syntactic identity of functional identification would make any semantic mismatch impossible.

However, while the discussion in the previous paragraph does have the logical consequence that functional identification always means exhaustive control, it does not entail that all cases of exhaustive control should be viewed as involving functional identification. That is to say, while cases like partial control must be analyzed in terms of anaphoric identification, I would not like to suggest that anaphoric control can never have strict semantics regarding the control relationship. In fact, for cases of equi for which Falk (2001) proposes functional identification, Dalrymple (2001) puts forward an analysis in terms of "obligatory anaphoric identification". While I agree with Falk (2001) in that functional identification is more appropriate for the English examples at hand, I also agree with him in that obligatory anaphoric control is a possibility in LFG (Falk 2001:138). So while functional identification is a unitary phenomenon in terms of the semantics of the control relationship, anaphoric identification may have different types, ranging from obligatory through "quasi-obligatory" (Haug 2013) to arbitrary. The analysis of specific constructions in specific languages should boil down to careful investigation of these particular constructions and languages.

Let us make an interim summary at this point. So far we have discussed two aspects of variation in control constructions from an LFG-perspective. On the one hand, a control construction may involve a thematic controller and be equi or a nonthematic controller and be raising. On the other hand, the control relationship may be anaphoric or functional. Since functional identification can be thought of as an analogue for movement, while anaphoric identification can be likened to PRO-based GB/MP approaches, the LFG-architecture provides theoretical space for both main Chomskyan approaches to equi, without having to choose

between them. Such a flexibility is a long-standing merit of LFG and has been noted to be in line with linguistic diversity (Levinson & Evans 2010). As a result, the following taxonomy emerges.

CONTROL-TYPE		**Example**
Thematicity of controller	**Nature of identification**	
Equi	Anaphoric identification	canonical control, *agree*-type
	Functional identification	canonical control, *try*-type
Raising	Anaphoric identification	NOT EXPECTED
	Functional identification	canonical raising

Table 2.
An LFG-taxonomy of control. (first version)

The reason for anaphoric raising being not expected is that such a configuration would lead to a violation of Semantic Coherence. The LFG architecture excludes such a construction. Actually such an analysis was argued by Coppock (2003) for Hungarian OF, which I have analyzed as prolepsis. I have argued extensively that a prolepsis-account is preferrable both on theoretical and empirical grounds. Since anaphoric raising has not been put forward for other constructions in any language, the taxonomy is legitimate at this point.

So how does prolepsis fit into this picture? I have argued that the proleptic constituent is a thematic argument of the main verb, so it is an equi-type construction. Also, in order to account for the possible variation of the number feature of the embedded predicate I analyzed it as involving anaphoric identification. Consequently, prolepsis is related to the *agree*-type constructions. From this perspective, prolepsis is a kind of object-equi (like sentences with *persuade, encourage, force,* etc.), utilizing an obligatory anaphoric identification.

The primary feature in which prolepsis differs from canonical control-constructions is that it involves a finite complement clause. In earlier stages of generative grammar, it had been thought that control involves only non-finite complements. However, this has changed and there have been described several constructions in various languages that necessitate positing control into finite complements. If this is so, then the taxonomy presented in Table 2 can be expanded so that it encompasses finite versions of control.

CONTROL-TYPE			Example
Thematicity of controller	**Nature of identification**	**Finiteness**	
Equi	Anaphoric	Finite complement	**PROLEPSIS**
		Non-finite complement	canonical control, *agree*-type
	Functional	Finite complement	
		Non-finite complement	canonical control, *try*-type
Raising	Anaphoric	Finite complement	not expected
		Non-finite complement	not expected
	Functional	Finite complement	
		Non-finite complement	canonical raising

Table 3.
An LFG-taxonomy of control. (expanded version)

But is there justification for making a parallel between prolepsis and equi-type control relations? I think there is.

Let us see here a prototypical Hungarian proleptic example and its f-structure. The proleptic object is a matrix clause argument and it is co-referent with the embedded clause subject. The co-reference is lexically induced by the lexical entry of the proleptic verb containing the equation OBJ INDEX={COMP$^+$ GF*} GF INDEX. Then compare this with a garden-variety object-equi structure from English: (55) and Figure 3.

Apart from the information-structural side (which is a by-product of Hungarian sentence-structure) the similarity is evident. In both cases, the main clause thematic object is anaphorically identified with the subject of the embedded clause. Although there are some important differences between the standard case of control and prolepsis, there are crucial parallels as well. Because of these similarities I think it is warranted to treat prolepsis as a special subtype of control.

(54) *Jánost mondtad, hogy jön.*
John.ACC said.2SG that(C) comes
'John you said that he will come.'

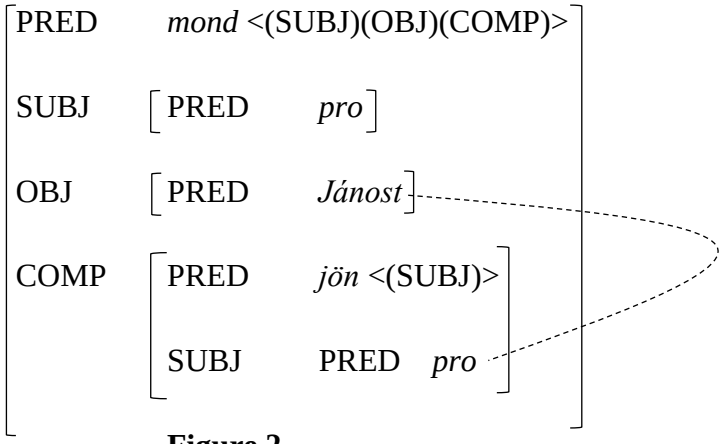

Figure 2.
F-structure of (54)

(55) *John persuaded Mary to go.*

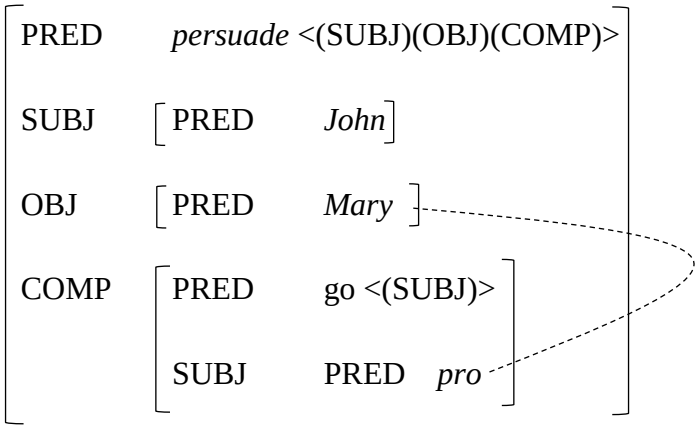

Figure 3.
F-structure of (55)

First let us see the differences. One difference is that equi is traditionally seen as involving non-finite complement clauses, while prolepsis involves a finite one. However, this standard view is false, control can involve certain finite embedded clauses (Landau 2013), so this is not a real difference, but only a typological variation. We will discover such cases in section 6.3.2 (and we will also clarify what is meant by a clause being "(non-)finite").

Another one is that standard control is a local dependency between a main clause argument and an immediately embedded clause. Prolepsis, on the other hand, can be long-distance.

(56) a *John persuaded Mary to kiss herself.*

b **John persuaded Mary to say that he should kiss herself.*

c *Jánost mondtad, hogy hallottad, hogy jön a partira.*
John.ACC said.2SG that(C) heard.2SG that comes the party.to
'John you said that you heard that he is coming to the party

A third difference is that in standard equi, the controlled element is always the embedded subject (57a), while prolepsis can involve a variety of grammatical functions (57b-c).

(57) a **I persuaded Mary for John to like.*

b *Két almát mondtál, hogy vettél.* (embedded OBJ)
two apple.ACC said.2SG.INDEF that(C) bought.2SG.INDEF
'Two apples you said that you bought.'

c *Párizst mondtad, hogy oda utazol.* (embedded OBL)
Paris.ACC said.2SG.DEF that(C) there travel.2SG.
'(About) Paris you said that you will go there.'

That prolepsis involves an anaphoric and not a functional link is probably relevant for these two differences. Anaphoric dependencies involve an f-structural *pro* in LFG, which is no different in terms of the PRED feature from ordinary pronouns. That is, the inherent distinction between PRO and *pro* in GB/MP is absent from LFG. As pronouns are not restricted to find their antecedents in the immediately containing clause, a long-distance anaphoric dependency is actually not that surprising.

As for the grammatical function of the controllee, in the theory of subjects in LFG framework in Falk (2006b), it is established that lexically specified functional control equations can only refer to subjects, because of a restriction by Universal Grammar. However, no such constraint has been put forward in connection with anaphoric dependencies. This can provide space for the conceptual possibility of prolepsis as anaphoric identification with a range of grammatical functions.

Now let us take stock of the similarities, apart from the analytical similarity shown in connection with Figure 2 and 3. Landau (2013), surveying the properties of equi, shows that in control, the controllee (the embedded subject) behaves as a bound variable. Consider the pair in (58). For context, imagine that there is a regulation that everyone in the army is only allowed to surrender if given an order to do so. Suppose there are two commanders, Bill and John. If self-ordering is sufficient for surrender, then Bill may surrender in the case of (58a) but not necessarily in the case of (58b). (More straightforward examples can be construed with subject control (e.g. *Only John claimed [that he won]/[claimed to have won])*, but I want to keep examples as parallel as possible.)

(58) a *Only Bill ordered himself to surrender.*

b *Only Bill ordered that he should surrender*.[82]

So while (58b) is ambiguous between the reading in which *Bill* is the only person who ordered *Bill* to surrender (the so-called "strict reading") and the one where nobody but *Bill* issued an order for himself/herself to surrender (the bound variable"/"sloppy") reading, (58a) only has the second interpretation. Prolepsis behaves like this, as attested by (59a), which can mean that *János* was the only person who claimed self-victory, not that nobody else claimed *János* to be the winner. Compare this with (59b), which is ambiguous.

(59) a *Csak János mondta magát, hogy (ő) nyert.*
only John said.2SG himself.ACC that(C) he won.3SG
'Only John said (of) himself that he had won.'

b *Csak János mondta azt, hogy (ő) nyert.*
only John said.2SG that.ACC that(C) he won.3SG
'Only John said that he had won.'

A similar phenomenon can be observed in elliptical contexts. In equi, the elided part is interpreted sloppily.

(60) a *Mary expected to attend the ceremony, and Sue did too (expect to attend the ceremony).*

b *Mary encouraged Paul to attend the ceremony, but not David (encourage to attend the ceremony).*

The unexpressed subjects of the elided parts are understood as *Sue* and *David*. So (60a) does not mean that *Sue* expected *Mary* to attend the ceremony, and (60b) cannot mean that *Mary* encouraged *David* so that *Paul* attends the ceremony.

Again, prolepsis shows this pattern: (61) cannot mean that "you said of Péter that János comes", the agent of *jön* ('come'*)* must be *Peter*.

(61) *Én Jánost mondtam, hogy jön, te pedig Pétert (mondtad, hogy jön).*
I John.ACC said.1SG that(C) comes you but Péter.ACC said.2SG that(C) comes
'John I said that he comes and you did too Peter.'

[82] Also, there is an irrelevant interpretation, where *he* refers to an outside person.

The similarities are rooted in the fact that some obligatory equi constructions use the same kind of anaphoric dependency that is utilized in prolepsis.

Consequently, I deem it motivated to locate the typological place of prolepsis in the realm of control, as an instance of obligatory anaphoric finite equi. Given Landau's (2015) analysis, where predication is involved in control, it may turn out in the long run that a convergence with Salzmann's (2006, to appear) predicative approach to prolepsis is more than a remote possibility.

In the remaining part of the dissertation, I will explore the taxonomy in Table 3 and present examples from various languages to fill it with. These sections do not intend to provide an exhaustive theoretical account for control in general. That would be far outside of the scope of this dissertation. Rather, I set out to provide an empirical overview from the perspective of the dissertation, though some analytical suggestions will nevertheless be made.

6.3.2 Control in the finite domain

In order to discuss control in the finite domain, first we have to clarify what is meant by the term "finite". Surprisingly, although the term has been used for a long time in the study of language and linguists in general seem to have an intuitive understanding of what it is, the exact meaning is elusive and is quite difficult to pin down.

There are at least four dimensions that have been evoked in the literature for the characterization of finiteness: tense, agreement, the ability to host an own subject, and the capability to assign nominative case to this subject. Much of the difficulty arises because these dimensions are largely independent and it is not clear whether a specific property on each of these dimensions is required for a clause to be classified as "finite", or partially right specifications are enough. Also it is not settled whether any of these dimensions should be regarded as a primary trait of finiteness. Even if a primary trait is delimited, there remains the question of how the other dimensions are related to this primary trait and to each other. Some progress has been made, but it cannot be said that a consensus has been reached. No answers for these pressing questions will be provided in this dissertation. What will be provided is an overview of structures that can be classified as finite according to some of these criteria and some remarks from the theoretical perspective of LFG.

Before briefly describing the aforementioned dimensions, it must be indicated at this point that although in some versions of Chomskyan syntax the option of finite control is a

difficulty,[83] this is not the case in LFG. As far as no other principle of LFG is violated, a clause having tense, agreement, etc. specifications has nothing to do with individual predicates instantiating control-equations in the f-structures. So if anything is expected by LFG regarding finite control constructions is that we should find them.

So let us briefly introduce the ingredients of finiteness. "Tense" is one such ingredient. Having a tense-specification on a predicate is usually taken as an indication of its finiteness. A complicating issue is that tense can be defined as a morphosyntactic property (overt tense-marking morphemes) or as "semantic tense", that is, being able to establish independent time reference. The first case is self-explanatory, as for the second, consider English infinitives. These are not marked for tense morphologically but since Stowell (1982) it is known that they can be divided into two groups along their ability to host independent temporal reference. Consider (62).

(62) a *#Yesterday, John tried/dared/managed/forgot/avoided/failed/got to solve the problem tomorrow.*

b *Yesterday, John wanted/agreed/preferred/arranged/hoped/refused to solve the problem tomorrow.*

The second property on our list is agreement, by which we mean morphological person/number marking on the predicate. According to the theory of Landau (2013), tense and agreement exhaust the list of finiteness-ingredients.

The property of hosting a subject is sometimes also mentioned as a defining trait of finite clauses (Subirats-Rüggeberg 1990). To make this dimension fully operational, one has to factor different conceptions of subjects into it. In GB/MP raising sentences do have a trace/copy in the subject position of infinitival clauses, while equi sentences are commonly analyzed as a PRO occupying these positions.[84] Also, some analyses regard the *for*-phrases associated with infinitivals (*I hope for Bill to win*) and the objects of "Exceptional Case Marking" sentences (*I believe John to be happy*) as some sort of subjects.[85] In LFG, a further question that arises regarding this criterion is that what the relevant level is for checking the presence of a subject: c- or f-structure. The subjects of XCOMPs are "empty" only in the sense that they are predicated

[83] For raising, the main clause being able to check case, tense and agreement features is assumed to block the movement of elements out of these clauses, unless other unchecked features (*wh*, focus, etc.) motivate the operation.

For equi, the controllee in standard GB/MP is PRO, an intrinsically zero nominal element. According to the "PRO theorem", PRO must be ungoverned, but the subject position of any finite clause is governed.

[84] Also, one must count with the "pro" subjects of pro-drop languages.

[85] Of course such cases may be argued not to be proper subjects. My aim is not to adjudicate on these issues, I just would like to highlight the surrounding issues.

from the outside via functional identification. Otherwise, the subject is f-structurally "fully there". (This is paralleled by the copy-theory of movement in Minimalism.) In anaphoric identification, the complement clause (bearing a COMP grammatical function) has its independent "pro" subject, which in LFG theory is no more special than the zero subjects of pro-drop languages, but this "pro" is absent in the c-structure. The reverse situation is less likely, since any c-structural element would contribute some attribute-value pair to the f-structure, unless it is excluded by some special mechanism.

Finally, in the theory of Cowper (2016) it is not enough that a clause hosts a subject, it is the subject bearing structural case that is the crucial diagnostics of finiteness. It is not entirely clear whether structural case only involves the nominative or subjects with accusative, dative or oblique cases are also relevant. For Cowper, finiteness has no semantic content itself so features like "tense" can only be indirectly associated with it.

After this familiarization with the concept of finiteness, let us investigate some finite control structures in various languages. At any rate, prolepsis is unquestionably finite since the complement clause has independent tense, agreement and can host its subject both in f- and c-structure.

The schedule is the following. First, we will look at English and some Bantu languages which have raising constructions out of finite clauses. Actually this is a small sample of languages with such constructions, for an extensive discussion, see Ademola-Adeoye (2010). Then there will be a discussion of finite equi-constructions in Hungarian, Greek, Serbo-Croatian, Finnish and Turkish. Finally, before concluding the chapter, a summary of these finite control constructions will be outlined.

6.3.2.1 Finite raising

The basic f-structural setup for a raising construction needs the following ingredients: a predicate with a non-thematic argument and an XCOMP. The athematic argument is functionally identified with the empty subject function of the XCOMP. The identification is encoded in the lexical entry of the predicate in the form of an equation like (SUBJ)/(OBJ)=(XCOMP SUBJ).[86] Since the defining property of XCOMPs is this functional identification and not their c-structural categorial expression (Asudeh 2002), the c-structural categorial status of the XCOMP is irrelevant: it can equally be a VP, an IP or a CP. In these latter cases, the clause may bear a full range of tense and agreement features. If the embedded

[86] In the languages surveyed, it was the matrix SUBJ that got identified with the embedded subject. However, Ademola-Adeoye (2010) provides a wide range of examples for finite object-Hyperraising structures.

subject is independently expressed, we have what the literature calls a "Copy raising" construction, if it is not, we have "Hyperraising". We will see that e.g. English instantiates the first type (63a), while Bantu languages the second one (63b).

(63) a *Richard$_i$ seems like he$_i$ smokes*

b *Chisaang'i chi-lolekhana mbo chi-kona*
animal SM-seem that(C) SM-sleep.PRES
'The animals seem to be sleeping.'

A complication with Copy raising is that the PRED feature of the embedded copy pronoun clashes with the PRED feature of the matrix subject. In this sense, Hyperraising is less problematic for an analysis, since the SUBJ of the complement clause is not and cannot be expressed. The question is that what kind of deeper reasons lie beneath these patterns.

My proposal is that the obligatoriness of a c-structurally filled subject position is crucial in this respect. That is to say, in English we must have a copy pronoun simply because the language is such that requires the subject position of an indicative clause to be filled. This is dubbed as the famous EPP-requirement in GB/MP. For this requirement, let us use the name "EPP", without commitment to the GB/MP-theory as a whole. The only way to satisfy the EPP and the lexically encoded functional identification equation in English is to insert a c-structural subject in the complement clause. Although this leads to a violation of Semantic Consistency, this solution is arguably better than the alternative, violating the EPP. In languages where the EPP is not as strong as in English, the consistency-violation is not necessary. Thus in Bantu (and in other languages with Hyperraising), the presence of the embedded subject is covert. There is no element in the sentence that would provide a conflicting PRED value. Since there is no independent motivation for the presence of such embedded subjects, the only thing their presence would contribute to the construction is a violation of Semantic Consistency. Thus, they are ruled out.

This leads us to a typological prediction: pro-drop languages (languages with no strong EPP requirement) should instantiate Hyperraising, while non-pro-drop languages (with strong EPP, like English) should have Copy raising. What I am saying is that Copy raising and Hyperraising are essentially the same structure: finite raising with functional identification. The differences then follow from language-particular properties. This seems to be on the right track: Ademola-Adeoye (2010:vi) makes the same assertion if we disregard the different theoretical persuasion:[87]

[87] I found Ademola-Adeoye's (2010) work after reaching the conclusion myself.

> An important empirical generalisation, first noted by Ura (1994), which is empirically supported by the data discussed in this thesis, is that if a language has Hyperraising or Hyper-ECM, it is also a pro-drop language. On the basis of this generalisation, it is argued that Hyperraising and Hyper-ECM constructions involve the use of resumptive *pro* in the embedded subject position, while languages with Copy raising and Copy ECM use overt pronouns. Apart from this difference, it is argued that these A-movement constructions are identical in all crucial respects.

While the cited works and my research radically differ in theoretical background, our conclusion is the same: finite versions of raising are possible, and the overt presence of the controlled subject is dependent on the status of the pro-drop parameter in a given language. This convergence is welcome. Let us now turn to the detailed discussion of these constructions.

6.3.2.1.1 English

In English, although there are differences among infinitival clauses in their capabilities of hosting an independent tense (as in 62), obligatory control is generally regarded as limited to non-finite contexts. There is one construction, however, that might be analyzed as a finite raising structure: the so-called Copy raising construction. It is demonstrated in (64). As can be seen from the example, Copy raising (CR) can occur with verbs related to perception. In Copy raising, these take a finite complement clause (with agreement and independent tense) headed by *like/as if* and there is an obligatory co-referent pronoun (the "copy pronoun") in this complement clause.

(64) *John$_i$ seems/appears/looks/feels/sounds like he$_{i/*j}$ is going to win the race.*

A complicating factor in CR is that in certain cases the main clause subject might be regarded as a thematic argument of the main verb. Although the most recent LFG-analysis of CR by Asudeh & Toivonen (2012) explicitly denies this possibility, claiming that the CR-subject can only bear a "semantic role" (a weaker relation than thematic roles), Landau (2011) convincingly argues that an argumental analysis is simpler and more satisfactory for these cases of CR. In fact, if these predicates assign a "perceptual source" thematic role to their subjects, a copy in the complement clause is no longer necessary. In this case, the concept of control is simply unnecessary.

(65) *John seems/appears/looks/feels/sounds like Mary is going to win the race.*

Of course, pragmatic difficulties may arise when the perception of the main clause subject cannot be plausibly used for inference about the event of the embedded clause, but these restrictions are not syntactic. The main predicates in such cases have dyadic lexical entries like (66), where the SUBJ is associated with a perceptual source thematic role and the COMP is a proposition.

(66) *seem/appear/…* V <(SUBJ)(COMP)>

So these cases must be disregarded as no-control scenarios. However, there are cases where the matrix clause subject is non-thematic. This should come as no surprise since the predicates under consideration all can occur with expletive *it* subjects.

(67) *It seems/appears/looks/feels/sounds like/that John is going to win the race.*

Landau (2011) shows that in cases when because of world-knowledge or contextual reasons, the main clause subject cannot possibly be construed as a perceptual source, a different scenario emerges: the resumptive element becomes obligatory.[88]

(68) *The problem$_i$ sounds like it$_{i/*j}$ is difficult to solve.*

So in the following discussion pertains only those cases of Copy raising that are of this second type, where an obligatory control relationship is involved. As the name suggests, this a raising-type of control relationship, the main clause subject does not bear any thematic relationship to the main verb. This can be seen from the fact that expletive-like elements and idiom chunks may participate in CR (Potsdam & Runner 2001).[89]

(69) a *It seems like it's raining harder than it is.*
b *There looks like there's gonna be a riot.*
c *The shit appears as though it's going to hit the fan very soon.*

The other property of CR that likens it to regular (non-finite) raising constructions is that it uses functional identification. Fuji (2005) observes that data from binding argues for reconstruction into the infinitival clause. From his theoretical perspective, this argues for A-movement. In LFG, this is a hallmark of functional identification.

[88] This is the "Psource-Copy Generalization": a copy is not necessary only if the subject is a Psource.
[89] There could be some dialectal variation in the judgment about these sentences.

(70) *[Stories about each other$_i$]$_j$ seem like they$_j$ have frightened [John and Mary]$_i$.*

Based on these, it could be plausible to represent CR as the following (simplified) f-structure. Figure 4 is essentially like a standard raising f-structure.

(71) *Richard$_i$ seems like he$_i$ smokes.*

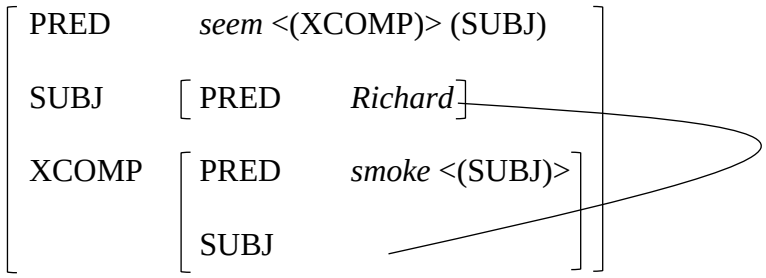

Figure 4.
A proposed f-structure of (71).

This intuitive approach is not without problems though. Let us begin with the lesser problem. In the original formulation of LFG's control-theory (Bresnan 1982), XCOMPs are VPs. They lack a structural subject-position, which enables them to having their subject specified from the outside via a functional identification. However, here there is a subject position in the complement clause, standardly assumed to be Spec/IP or TP in English. However, if one follows Asudeh (2002) in abandoning the XCOMP-VP association and assuming that "the defining property of XCOMP is not its c-structural category, but rather whether it contains a grammatical function that is the target of a functional control equation", this problem can be bypassed.

A more serious problem is that although I have conveniently left the XCOMP SUBJ empty in Figure 4, allowing it to be targeted by functional identification, in fact the attribute-value pairs contributed by the pronoun in the complement clause should be there, including a PRED feature, which is barred from functional identification. Thus, as it stands now, an analysis like Figure 4 results in violation of Semantic Consistency.

As an alternative, one may adopt the analysis provided by Asudeh (2002). In that, the main clause subject is not directly associated with the embedded copy, but the prepositional connective elements *like/as* function as predicates that can take a thematic subject. Then this thematic subject is anaphorically identified with the copy pronoun. Thus, the f-structure would be the following (Figure 5).

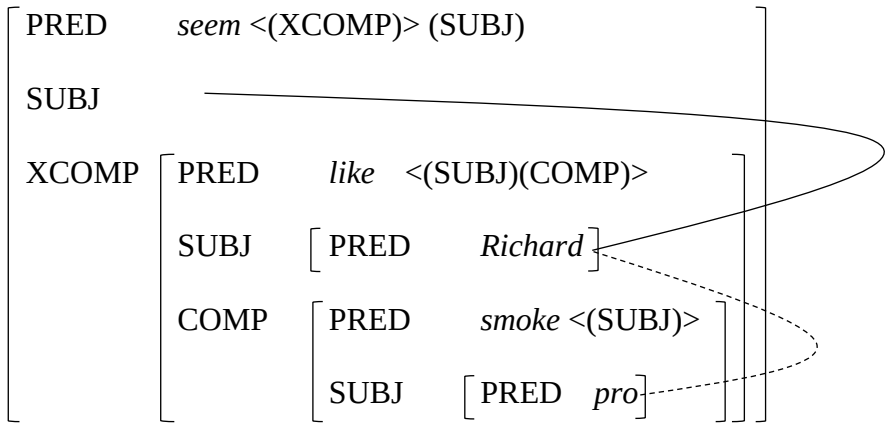

Figure 5.
Asudeh's (2002) f-structure for (71).

This alternative is not without problems either. The root of its problems is the unintuitive lexical entry for the intermediary predicate *like*. Although Asudeh (2002) argues for the desirable conceptual unity of this *like* with the usual prepositional use of *like* (*John is like a monster*), the unification does not survive a closer scrutiny. The usual *like* has an object, as can be seen from the accusative form of the personal pronoun after it (*John is like him/*he*). Also the kind of thematic role that Asudeh's (2002) *like* would assign to this SUBJ is quite obscure. Even if these problems could be evaded, this *like* would certainly not be the same lexical entry as the usual *like*, since an obligatory control equation should be added to this instance of *like* to ensure the referential identity of its SUBJ and the SUBJ of its COMP. An equation of this sort is obviously lacking in the lexical entry of the usual *like*. Alternatively, as in Asudeh (2012), the dependency with the copy pronoun is not established by such an equation, but falls out from the semantic composition of the sentence. In this case, however, the pronoun constitutes an extra semantic resource, which has to be removed by "manager resources". This is essentially the same kind of problem that my proposal faces regarding Semantic Consistency, which means that Asudeh's proposal has no advantage over mine from this aspect.

Additionally, to handle cases like (69), an alternative lexical entry for *like* is stipulated by Asudeh (2002), which takes a nonthematic subject and an XCOMP complement. Finally, in subsequent work, Asudeh & Toivonen (2012) argue that even though CR subjects do not receive a thematic role, they receive a "semantic role" as perceptual sources. How this semantic role works in relation to expletives and idiom chunks remains a mystery. If thematic adjuncts can be thought of as kinds of elements with semantic but not thematic roles (as Rákosi 2012 suggests), then semantic roles are not compatible with nonreferential elements. (72) illustrates this with experiencers for which a thematic adjunct-analysis is justified (Rákosi 2006b, 2012).

(72) a *#It seemed important for the shit to hit the fan.*

b *#It is important for the cat to get his tongue.*

Based on such considerations and because of conceptual simplicity I argue that an analysis of English CR which results in the f-structure in Figure 4 is a viable alternative.

Let me make some remarks about the lexical entries of Copy raising verbs and about some of their c-structural consequences. As already mentioned, the lexical entries are essentially the same as the lexical entry for the standard raising versions of these verbs, with an <(XCOMP)>(SUBJ) grammatical function requirement. However, in these uses, the XCOMP is realized as a CP[90], while in CR it is realized by prepositional elements (*like*, *as if*, etc.). This could be a problem on the view that PPs are canonically associated with OBL functions. However, as mentioned in section 2.3, there is a debate in LFG about the status of COMP and XCOMP, some arguing that they are special OBLs. If these proposals are on the right track, it should come as no surprise that these grammatical functions may have PPs as their c-structural realizations.

As a final point, as mentioned at the beginning of this section, there exists a version of these verbs where the subject is thematic (having a perceptual source thematic role) and there is also a propositional COMP function. A similar subcategorization, although with a non-thematic SUBJ, is independently required for sentences like (73).

(73) *It seems that Richard smokes.*

In an attempt to provide some links between these lexical entries, I hazard the suggestion that a perceptual source element is a conceptual possibility for these verbs, which either can be realized as a thematic adjunct or a genuine argument. The thematic adjunct is marginally possible in English, but such entities are fully productive for example in Swedish in the form of *på*-PPs (Asudeh & Toivonen 2012).

(74) a *??It seems from Richard that he smokes.*

b *Dat verkar på Tom som om han har vonnit.*
it seems on Tom as if he has won
'Tom gives the impression that he has won.'

[90] I follow Falk (2002:139-140) in assuming that infinitival *to* is a complementizer.

It could be that the perceptual source thematic adjunct becomes integrated into the predicate in a lexical/argument-structure process similar to the one I proposed for prolepsis in section 5.2.3, where an alternation like this happens with the subject matter semantic dependent. This means that CR-subjects, just like proleptic objects are derived arguments in the sense of Toivonen (2013).

6.3.2.1.2 Bantu

As Carstens & Diercks (2013) note, many Bantu languages exhibit apparent raising out of a finite clause. Such configurations are often referred to as "Hyperraising". The finiteness of these clauses is manifested by the possible agreement and tense features inside them. Zeller (2006) shows one such construction in Nguni, while Carstens & Diercks (2013) demonstrate these with Lubukusu and Lusaamia. Nguni Hyperraising is restricted to one predicate, the modal-like *fanele* ('ought to'), while Lubukusu and Lusaamia involve standard raising predicates like *lolekhana* ('seem') and *bonekhana* ('appear'). Another difference between these languages is that Nguni Hyperraising takes place from a subjunctive complement clause, while Lubukusu and Lusaamia raise out of an indicative clause. Traditionally subjunctives are regarded as "less finite" than indicative clauses, since subjunctives often have limited tense-assigning capabilities (they are often restricted to future-orientation or a tense dependent on the main verb). Although according to Zeller (2006), the tense of the subjunctive clause in the Nguni Hyperraising construction is independent, it is safe to say that the structures in Lubukusu and Lusaamia are even "more finite" than the Nguni ones.

Apart from these differences, a similar set of data from these languages supports the hypothesis that these languages do have finite raising constructions. The Hyperraising sentences (the "b" examples below) have non-raising alternates (the "a" examples) with truth-conditional equivalence, so the in the "b" sentences, the main verb subject can be "reconstructed" into the complement clause. (The glosses are somewhat simplified for more focus on substantial issues. E.g. representation of the complex agreement system is omitted. The reader is referred to the cited works for full glosses.)

(75) a *Ku-fanele ukuthi uMdu a-khulume isiZulu name.* (Nguni)
LOC-ought that(C) Mdu SM-speak.SBJV Zulu with.1SG
'Mdu should speak Zulu with me.'

b *UMdu u-fanele ukuthi a-khulume isiZulu na-mi.*
Mdu SM-ought that(that) SM-speak. SBJV Zulu with.1SG
'Mdu should speak Zulu with me.'

(76) a *Ka-lolekhana mbo chisaang'i chi-kona.* (Lubukusu)
SM-seem that animal SM-sleep.PRES
'It seems that the animals are sleeping.'

b *Chisaang'i chi-lolekhana mbo chi-kona*
animal SM-seem that(C) SM-sleep.PRES
'The animals seem to be sleeping.'

(77) a *Bi-bonekhana koti eng'ombe chi-ng'were amachi.* (Lusaamia)
SM-appear that(C) cow SM-drink water
'It appears that the cows drank the water.'

b *Eng'ombe chi-bonekhana chi-ng'were amachi.*
cow SM-appear SM-drink water
'The cows appear to have drunk the water.'

Both Zeller (2006) and Carstens & Diercks (2013) show data from passivization in support of the claim that these constructions are parallel to traditional raising constructions. Zeller (2006) demonstrates the truth-conditional equivalence of sentence pairs where an embedded object is passivized into the main clause through a raising verb (78). This is a characteristic of raising constructions, see the truth-conditional equivalence of English raising sentences in (79a-b). The same is not true for equi sentences (79a'-b').

(78) a *Udokotela u-fanele ukuthi a-bhek-e isiguli.*
doctor SM-ought that(C) SM-examine-SBJV patient
'The doctor must examine the patient.'

b *Isiguli si-fanele ukuthi si-bhek-w-e ng-udokotela*
patient SM-ought that(C) SM-examine-PASS- SBJV by-doctor
'The patient must be examined by the doctor.'

(79) a *John seems to visit Mary.* a' *John tries to visit Mary.*

b *Mary seems to be visited by John.* b' *Mary tries to be visited by John.*

Besides, Carstens & Diercks (2013) show that under the assumption that *suubil* ('believe') in Lubukusu does not assign a thematic role to its object, the thematic role of *omukeni* ('guest') can only originate from the embedded verb, proving the ties to the embedded clause (80).

(80) *Omukeni ka-asuubil-wa mbo k-ola*
guest SM-believe.PAST-PASSIVE that(C) SM-PAST.arrive
'The guest was believed to have arrived.'

Finally, Zeller (2006) shows that idiom chunks may be “raised” in Nguni. Although Carstens & Diercks (2013) does not provide such data, it can be safely assumed that the same is true for Lubukusu and Lusaamia.

(81) a *Izandla zi-fanele ukuthi zi-gez-an-e.*
hand SM-ought that(C) SM-wash-REC-SBJV
‘It’s vital that two people do something for each other.’ Lit.: ‘It’s vital that one hand washes another.’

In the previous section, for English Copy raising it constituted an analytical difficulty that the embedded pronoun provided a PRED feature, which should have been identified with the main clause subject. Neither Zeller (2006) nor Carstens & Diercks (2013) mentions the possibility of overt embedded subjects. In fact, according to personal communication with Vicky Carstens, such elements are unlikely to occur. Consequently, the consistency-problem associated with English CR does not arise in Bantu so these structures can unproblematically be represented along the lines of Figure 4 in the previous section.

6.3.2.2 Finite equi

As a reminder, equi constructions are those constructions where the controller is a thematic argument of the main verb. It enters a dependency with the (usually covert) subject of a complement clause. Just like raising, equi is canonically associated with non-finite complements. In spite of this widespread generalization, as Landau (2013:88) notes, “the existence of finite control was recognized already in the 1980s as a pervasive feature of the Balkan languages”. As we will see apart from the Balkan cases, a number of other languages also display such configurations, including Hungarian, Greek and Turkish.

6.3.2.2.1 Hungarian

There are two phenomena in Hungarian that are relevant for the topic of finite equi. One is the existence of inflected infinitives; the other is the possibility to have overt subjects in certain infinitival clauses.

Inflected infinitives only occur if the controller bears dative case and has an experiencer thematic relation to the predicate. According to Rákosi (2006b), inflection on the infinitive is possible if the infinitival clause hosts its own subject, which can be translated as anaphoric identification with a controlled *pro* subject. This is also in line with Falk (2001:138) who states

that since "anaphoric control involves an ordinary COMP function, an overt SUBJ should be an option."

However, there is another logical possibility: the case of no-control. That is, it could happen that the dative dependent fully belongs to the complement clause, serving as its subject and therefore no control relation is present in the sentence. To illuminate the issue let us take the predicate *fontos* ('important'). The two conceptually possible analyses for a sentence with this predicate and an inflected infinitive are shown in (82). (The "PROs" in the string are only for expository purposes.)

(82) a *Fontos* *Jánosnak$_i$* *[PRO$_i$* *hazamen-ni-e]*.
important John.DAT home.go-INF-3SG
'It is important for John to go home.'

b *Fontos [Jánosnak hazamennie]*.

Rákosi (2006b), agreeing with É. Kiss (2002), shows that (82b) is the proper analysis. This is supported by the observation that (82) does not necessarily mean that *János* ('John') judges it to be important to go home, it could someone else's opinion. This stems from the fact that the dative is not an obligatory dependent of *fontos* ('important'). *Fontos* (and many other predicates with experiencer semantics) are essentially monadic in nature. This is reflected in the bracketing of (82b). However, they may license a dative dependent as a thematic adjunct. The thematic adjunct may then control the embedded subject, but then the infinitival agreement is not possible. So if one wants a construal where it is necessarily *János* who judges his going home as important, one can only get that with a non-inflecting infinitive (this crucial interpretational difference is mirrored by the translation). Note that *János* is not an argument, the morphology of such a dependent is not fixed, a PP-alternative with *számára* ('for') is possible (Rákosi 2006b).

(83) *Fontos* *Jánosnak$_i$/ János$_i$* *számára [PRO$_i$* *hazamen-ni-(*e).*
important John.DAT John for.3SG home.go-INF-3SG
'John judges it to be important for him to go home.'

Because of this, this class of predicates is not really suitable for investigating finite control constructions in Hungarian. However, there exists a number of predicates which do take a dative constituent as an argument which establishes a control relationship with the infinitival clause. Such predicates are *számít* ('matter'), *derogál* ('feels derogatory to do') or *nem akaródzik* ('does not feel like'), *sikerül* ('succeed'). According to Rákosi (2006b), the datives associated with

these predicates are arguments. This for example can be seen from the fact that such datives are not replaceable with *számára*-PPs (*számít/derogál/nem akaródzik Jánosnak/ *János számára*). With these, the dative is always a matrix clause element, so the structural representation should always be like along the lines of (82a). This in turn predicts that no inflected infinitive is expected for these predicates under Rákosi's (2006b) assumptions. On the other hand, since these are subject clauses, anaphoric control is expected for them, since functional identification as a control relation is uniquely associated with the XCOMP grammatical function. So there might be grounds for expecting inflection. It turns out that infinitival inflection is possible with these predicates. Although Rákosi (2006b:219) marks a sentence like (84a) ungrammatical, I do not share this intuition. Also, the examples in (84b-c) are entirely acceptable.

(84) a *Jánosnak derogál velem mutatkoz-ni-(a).*
János.DAT feels.derogatory with.me show.up-INF-3SG
'It feels derogatory for John to show up with me.'

b *Jánosnak nem akaródzik hazamen-ni(-e).*
János.DAT not want.MIDDLE.3SG home.go-INF-3SG
'John does not feel like going home.'

c *Jánosnak sikerült hazamen-ni(-e).*
János.DAT succeeded home.go-INF-3SG
'Going home was successful for John.'

It must be added that even Rákosi (2006b:218, footnote 13) admits that if the dative argument is implicit, the sentences are quite acceptable. While such an observation may have some merit to it, I doubt that it would boil down to any fundamental configurational difference. In my view, inflected infinitives like (84) are acceptable, be the controller implicit or explicit.

Recall that the anaphoric dependency in proleptic OF enabled an agreement-mismatch between the matrix proleptic argument and the embedded predicate. As it turns out, a similar phenomenon may be observed with the constructions in (85). Note however that in contrast to OF, quantification in this case actually bars agreement-mismatch.

(85) a *A lányoknak derogált hazamen-ni-e / hazamenniük.*
the girls.DAT felt.derogatory home.go-INF-3SG home.go-INF-3PL
'It felt derogatory for the girls to go home.'

b *A összes lánynak derogált hazamen-ni-e / *hazamenniük.*
the every girl.DAT felt.derogatory home.go-INF-3SG home.go-INF-3PL
'It felt derogatory for every girl to go home.'

An open question concerns the optionality of infinitival agreement. Cases like (84) suggest that it is entirely optional, without noticeable difference in meaning. On the other hand, in the case of thematic adjunct controllers, there is a noticeable semantic difference between the sentences with and without agreement. Although the two cases are differentiated by the argument or adjunct nature of the controller, the control-mechanism itself is the same, so the cause of the difference remains obscure. I leave this for further research.

The other constellation in which finiteness of a control clause is relevant is the possibility to have overt subjects in infinitival clauses. It was Szabolcsi (2009) who observed that if the infinitival subject is under the scope of some operator, it is possible to pronounce the subject. Note that in the absence of the focus operator, this is ungrammatical. (86a) shows this with a raising structure (Szabolcsi convincingly shows that there is a raising construal for *(el)kezd* ('begin')), while (86b) is an example with equi.

(86) a *János elkezdett [*(csak) ő kapni kitüntetéseket].*
John began.3SG only he get.INF awards.ACC
'It began to be the case that only John got awards.'

b *János szeretne [*(csak) ő énekelni].*
John would.like.3SG only he sing.INF
'John would like it to be the case that only he sings.'

Interesting data may be gained if one puts together the two Hungarian constructions under discussion. One observation is that these overt subjects cannot license infinitival agreement.

(87) a **János elkezdett csak ő kap-ni-a kitüntetéseket.*
John began.3SG only he get-INF-3SG awards.ACC

b **János szeretne csak ő énekel-ni-e.*
John would.like.3SG only he sing-INF-3SG

If Rákosi (2006b) is right in that a proper subject in the complement clause enables the agreement to appear, this might suggest that these subjects are somehow deficient. Note also that in the case of *sikerül* ('succeed') or *fontos* ('important'), where the controller is in the dative, the overt infinitival subject also bears this case. There, the possibility to license infinitival agreement is not impaired by the overt subject.

(88) *Jánosnak sikerült/ fontos volt csak neki hazamenni(e).*
János.DAT succeeded.3SG important was only him.DAT home.go.INF(.3SG)
'John succeeded in that only he went home.'/ 'It was important for John that only he goes home.'

Szabolcsi (2009) notes that it is quite cumbersome to construe sentences with the meanings of (86) and (88) but without the overt nominative in the infinitival clause, so it might be that these overt subjects are some last-resort elements. This somehow could lead to their less than full f-strucural presence. There are technical ways in LFG to remove unwanted f-structural elements, such as the use of the restriction operator (Kaplan & Wedekind 1993) or the use of "manager resources" (Asudeh 2010), but working out an analysis is left for future research.

I conclude this section with a few remarks about the interaction of overt infinitival subjects with predicates that take only an optional dative thematic adjunct. This produces some interesting data since the word-order change triggered by the focus-operator exposes some configurational issues. Consider the data in (89).

(89) a *Fontos [csak Jánosnak hazamen-ni-*(e).]*
important only John.DAT home.go-INF-3SG
'It is important that only John goes home.'

b *Csak Jánosnak fontos hazamen-ni-(e).*
'It is important that only John goes home.'/ It is important only for John that he goes home.

Given the compulsory pre-predicate position of focussed elements in Hungarian *csak Jánosnak* ('only John.DAT') can only belong to the complement clause in (89a), which in turns triggers the compulsory agreement. In (89b), we can have two parses. One is similar to the one already mentioned, the difference being that the dative is scrambled to the beginning of the sentence. Also, *csak Jánosnak* ('only John.DAT') may be a thematic adjunct associated with *fontos* ('important') and it controls into the complement clause. In this case, no agreement is necessary.

Also, if the dative is a thematic adjunct, belonging to the main clause, an additional dative can appear in the infinitival clause as its overt subject. In (90) the agreement would indicate the presence of *János* in the complement clause, leaving no position for the other dative subject *csak neki* ('only for him').

(90) *Fontos Jánosnak csak neki hazamen-ni-(*e).*
important John.DAT only for.him home.go-INF-3SG
'It is important for John that only he should go home.'

6.3.2.2.2 Greek

Greek is one of the first languages for which a finite control analysis was posited. A recent overview of Greek finite equi is to be found in Spyropoulos (2007), on which this section is based. Greek allows equi with subjunctive complements, which may count as finite because of the full range of agreement marking in them. However, their temporal properties are restricted: according to Spyropoulos (2007), they are either fixed in temporal reference (a future-oriented reading) or have "anaphoric tense", which is determined by the temporal characteristics of the main clause. The first type is called "dependent subjunctives" (DS) by Spyropoulos (2007), while the second one is labelled as "anaphoric subjunctives" (AS). I adopt this terminology.

AS shows what one would expect in the case of obligatory anaphoric control. The anaphoric nature is reflected in the possibility of an overt subject in certain contexts (91a), while the obligatory nature is manifest in the impossibility of partial control (91b).

(91) a *o janis$_i$ kseri na xorevi ki aftosi/$_{*j}$ kalo tsamiko*
the John.NOM know SBJV dance.3SG and he good tsamiko
'John knows how to dance tsamiko well too.'

b **i zoi emaθe na kolimbane*
the Zoe learned.3SG SBJV swim.3PL
Intended, approx.: 'Zoe learned to swim with others.'

DS displays some interesting differences. Apart from licensing the kind of overt subjects that can be seen in (91a), it also allows an unusual version of partial control. Normally, in partial control, only semantic, but not syntactic plurality is licensed, see (92a), where the reciprocal cannot appear. Unlike this, in Greek DS, syntactic plurality is also licensed (92b), as can be seen from the plural agreement on *pane* ('go.3PL').

(92) a *John told Mary that he preferred to meet (*each other) at 6.*

b *episa ti maria na pane a psonia tin triti*
persuaded.1SG the Maria.ACC SBJV go.3PL for shopping the Tuesday
Intended, approx.: 'I persuaded Maria that they should go for shopping on Tuesday.'

This argues for a looser version of anaphoric control. I think the reason for this possibility is that anaphoric control may come in various degrees of strictness, as it was already mentioned in section 6.3.2. The number feature of the complement subject seems to be a point of variation.

That Greek DS is on the lenient side of anaphoric control is very plausible in the light of the fact that control can be suspended with certain DS predicates.[91]

(93) *o janis prospathise na erθun, ala afti ðen ta kataferan*
the John tried.3SG SBJV come.3PL but they NEG them manage.3PL
'John tried for them to come, but they didn't make it'

On the basis of (93) Spyropoulos (2007) notes that it is not without merit to argue that Greek DS are simply a no-control scenario. However, he maintains the distinction for two reasons. One, not every DS predicate allows control suspension, it is dependent on the semantics of the individual predicates. Two, if the feature specification of the matrix controller matches that of the controllee, independent reference is not possible (as in 94a), unlike genuine cases of no-control (94b). I follow his conclusion and regard Greek DS as a "liberal" type of anaphoric control.

(94) a *o janis$_i$ prospaθise na PRO$_{i/*j}$ ftasi noris*
the John tried.3SG SBJV arrives early
'John tried to arrive early.'

b *o janisi$_i$ elpizi na PRO$_{i/j}$ ftasi noris*
the John hopes SBJV arrives early
'John hopes that s/he will arrive early.'

Based on the flexible nature of Greek subjunctive control, the availability of this special type of partial control, the possibility of overt subjects and control into arguably indicative clauses, Spyropoulos (2007:174) suggests that control in Greek does not involve a "PRO", but a "*pro*". Hence, he proposes that "the hypothesis that the controlled null subject of Greek (…) subjunctives is a controlled "*pro*" breaks the strict association between control and PRO. It implies that control is not a property of PRO only and that other categories can also be controlled". As the "PRO" vs. "*pro*" distinction is nonexistent in LFG to begin with, such a proposal is certainly a promising step towards theoretical convergence.

[91] According to Spyropoulos (2007:166), "there is a variation with respect to the availability of control suspension, which depends on the semantics of the matrix predicate."

6.3.2.2.3 Serbo-Croatian

It was Zec (1987) who first called attention to the possibility of finite control in Serbo-Croatian. Like Greek, it involves subjunctive clauses, with agreeing verb forms inside. Tense is constrained to present forms.

(95) a *Petar$_i$ je pokusao da PRO$_{i/*j}$ dodje na vreme.*
Peter AUX tried.3SG that(C) comes on time
'Peter tried to come on time.'

b *Ana je naterala Mariju$_i$ da PRO$_{i/*j}$ dodje.*
Anna AUX forced.3SG Maria.ACC that(C) comes
'Anna forced Marija to come.'

That these kind of sentences involve anaphoric, rather than functional identification is shown by two facts. One, an overt complement subject is possible in the form of an emphatic pronoun (96). This is comparable to the case of overt infinitival subjects in Hungarian.

(96) a *U takvim situacijama se odman pokusa da se pobegne.*
in such situations one immediately tries that(C) one escapes
'In such situations one immediately tries to escape.

b *Ana$_i$ je naterala Mariju$_j$ da one$_{*i/j}$ dodje.*
Anna AUX forced.3SG Maria.ACC that(C) she comes
'Anna forced Marija that she should come.'

Two, split control, where the controllee is controlled by two matrix clause elements, is possible (97). This is not the same as the previously mentioned partial control, since now the matrix arguments together exhaustively specify the controllee. Also, the licensing of syntactic plurality is the standard, rather than an exception (Landau 2013:172-174). Like partial control, this is unexpected under functional identification, which states the full identity of strictly two f-structural elements. From an LFG perspective, this can be seen as another type in the possible manifestations of anaphoric control.

(97) *Petar je naterao Mariju da zajedo pobegnu.*
Peter AUX advised.3SG Maria.ACC that(C) together escape.3PL
'Peter advised Mary to escape together.'

These properties argue for an independent, controlled anaphor in a COMP grammatical function required by these control predicates.

6.3.2.2.4 Turkish

In Turkish, there is a set of verbs which occur with a direct object and a complement clause, the subject of which is controlled by the direct object. This is the description of a standard object-control structure. The only peculiarity of the Turkish construction is that the complement clause has full tense and agreement specifications. In other words, this is a finite object-control structure.

(98) *Ben Ali-yi yarın balığı yiyecek sanıyordum.*
I Ali-ACC tomorrow fish eat.FUT.3SG assumed.1SG
'I thought that Ali will eat the fish tomorrow.'

Sentences like (98) had been analyzed as Exceptional Case Marking-constructions, with *Ali-yi* ('Ali-ACC') being part of the complement clause (e.g. Aygen 2002), or as *Ali-yi* having moved from the embedded clause into the matrix clause (e.g. Özsoly 2001). However, Ince (2006) shows that neither analysis is correct: in Minimalist framework, the analysis proposed is that the accusative element is base generated as the main clause object. I just cite one piece of evidence for each claim, the reader is referred to Ince's (2006) work for a full discussion.

As (99a) shows "Turkish does not permit scrambling of any element out of embedded finite clauses to the right periphery of matrix clauses" (Ince 2006:210), but the accusative element can be scrambled to the right periphery, it must be part of the main clause (99b).

(99) a **Dinleyiciler biz içtik sanıyorlar viski-yi.*
auditors we drank.1PL assuming.3PL whisky-ACC
Intended: 'The auditors believe we drank the whisky.'

b *Hasan Ayşeden nefret ediyoruz sanıyor biz-i.*
Hasan Aysen.ABL hatred do.3PL assuming.1PL we-ACC
'Hasan thinks that we hate Ayşe.'

That the accusative element is base-generated in the matrix clause is can be seen by using a familiar test: idiom-chunks, as they are anomalous in this construction. That is to say, the object is a thematic argument of the main verb.

(100) a *Alinin anası bellendi.*
Ali.GEN his.mother was.screwed.3SG
Idiomatic: 'They really messed up Ali.' Literal: 'They raped Ali's mother.'

b *Alinin anası-nı bellendi sanıyordum.*
Ali.GEN his.mother-ACC was.screwed.3SG assumed.1SG
Only idiomatic: 'I thought that Ali's mother was raped.'

Ince (2006) considers a proleptic analysis for this construction, but rejects it on the grounds that unlike prolepsis, it is restricted to embedded subjects, so it must involve a "PRO" rather than a "*pro*". From my perspective, it is true that prolepsis is much more flexible than standard object control (as discussed in section 6.3.1) in that it allows for nonsubjects to be controlled and a greater distance between the controller and the controllee, but the fundamental analytical idea is the same: a thematic matrix object identified with an embedded argument. Since Ince (2006) mentions no possibility of overt embedded subjects or partial control, I assume this is a case of functional identification. This in turn gives some explanation for its restrictedness, given Falk's (2006) approach to functional identification, mentioned earlier.

6.3.2.3 Summary of finite control

Now that we have surveyed a number of languages with finite control constructions it is time to put the pieces together and by doing so, assemble the bigger picture of control-typology from an LFG-perspective.

The starting point has been that the LFG architecture provides a solid framework and points of references for building a typology like this. From the earliest days, the two fundamental axes of variation have been recognized: the thematic relationship of the controller to the main verb, and the nature of the identification process. If the controller element is a thematic argument, we talk about "equi" structures, if it is athematic, we talk about "raising". The nature of the identification may be either functional, which is the strict identity of two f-structural elements or anaphoric, which is a referential dependency. The former is instantiated into the open complement function XCOMP, while the latter involves the closed complement function COMP. Furthermore, while functional identity always requires full syntactic and semantic matching, the anaphoric link may be of various strength, ranging from strict obligatory control, through "quasi-obligatory" control (in English, Haug 2013) to weak/no-control (as in Greek DS). What I have done is supplement this taxonomy with another fundamental concept of grammar, finiteness.

Now we are able to present a populated version of Table 3 from section 6.3.1.

CONTROL-TYPE			Example
Thematicity of controller	**Nature of identification**	**Finiteness**	
Equi	Anaphoric identification	Finite complement	argumental prolepsis, inflected infinitives in Hungarian, Greek and Serbo-Croatian control
		Non-finite complement	"*agree*-type" canonical control
	Functional identification	Finite complement	Turkish object control
		Non-finite complement	"*try*-type" canonical control, standard Hungarian control
Raising	Anaphoric identification	Finite complement	not expected
		Non-finite complement	not expected
	Functional identification	Finite complement	Copy raising in English, Bantu Hyperraising
		Non-finite complement	canonical raising

Table 4.
An LFG-taxonomy of control. (final version)

The starting point was that although prolepsis differs from standard cases of equi in a number of important properties (distance, grammatical functions), they share the fundamental analytical idea: a thematic grammatical function in the main clause is anaphorically identified with an embedded one. Although the strictness of the anaphoric link varies in the constructions in the slot, it transpires in the variations of number features of the controllee (singular/plural variation in certain cases of proleptic OF, partial control in Greek, split control in Serbo-Croatian). Also, the complement clause is a full, independent clause (COMP), hosting its own subject. I have also identified a finite version of functionally identified equi in Turkish. These are straightforward extensions of the already established cases of canonical equi into the finite domain.

Raising has also been shown to cut across the distinction of finiteness. Finite raising structures exist, not just in exotic Bantu languages, but arguably in English too, in the form of Copy raising, which is actually Hyperraising for languages with compulsory overt subjects. The LFG architecture readily accommodates the typology. Moreover, the slot that is unexpected given the theoretical assumptions of LFG (raising with anaphoric identification) is still empty, which is a welcome result for this research paradigm.

6.4 Conclusion of chapter 6

This chapter has provided some theoretical and cross-linguistic perspective of the left-peripheral discourse-related constructions that have been the topic of the dissertation. Three main configurations were distinguished: Fronting proper, left-dislocations and prolepsis.

In proper fronting constructions, a left-peripheral element is "extracted" from a sentence-internal position, but it maintains strong syntactic ties to its original position via functional identification. The configurations in this category are English Topicalization and case-retaining Hungarian Operator fronting ("fronting" OF). Additionally, certain Finnish and German fronting constructions were also argued to be manifestations of Fronting proper. There is a tendency for such structures to be interpreted contrastively, but this is not strictly necessary.

In Left-dislocation constructions, the peripheral discourse-prominent element may be related to a pronominal associate. There are two subtypes. One is syntactically integrated left-dislocations, where the link between the fronted element and the pronoun is grammatically encoded and enforced: these are Hungarian Contrastive and Non-contrastive left-dislocation. German Weak pronoun left-dislocation may also be located here. The other left-dislocation type comprises of "hanging topics": some "radically" left peripheral element is generated as a "syntactic orphan" and is only related to the host sentence via pragmatic inference. English Left-dislocation and Hungarian Contrastive focus left-dislocation are the prime examples for this. Additional examples are Topic fronting in Hungarian and German Hanging topic left-dislocation.

Finally, in the case of prolepsis, some element is inserted as a matrix-clause argument and is related to an embedded grammatical function via obligatory anaphoric identification, as it was argued to happen in case-switching Hungarian Operator fronting. This construction could be viewed as a subcase of control.

Taking another, broader perspective, it seems that a left-peripheral discourse-related constituent may be related to the clause in the following ways:

- syntactic dependency
    - functional identification
    - anaphoric identification (structurally or lexically encoded)[92]
- pragmatic dependency ("syntactic orphans")

Table 5 summarizes the picture.

	Syntactically integrated	**Pronoun associate**	**Nature of association**
TOP-Eng	yes	no	functional
LD-Eng	no	not necessarily	pragmatic
CTLD-Hun	yes	yes (by definition)	anaphoric (structurally encoded)
NCLD-Hun	yes	yes (by definition)	anaphoric (structurally encoded)
CFLD-Hun	no	yes (by definition)	pragmatic
"Fronting" OF	yes	no	functional
Proleptic OF	yes	possibly (usually no)	anaphoric (lexically encoded)

Table 5.
Clause-initial discourse-related constructions.

The emerging picture is not unlike what den Dikken (2010) presents as the "typology of A'-dependencies". Disregarding Left-dislocations (as they lie outside of his concern), he claims that A'-dependencies may be established in three ways: successive-cyclic movement via *v*P–edges, resumptive prolepsis, and scope marking. While in this dissertation there is no correlate for scope-marking constructions, the first two may be easily related to the LFG-view presented here. Functional identification is conceptually the LFG-correlate of successive-cyclic movement (though the technical details are obviously different). Den Dikken (2010) is not particularly explicit about the details of prolepsis, he largely follows Salzmann (2006), but adds that that the proleptic argument may be generated either as an adjunct or as an argument ("the proleptic XP originates in the matrix clause, either inside or outside the VP"). For prolepsis,

[92] In the case of left-dislocations, the anaphoric link between the contentful element and its associate pronoun is established via the rules of the sentence structure, while in prolepsis, the dependency is part of the lexical entries of the individual predicates.

Salzmann's (2006, to appear) Minimalist and my LFG-based implementation differs substantially (predication vs. anaphoric control), though convergence of the approaches in the future is possible, given Landau's (2015) predicative theory of control.

What is not discussed in den Dikken (2010) are the extrasentential syntactic orphans at the left-periphery, which are only related to the host sentence via pragmatic inference.

CHAPTER 7

CONCLUSION AND FUTURE PERSPECTIVES

In this dissertation I have investigated several clause-initial ("fronting") discourse-related sentence types in English (Topicalization, Clause-initial adjuncts, Left-dislocation) and Hungarian (Left-dislocation, Operator Fronting) within the framework of Lexical-Functional Grammar.

After a general outline of the aims and scope of the dissertation in Chapter 1, Chapter 2 presented the theoretical framework, Lexical-Functional Grammar. The most attention was given to information-structure, for which a new feature-based taxonomy was offered. This new system is aimed to be a fruitful amalgamation of earlier theoretical taxonomical work. It is based on the notions of relational newness, discourse-structuring capacity and contrastiveness (as a subcase of the discourse-structuring feature) and these features classify the following i-structural categories: Topic, Contrastive topic, Information focus, Contrastive focus, Completive information and Background information.

In Chapter 3, I outlined my view of the relevant Hungarian grammar. The basis for it was the recent work of Laczkó (2014a, 2014b, 2015), which in turn is an adaptation of É. Kiss (1992). Hungarian sentence-structure is seen as a hybrid-model: a hierarchical preverbal field and a flat postverbal area. The preverbal field is the discourse-structural prominent part of the sentence. Under an exocentric S-node we find an iterative "topic-field", quantifiers, and a Spec/VP "focus-position". I supplemented Laczkó's framework with my proposals about information-structure.

As for subordinate clauses, I argued contra the standard view (put forward in Kenesei (1992/1994) that the optional associate pronouns of subordinate clauses are not expletives, but contentful demonstratives, as Tóth (2000) first argued. I supported my view with theoretical, empirical and typological arguments.

In Chapter 4, I investigated the left-peripheral discourse-related constructions in English: Topicalization, Clause-initial adjuncts and Left-dislocation. The first was argued to be a genuine fronting construction, involving functional identification and the connectivity-effects that follow from this. Information-structurally, it marks +CONTRASTIVE categories. The analysis of Clause-initial adjuncts was split between event-internal adjuncts which behave like topicalized elements and frame-setting ones, which are "base-generated" on the left-periphery and do not engage in clause-internal dependencies. Unlike TOP-Eng, CIADJ-Engs are not

necessarily contrastive. The third structure, English Left-dislocation was argued to be an extrasentential syntactic "orphan", functionally unintegrated into the host sentence.

Chapter 5 was about the analysis of the Hungarian constructions: various Left-dislocations and Operator fronting. Both turned out to involve various subtypes. LD-Hun was split between the syntactically integrated Contrastive topic/ neutral Topic-related structures and the focus related version, which should receive an extra-clausal analysis like LD-Eng.

Operator fronting constructions involve a dichotomy of proper fronting constructions, which involve functional identification and thus are syntactic correlates of TOP-Eng structures and prolepsis, in which a thematic matrix argument is anaphorically identified with some embedded function. Information-structurally, OF is more uniform: it is involved in the marking of +D-STRUCTURING categories, except for neutral Topics. A detailed analysis, including various levels of linguistic representation was provided for each of the constructions discussed.

Finally, chapter 6 discovered the cross-linguistic and theoretical perspectives of the constructions discussed throughout the dissertation. First, I examined some proper fronting constructions and showed that it is not unusual for languages to dislocate constituents to the left periphery of the clause, in order to give them discourse-prominence. Such processes are often accompanied by contrastive interpretations (as in Finnish and certain German configurations) but this is not necessary (as in Hungarian OF or other structures in German). German was also invoked as a basis for comparison of constructions with a pronominal associate ("Left-dislocation"). It was shown that a binary split of Left-dislocation is common: the left-peripheral element may either be a "hanging topic" or it may be syntactically plugged into the host sentence.

Proleptic structures were also surveyed in a number of languages and then I investigated how prolepsis fits into the space of theoretical syntax. I concluded that from an LFG-perspective, prolepsis may be looked at as a specific subtype of "control": finite, long-distance obligatory anaphoric equi. From this stance, I explored the landscape of control constructions in general, putting the focus on less-studied, finite control constructions (including both raising- and equi-type structures).

A number of issues remain undiscovered and thus potential targets for future research. Apart from the specific issues that were mentioned throughout the dissertation, a number of additional points are to be added here.

For instance, it is to be seen how the different taxonomies that I have presented (information-structure, control) can withstand the test of time. As I have identified Topicalization as a construction that makes explicit use of the feature +CONTRASTIVE, the Hungarian preverbal position (the "focus-position") as a locus for [+NEW ∨ +D-STRUCTURING],

hopefully other constructions may be identified which refer to other possibities e.g. +D-STRUCTURING or –NEW.

As for control, it is predicted that certain slots of the taxonomy (anaphoric raising) should remain empty. Time will tell whether they do so. Also, more research on the relationship of control and prolepsis is needed (especially in an LFG framework) to consolidate my proposal of a unified approach for them. Furthermore, anaphoric identification has been argued to have different subtypes, on the basis of the strictness of the connection between the controller and controllee. The precise characterization of these subtypes, their relatedness and the overall ramifications of this for the LFG-theoretic anaphoric control are yet to be investigated.

I have argued extensively why a proleptic account of the kind presented here is plausible for certain cases of Operator fronting, and I gave an outline of similar constructions in a number of languages. However, an essential part of the analysis is clouded at this point. It was claimed that an argument-structural process can turn certain thematic adjuncts into derived arguments (as a delative dependent turns into a proleptic object, or perceptual source adjuncts turn into real subject arguments in Copy raising), but exactly how, why and when, with what constraints such argument structural processes may happen is opaque.

It is also not clear at this point what the precise analysis of non-integrated syntactic orphans should be. LFG's multi-level correspondence architecture offers a conceptually appealing framework for such an analysis. To do so, detailed work on the formal representation of discourse-structure should be carried out.

Finally, the lineup of clause-initial discourse-related constructions in the target-languages is not complete, especially regarding English. I have not discussed Inversion (*rarely do I see John*), *it*-clefts (*it was John that I invited*) or *wh*-clefts (*what I want is a nice cup of tea*). All of these have interesting syntactic and information-structural properties which are yet to be explored.

REFERENCES

Ademola-Adeoye, Feyisayo Fehintola (2010). *A Cross-linguistic Analysis of Finite Raising Constructions*. Doctoral dissertation, University of KwaZulu-Natal.

Alba-Salas, Josep (2004). Lexically selected expletives: evidence from Basque and Romance. *SKY Journal of Linguistics* 17: 35-100.

Alsina, Alex (2008). A theory of structure-sharing: focusing on long-distance dependencies and parasitic gaps. In Butt, Miriam and Tracy H. King (eds.), *Proceedings of the LFG08 Conference* (pp. 5–25). Stanford: CSLI Publications.

Alsina, Alex, KP Mohanan & Tara Mohanan (2005). How to get rid of the COMP. In Butt, Miriam and Tracy H. King (eds.), *Proceedings of the LFG05 Conference* (pp. 21-41). Stanford: CSLI Publications.

Alrenga, Peter (2005). A sentential subject asymmetry in English and its implications for complement selection. *Syntax* 8(3): 175-207.

Ariel, Mira (1988). Referring and accessibility. *Journal of Linguistics* 24: 67-87.

Asudeh, Ash (2009). Adjacency and locality: A constraint-based analysis of complementizer-adjacent extraction. In Butt, Miriam & Tracy H. King (eds.), *Proceedings of the LFG09 Conference* (pp. 106-126). Stanford: CSLI Publications.

Asudeh, Ash (2011). Towards a unified theory of resumption. In Rouveret, Alain (ed.), *Resumptive Pronouns at the Interfaces* (pp. 121-187). Amsterdam: John Benjamins.

Asudeh, Ash (2012). *The logic of pronominal resumption*. Oxford: Oxford University Press.

Asudeh, Ash & Ida Toivonen (2012). Copy raising and perception. *Natural Language and Linguistic Theory* 30(2): 321-380.

Aygen, Gülşat (2003). Extractibility and the nominative case feature on tense. In Özsoy, A. Sumru, Didar Akar, Mine Nakipoglu-Demiralp, Eser E. Taylan, and Aksu Koc (eds.), *Studies in Turkish Linguistics* (81-94). Istanbul: Bogazici University Press.

Baloghné Nagy Gizella (2013). *Left dislocation in Optimality Theory*. Doctoral dissertation, ELTE.

Bejar, Susana & Diane Massam (1999). Multiple Case Checking. *Syntax*, 2(2): 65-79.

Bennis, Hans (1986). *Gaps and dummies*. Dordrecht: Foris.

Belletti, Adriana & Luigi Rizzi (1988). Psych-verbs and Θ-theory. *Natural Language and Linguistic Theory* 6: 293-352.

Berman, Judith (2001). On the cooccurrence of *es* with finite clause in German: an LFG analysis. In Meurers, Walt D. and Kiss Tibor (eds.), *Constraint-Based Approaches to Germanic Syntax* (pp. 7-30). Stanford: CSLI Publications.

Biberauer, Theresa (2008). Semi null-subject languages, expletives and expletive *pro* reconsidered. In Lash, Elliott, Yi An Li and Thomas Rainsford (eds.), *Cambridge Occasional Papers in Linguistics 4* (pp. 24-43). Cambridge: University of Cambridge.

Boeckx, Cedric, Norbert Hornstein & Jairo Nunes (2010). *Control as movement*. Cambridge: Cambridge University Press.

Bresnan, Joan (1978). A realistic transformational grammar. In Halle, Morris, Joan Bresnan and George A. Miller (eds.), *Linguistic Theory and Psychological Reality*. Cambridge, MA: MIT Press.

Bresnan, Joan (ed.) (1982). *The Mental Representation of Grammatical Relations*. Cambridge, MA: MIT Press.

Bresnan, Joan, Ash Asudeh, Ida Toivonen & Stephen Wechsler (2016). *Lexical-functional syntax*. Wiley Blackwell.

Bresnan, Joan & Jonni M. Kanerva (1989). Locative inversion in Chichewa: a case study of factorization in grammar. *Linguistic Inquiry* 20(1): 1-50.

Breul, Carsten (2004). *Focus structure in generative grammar*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Brody, Michael (1990). Remarks on the order of elements in the Hungarian focus field. In Kenesei István (ed.), *Approaches to Hungarian Vol. 3* (pp. 95-122). Szeged: JATE.

Browning, Maggie A. (1996). CP recursion and that-t effects. *Linguistic Inquiry* 27(2): 237-255.

Butt, Miriam & Tracy H. King (1996). Structural topic and focus without movement. In Butt, Miriam and Tracy H. King (eds.), *Proceedings of the LFG96 Conference*. Stanford: CSLI Publications.

Butt, Miriam & Tracy H. King (1999). *The Status of Case*. University of Konstanz and Xerox PARC.

Butt, Miriam, Tracy H. King, María-Eugenia Niño & Frédérique Segond (1999). *A Grammar-Writer's Cookbook*. Stanford: CSLI Publications.

Büring, Daniel (1997). *The Meaning of Topic and Focus – The 59th Street Bridge Accent*. London: Routledge.

Büring, Daniel (2003). On d-trees, beans, and b-accents. *Linguistics & Philosophy* 26(5): 511–545.

Cardinaletti, Anna & Michal Starke. (1994) The typology of structural deficiency. On the three grammatical classes. *University Venice Working Papers in Linguistics* 4, 41-109.

Carstens, Vicki & Michael Diercks (2013). Parameterizing case and activity: hyper-raising in Bantu. In Kan, Seda, Claire Moore-Cantwell, Robert Staubs (eds.), *Proceedings of the 40th Annual Meeting of the North East Linguistic Society* (pp. 99-118). Amherst: University of Massachusetts Graduate Linguistic Student Association.

Cattell, Ray (1978). On the source of interrogative adverbs. *Language* 54: 61-77.

Chafe, Wallace R. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Charles Li (ed.), *Subject and Topic* (pp.27-55). New York: Academic Press.

Choi, Hye-Won (1996). *Optimizing structure in context*. Doctoral dissertation, Stanford University.

Chomsky, Noam (1981). *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, Noam (1986). *Knowledge of Language, its Nature, Origin, and Use*. New York: Praeger.

Cook, Philippa & John Payne (2006). Information structure and scope in German. In Butt, Miriam and Tracy H. King (eds.), *Proceedings of the LFG06 Conference* (pp. 124-144). Stanford: CSLI Publications.

Coppock, Elizabeth (2003). Sometimes it's hard to be coherent. In Butt, Miriam & Tracy H. King (eds.), *Proceedings of the LFG03 Conference* (pp. 126-143). Stanford: CSLI Publications.

Cowper, Elizabeth (2016). Finiteness and Pseudofiniteness. In Eide, Kristin M. (ed.), *Finiteness matters: On finiteness related phenomena in natural languages* (pp. 47-77). Amsterdam: Benjamins.

Dalrymple, Mary (1993). *The Syntax of Anaphoric Binding*. Stanford: CSLI Publications.

Dalrymple, Mary (2001). *Lexical Functional Grammar*. San Diego: Academic Press.

Dalrymple, Mary & Irina Nikolaeva (2011). *Objects and information structure*. Cambridge: Cambridge University Press.

Dalrymple, Mary & Louise Mycock (2011). The prosody-semantics interface. In Butt, Miriam and Tracy H. King. (eds.), *Proceedings of the LFG11 Conference* (pp. 173-193). Stanford: CSLI Publications.

Dalrymple, Mary, Ronald M. Kaplan, John T. Maxwell III & Annie Zaenen (eds.) (1995). *Formal issues in Lexical-Functional Grammar*. Stanford: CSLI Publications.

Davies, William D. (2001). Against raising in Madurese (and other Javanic languages). *Proceedings of the Chicago Linguistics Society* 36: 57-69.

Davies, William D. (2005). Madurese prolepsis and its implications for a typology of raising. *Language* 81: 645–665.

Davies, William D. & Stanley Dubinsky (2004). *The grammar of raising and control – A course in syntactic argumentation*. Blackwell Publishing.

Davies, William D. & Stanley Dubinsky (2009). On the Existence (and Distribution) of Sentential Subjects. In Gerdts, Donna B., John C. Moore and Maria Polinsky (eds.), *Hypothesis A/Hypothesis B – Linguistic Explorations in Honor of David M. Perlmutter* (pp. 111–128). Cambridge, MA: MIT Press.

Declerck, Renaat (1988). *Studies on Copular Sentences, Clefts and Pseudoclefts*. Dordrecht: Foris.

den Dikken, Marcel (2010) *On the strategies for forming long A'-dependencies: Evidence from Hungarian*. Unpublished manuscript, Ms. CUNY Graduate Center.

den Dikken, Marcel & Surányi Balázs (to appear). Contrasting contrastive left dislocation explications. Linguistic Inquiry.

Emonds, Joseph E. (1976). *A Transformational Approach to English Syntax*. New York: Academic Press.

É. Kiss Katalin (1992). Az egyszerű mondat szerkezete In Kiefer Ferenc (ed.), *Strukturális magyar nyelvtan 1. – Mondattan* (pp. 79-177). Budapest: Akadémiai Kiadó.

É. Kiss Katalin (ed.) (1995). *Discourse Configurational Languages*. New York: Oxford University Press.

É. Kiss Katalin (1998). Identificational Focus versus Information Focus. *Language* 74: 245-273.

É. Kiss Katalin (2002). *The Syntax of Hungarian.* Cambridge: Cambridge University Press.

É. Kiss Katalin (2005). Eseményszerkezet és topik-predikátum tagolás a magyar mondatban. *Nyelvtudományi Közlemények* 101: 160–172.

É. Kiss Katalin & Pintér Lilla (2014). Identificational focus revisited. Paper presented at CL50 (the 50th anniversary meeting of the Chicago Linguistic Society).

Enc, Mürvet (1991). The semantics of specificity. *Linguistic Inquiry* 22.1: 1-25.

Ernst, Thomas (2002). *The syntax of adjuncts*. Cambridge: Cambridge University Press.

Erteschik-Shir, Nomi (2007). *Information structure: The syntax-discourse interface*. Oxford: Oxford University Press.

Erteschik-Shir, Nomi & Shalom Lappin (1979). Dominance and the functional explanation of island-phenomena. *Theoretical Linguistics* 6: 45-81.

Erteschik-Shir, Nomi & Shalom Lappin (1983). Under stress: a functional explanation of English sentence-stress. *Journal of Linguistics* 19: 419-453.

Falk, Yehuda N. (2001). *Lexical-Functional Grammar*. Stanford: CSLI Publications.

Falk, Yehuda N. (2006a). On the representation of case and agreement. In Butt, Miriam and Tracy H. King (eds.), *Proceedings of the LFG06 Conference* (pp. 184-201). Stanford: CSLI Publications.

Falk, Yehuda N. (2006b). *Subjects and Universal Grammar: An explanatory theory*. Cambridge: Cambridge University Press.

Falk, Yehuda N. (2007). Do we wanna (or hafta) have empty categories? In Butt, Miriam and Tracy H. King (eds.), *Proceedings of the LFG07 Conference* (pp. 184-197). Stanford: CSLI Publications.

Featherston, Samuel (2004). Bridge verbs and V2 verbs – the same thing in spades? *Sprachwissenschaft* 23: 181-209.

Forst, Martin (2006). COMP in (parallel) Grammar Writing. In Butt, Miriam and Tracy H. King (eds.) *Proceedings of the LFG06 Conference* (pp. 222-239). Stanford: CSLI Publications.

Frascarelli, Mara (2007). Types of topics in German and Italian. In Winkler, Susanne & Kerstin Schwabe (eds.), *On Information Structure, Meaning and Form* (pp. 87-116). Amsterdam, Philadelphia: John Benjamins.

Frey, Werner (2003). Syntactic conditions on adjunct classes. In Lang, Ewald, Claudia Maienborn and Catherine Fabricus-Hansen (eds.), *Modifying adjuncts* (pp. 163-209). Berlin, New York: Mouton de Gruyter.

Frey, Werner (2005). Pragmatic properties of certain German and English left peripheral constructions. *Linguistics* 43: 89-129.

Frey, Werner (2010). Ā-movement and conventional implicatures: About the grammatical encoding of emphasis in German. *Lingua* 120(6): 1416-1435.

Gazdik Anna (2011). *Multiple questions in French and Hungarian.* Doctoral dissertation, ELTE.

Gécseg Zsuzsanna (2001). A kontrasztív topik szintaxisáról és szemantikájáról. *Magyar Nyelv*, 2001. September, December.

Gécseg Zsuzsanna (2013). Adalékok az információs szerkezet és mondatszerkezet összefüggéseinek kérdéséhez. Magyar Nyelv 109(1): 32-49.

Gécseg Zsuzsanna & Kiefer Ferenc (2009). A new look at information structure in Hungarian. *Natural Language and Linguistic Theory* 27: 583-622.

Gerőcs Mátyás, Babarczy Anna & Surányi Balázs (2014). Exhaustivity in focus: experimental evidence from Hungarian. In Emonds, Joseph and Markéta Janebová (eds.), Language use and linguistic structure: Olomouc Modern Language Series 3 (pp. 181-195). Palacky University Press.

Gervain Judit (2002). *Linguistic methodology and microvariation in language: The case of operator-raising in Hungarian*. MA Thesis, University of Szeged.

Gervain Judit (2004). The syntax and semantics of the resumptive dependency in Hungarian focus-raising constructions. In Shaer, Benjamin, Werner Frey & Claudia Maienborn (eds.), *Proceedings of the Dislocated Elements Workshop* (ZAS Papers in Linguistics, 35:1) (pp. 111-131). Berlin: ZAS.

Gervain Judit (2009). Resumption in Focus(-Raising). *Lingua* 119: 687-707.

Gregory, Michelle L. & Laura A. Michaelis (2001). Topicalization and left-dislocation: A functional opposition revisited. *Journal of Pragmatics* 33: 1665-1706.

Grohmann, Kleanthes K. (2003). *Prolific domains: on the anti-locality of movement dependencies*. Amsterdam: John Benjamins.

Gundel, Jeanette K. (1974). *The role of topic and comment in linguistic theory*. Doctoral dissertation, University of Texas.

Gundel, Jeanette K. (1985). 'Shared knowledge' and topicality. *Journal of Pragmatics* 9(1): 83-107.

Gundel, Jeanette K. (1999). Topic, focus, and the grammar-pragmatics interface. In Alexander, Jim, Na-Rae Han & Michelle Minnick Fox (eds.), *Proceedings of the 23rd Annual Penn Linguistics Colloquium* (University of Pennsylvania Working Papers in Linguistics 6.1)(pp. 185-200). University of Pennsylvania.

Gundel, Jeanette K., Mamadou Bassene, Bryan Gordon, Linda Humnick & Amel Khalfaoui (2010). Testing predictions of the Givenness Hierarchy framework: A crosslinguistic investigation. *Journal of Pragmatics* 42: 1770–1785.

Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski (1993). Cognitive status and the form of referring expressions in discourse. *Language* 69:274-307.

Gutzmann, Daniel, Katharina Hartmann & Lisa Matthewson (2017). Verum focus is verum, not focus. Submitted, manuscript available online: https://www.danielgutzmann.com/work/verum-focus-is-verum-not-focus.

Gyuris Beáta (2003). *The semantics of contrastive topics in Hungarian*. Doctoral dissertation, ELTE.

Gyuris Beáta (2009). Quantificational contrastive topics with verum/falsum focus. *Lingua* 119: 625-649.

Haegeman, Liliane (1991). Parenthetical adverbials: The radical orphan approach. In Chiba, Shuji et al. (eds.), *Aspects of modern English linguistics: Papers presented to Masatomo Ukaji on his 60th birthday* (pp. 232–254). Tokyo: Kaitakusha.

Haegeman, Liliane (2003). Notes on Long Adverbial Fronting in English and the Left Periphery. *Linguistic Inquiry* 34: 640-649.

Hankamer, Jorge & Ivan A. Sag (1976). Deep and surface anaphora. *Linguistic Inquiry* 7:391-426.

Haug, Dag (2013). Partial Control and the Semantics of Anaphoric Control in LFG. In Butt, Miriam & Tracy H. King (eds.), *Proceedings of the LFG13 Conference* (pp. 274–294). Stanford: CSLI Publications.

Hedberg, Nancy (2000). The referential status of clefts. *Language* 76: 891-920.

Hegedűs Veronika (2013). *Non-Verbal Predicates and Predicate Movement in Hungarian.* PhD dissertation, Tilburg University. Utrecht: LOT Publications.

Higgins, Roger (1981). Proleptic objects and verbs of perception in Zacapoaxtla Nahuat. *Texas Linguistic Forum* 18: 69-88.

Hoekstra, Teun (1983). The distribution of sentential complements. In Bennis, Hans & Wus van Lessen Kloeke (eds.), *Linguistics in the Netherlands* (pp. 93-103). Dordrecht: Foris.

Hoji, Hajime (2005). Major object analysis of the so-called raising-to-object constructions in Japanese. Paper presented at New Horizons in the Grammar of Raising and Control Workshop. LSA Summer Institute, Harvard University.

Hong, Ki-Sun (1997). Yenge-wa kwuke-uy insang kwumwnun pikyo pwunsek (Subject-to-object raising constructions in English and Korean). *Language Research* 33: 409-434.

Hong, Soo-Min (2005). *"Exceptional" case-marking and resultative constructions*. Doctoral dissertation, University of Maryland.

Horváth Júlia (2010). 'Discourse features', syntactic displacement and the status of contrast. *Lingua* 120: 1346–1369.

Ince, Atakan (2006). Direct Complement Clauses as Object Control Structures in Turkish. In Bainbridge, Erin and Brian Agbayani (eds.), *Proceedings of the thirty-fourth Western Conference On Linguistics* (pp. 208-221). Department of Linguistics, California State University, Fresno.

Ingria, Robert J. P. (1981). *Sentential complementation in Modern Greek*. Doctoral dissertation, MIT.

Jackendoff, Ray (1972). *Semantic interpretation in generative grammar*. Cambridge: MIT Press.

Jánosi Adrienn (2013). *Long split focus constructions in Hungarian with a view on speaker variation*. Doctoral dissertation, CRISSP.

Kálmán László (1985). Word Order in Neutral Sentences. In Kenesei István (ed.), *Approaches to Hungarian* 1 (pp.13–23). Szeged: JATE.

Kálmán László (ed.) (2001). *Leíró magyar nyelvtan (Mondattan I.)*. Budapest: MTA-ELTE.

Kaplan, Ronald M. & Jürgen Wedekind (1993). Restriction and Correspondence-Based Translation. In *Proceedings of the 6th Meeting of the EACL*. European Chapter of the Association of Computational Linguistics, University of Utrecht.

Kenesei István (1992). Az alárendelt mondatok szerkezete. In Kiefer Ferenc (ed.), *Strukturális magyar nyelvtan II.: Mondattan* (pp. 529-714). Budapest: Akadémiai Kiadó.

Kenesei István (1994). Subordinate clauses. In Kiefer Ferenc and É. Kiss Katalin (eds.), *The syntactic structure of Hungarian* (pp. 141-165). San Diego: Academic Press.

Kenesei István (2006). Focus as identification. In Molnár Valéria and Susanne Winkler (eds.), *The architecture of focus* (Studies in generative grammar 82) (pp. 137-168). Berlin, New York: Mouton de Gruyter.

Kenesei István, Robert M. Vago & Fenyvesi Anna (1998). *Hungarian*. Taylor & Francis: Routledge, London.

Kibort, Anna (2007). Extending the applicability of Lexical Mapping Theory. In Butt, Miriam and Tracy H. King (eds.), *Proceedings of the LFG08 Conference* (pp. 250–270). Stanford: CSLI Publications.

Kibort, Anna (2008). On the syntax of ditransitive constructions. In Butt, Miriam and Tracy H. King (eds.), *Proceedings of the LFG08 Conference* (pp. 312–332). Stanford: CSLI Publications.

King, Tracy H. (1997). Focus domains and information structure. In Butt, Miriam and Tracy H. King (eds.), *Proceedings of the LFG97 Conference*. Stanford: CSLI Publications.

Kiparsky, Paul & Carl Kiparsky (1970). Fact. In Bierwisch, Manfred and Karl Heidolph (eds.), *Progress in Linguistics* (pp. 143-173). The Hague: Mouton.

Kotzoglou, George & Dimitra Papangeli (2007). Not really ECM, not exactly control: The "quasi-ECM" construction in Greek. In Davies, William D. and Stanley Dubinsky (eds.), *New horizons in the analysis of control and raising* (pp. 111–131). Dordrecht: Springer.

Krifka, Manfred (2008). Basic notions of information structure. *Acta Linguistica Hungarica* 55: 243-276.

Krifka, Manfred & Renate Musan (2012). Information structure: Overview and linguistic issues. In Krifka, Manfred and Renate Musan (eds.), *The expression of Information structure* (pp. 1-43). Berlin: De Gruyter Mouton.

Kuno, Susumo & Ken-ichi Takami (1993). *Grammar and Discourse Principles: Functional Syntax and GB Theory*. Chicago: Chicago University Press.

Laczkó Tibor (1995). *The syntax of Hungarian noun phrases – A Lexical-Functional Approach*. (Metalinguistica 2). Frankfurt am Main: Peter Lang.

Laczkó Tibor (2012). On the (un)bearable lightness of being an LFG style copula in Hungarian. In Butt, Miriam and Tracy Holloway King (eds.), *Proceedings of the LFG12 Conference* (pp. 341-361). Stanford: CSLI Publications.

Laczkó Tibor. (2014a). The essentials of an LFG-analysis of Hungarian finite sentences. In Butt, Miriam and Tracy H. King (eds.), *Proceedings of the LFG14 Conference* (pp. 325-345). Stanford: CSLI Publications.

Laczkó Tibor. (2014b). An LFG analysis of preverbs in Hungarian. In Butt, Miriam and Tracy H. King (eds.), *Proceedings of the LFG14 Conference* (pp. 346-366). Stanford: CSLI Publications.

Laczkó Tibor (2015). On operators in the preverbal domain of Hungarian finite sentences. In Butt, Miriam and Tracy H. King (eds.) *Proceedings of the LFG15 Conference* (pp. 187-207). Stanford: CSLI Publications.

Lambrecht, Knud (1994). *Information Structure and Sentence Form – Topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.

Lambrecht, Knud (2001). Dislocation. In Haspelmath, Martin et al. (eds.), *Language typology and language universals* (pp. 1050-1078). Berlin/New York: Walter de Gruyter.

Landau, Idan (2011). Predication vs. aboutness in copy raising. *Natural Language and Linguistic Theory* 29: 779-813.

Landau, Idan (2013). *Control in generative grammar – A research companion*. Cambridge: Cambridge University Press.

Landau, Idan (2015). *A Two - Tiered Theory of Control*. Cambridge, MA: MIT Press.

Landau, Idan (2016). Agreement at PF: an Argument from Partial Control. *Syntax* 19(1): 79-109.

Lang, Ewald, Claudia Maienborn & Catherine Fabricus-Hansen (eds.) (2003). Modifying adjuncts. Berlin, New York: Mouton de Gruyter.

Levinson, Stephen C. & Nicholas Evans (2010). Time for a sea-change in linguistics: Response to comments on 'The Myth of Language Universals'. *Lingua* 120: 2733-2758.

Lødrup, Helge (2012). In search of a nominal COMP. In Butt, Miriam and Tracy H. King (eds.), *Proceedings of the LFG12 Conference* (pp. 383-404). Stanford: CSLI Publications.

Lipták Anikó (1998). A magyar fókuszemelések egy minimalista elemzése. In Büky, László and Maleczki Márta (eds.), *A mai magyar nyelv leírásának újabb módszerei 3* (pp. 93-116). Szeged: University of Szeged.

Lipták Anikó (2011). The structure of the topic field in Hungarian. In Beninca, Paola and Nicola Munaro (eds.), *Mapping the left periphery* (pp. 163-198). Oxford: Oxford University Press.

Lipták Anikó (2012). Correlative topicalization. *Acta Linguistica Hungarica* 59(3): 245-302.

López, Luis (2009). *A Derivational Syntax for Information Structure.* Oxford: Oxford University Press.

Maienborn, Claudia (2001). On the Position and Interpretation of Locative Modifiers. *Natural Language Semantics* 9(2): 191-240.

Manzini, Rita M. (2014). Grammatical categories: Strong and weak pronouns in Romance. *Lingua* 150: 171-201.

Martins, Ana Maria and Jairo Nunes (2010). Apparent Hyper-raising in Brazilian Portuguese: Agreement with Topics across a finite CP. In Phoevos, Panagiotidis (ed.), *The Complementiser Phase: Subjects and Wh-dependencies* (pp. 143-163). Oxford: Oxford University Press.

Massam, Diane (1985). *Case Theory and the Projection Principle*. Doctoral Dissertation, MIT.

Mikkelsen, Line (2011). Copular clauses. In Maienborn, Claudia, Klaus von Heusinger and Paul Portner (eds.), *Semantics: An International Handbook of Natural Language Meaning – volume 2* (pp. 1805-1829). Berlin: Mouton de Gruyter.

Molnár Valéria (2002). Contrast – from a contrastive perspective. *Language and Computers* 39: 147-162.

Molnár Valéria (2006). On different kinds of contrast. In Molnár Valéria and Susanne Winkler (eds.), *The architecture of focus* (Studies in generative grammar 82) (pp. 197-233). Berlin, New York: Mouton de Gruyter.

Molnár Valéria & Susanne Winkler (2010). Edges and gaps: Contrast at the interfaces. *Lingua* 120: 1392–1415.

Moro, Andrea (1997). *The raising of predicates*. Cambridge: Cambridge University Press.

Mycock, Louise (2006). *The Typology of Constituent Questions: a Lexical-Functional Grammar analysis of 'wh'-questions*. Doctoral dissertation, University of Manchester.

Mycock, Louise (2013). Discourse Functions of Question Words. In Butt, Miriam and Tracy H. King (eds.), *Proceedings of the LFG13 Conference* (pp. 419–439). Stanford: CSLI Publications.

Mycock, Louise & John Lowe (2013). The Prosodic marking of discourse functions. In Butt, Miriam and Tracy H. King (eds.), *Proceedings of the LFG13 conference* (pp. 440–460). Stanford: CSLI Publications.

Needham, Stephanie & Ida Toivonen (2011). Derived arguments. In Butt, Miriam and Tracy H. King (eds.), *Proceedings of the LFG11 Conference* (pp. 401-421). Standford: CSLI Publications.

Neeleman, Ad & Hans van de Koot (2008). Dutch scrambling and the nature of discourse templates. *Journal of Comparative Germanic Linguistics* 11: 137–189.

Neeleman, Ad, Elena Titov, Hans van de Koot & Reiko Vermeulen (2009). A syntactic typology of topic, focus and contrast. In van Craenenbroeck, Jeroen (ed.), *Alternatives to Cartography* (pp. 15-52). Berlin: Mouton de Gruyter.

Onea, Edgar (2009). Exhaustiveness of Hungarian Focus. Evidence from Hungarian and German. In Riester, Arndt and Edgar Onea (eds.), *Focus at the Syntax-Semantics Interface* (Working Papers of the SFB 732: Incremental Specification in Context), volume 3 (pp. 53-68). University of Stuttgart.

Özsoy, A. Sumru (2001). On small clauses, other bare verbal complements and feature checking in Turkish. In Erguvanlı, Eser and Eser E. Taylan (eds.), *The verb in Turkish* (pp. 213-237). Amsterdam: John Benjamins Publishing Company.

Philippaki-Warburton, Irene (1987). The theory of empty categories and the pro-drop parameter in Modern Greek. *Journal of Linguistics* 23(2): 289-318.

Pesetsky, David (1987). Binding Problems with Experiencer Verbs. *Linguistic Inquiry* 18(1): 126-140.

Pesetsky, David (1995). *Zero syntax: Experiencers and cascades*. Cambridge, Massachusetts/London, England: The MIT Press.

Pintér Lilla (2016). A magyar szerkezeti fókusz kimerítő értelmezésének kísérletes vizsgálata. In Balázs József, Bojtos Anita, Paár Tamás, Tompa Zsófia, Turi Gergő and Vadász Noémi (eds.), Studia Varia (pp. 191–212). Budapest: Pázmány Péter Katolikus Egyetem Bölcsészet- és Társadalomtudományi Kar.

Pitteroff, Marcel, Artemis Alexiadou & Silke Fischer (2015). *On partial control in German*. Generative Grammatik des Südens.

Prince, Ellen F. (1981a). Towards a taxonomy of given-new information. In Cole, Peter (ed.), *Radical Pragmatics*. New York: Academic Press.

Prince, Ellen F. (1981b). Topicalization, focus-movement, and Yiddish-movement: A pragmatic differentiation. In Alford, Danny et al. (eds.), *Proceedings of the Seventh Annual Meeting of the Berkeley Linguistics Society* (pp. 249–264). Berkeley: Berkeley Linguistics Society.

Prince, Ellen F. (1998). On the limits of syntax, with reference to left-dislocation and topicalization. In Cullicover, Peter and Louise McNally (eds.), *Syntax and Semantics 29 - The limits of syntax*. New York: Academic Press.

Prince, Ellen F. (1999). How not to mark topics: 'Topicalization' in English and Yiddish. In *Texas Linguistics Forum* (chapter 8). Austin: University of Texas.

Prince, Alan & Paul Smolensky (1993). *Optimality theory: Constraint interaction in generative grammar (RuCCS Technical Report #2).* Rutgers University Center for Cognitive Science.

Postal, Paul M. & Geoffrey K. Pullum (1988). Expletive noun phrases in subcategorized positions. *Linguistic Inquiry* 19: 635-670.

Rákosi György (2006a). On the need for a more refined approach to the argument-adjunct distinction: the case of dative experiencers in Hungarian. In Butt, Miriam and Tracy H. King (eds.), *Proceedings of the LFG06 Conference* (pp. 416-436). Stanford: CSLI Publications.

Rákosi György (2006b). *Dative experiencer predicates in Hungarian*. Doctoral dissertation, University of Utrecht.

Rákosi György (2012). Non-core participant PPs are adjuncts. In Butt, Miriam and Tracy H. King (eds.), *Proceedings of the LFG12 Conference* (pp. 524– 543). Stanford: CSLI Publications.

Rákosi György & Laczkó Tibor (2005). Verbal category and nominal function – Evidence from Hungarian Subject Clauses. In Butt, Miriam and Tracy H. King (eds.), *Proceedings of the LFG05 Conference* (pp. 353-370). Stanford: CSLI Publications.

Reinhart, Tanya (1981). Pragmatics and linguistics: an analysis of sentence topics. *Philosophica* 27:53-94.

Reinhart, Tanya (2001). Experiencing derivations. In Hastings, Rachel, Brendan Jackson and Zsófia Zvolenszky. (eds.), *Proceedings of Semantics and Linguistic Theory 11*. Ithaca, NY: CLC Publications.

Reinhart, Tanya (2002). The theta system – an overview. *Theoretical Linguistics* 28:229-290.

Repp, Sophie (2010). Defining 'contrast' as an information-structural notion in grammar. *Lingua* 120: 1333–1345.

Roberts, Craige (1996). Information structure in discourse: towards an integrated formal theory of pragmatics. In Yoon, Jae-Hak, Andreas Kathol and Craige Roberts (eds.), *Papers in semantics* (Working Papers in Linguistics 49). The Ohio State University.

Rodrigues, Cilene (2004). Impoverished morphology and A-movement out of Case domains. Doctoral dissertation, University of Maryland.

Rodman, Robert (1974). On left dislocation. *Papers in Linguistics* 7: 437-466.

Rooth, Mats E. (1985). Association with focus. Doctoral dissertation, University of Massachusetts, Amherst, MA.

Ross, John R. (1967). *Constraints on variables in syntax*. Doctoral dissertation, Massachusetts Institute of Technology.

Rothstein, Susan D. (1995). Pleonastics and the interpretation of pronouns. *Linguistic Inquiry* 26: 499-529.

Salzmann, Martin (2006). *Resumptive Prolepsis: A Study in indirect A'-dependencies*. Utrecht: LOT (LOT Dissertation Series 136).

Salzmann, Martin (to appear). Prolepsis. In Everaert, Martin & Henk van Riemsdijk (eds.), *The Companion to Syntax, 2nd edition*. Wiley-Blackwell. Available online: http://ling.auf.net/lingbuzz/002001, accessed: 30 October 2017.

Sells, Peter (1984). *Syntax and Semantics of Resumptive Pronouns*. Doctoral dissertation, University of Massachusetts.

Sells, Peter (2006). Using subsumption rather than equality in functional control. In Butt, Miriam and Tracy H. King, (eds.), *Proceedings of the LFG06 Conference* (pp. 455-473). Stanford: CSLI Publications.

Shaer, Benjamin (2004). On the syntactic status of certain fronted adverbials in English. In Nowak, Pawel, Corey Yoquelet & David Mortensen (eds.), *Proceedings of the Berkeley Linguistics Society 29* (pp. 379–390). Berkeley, CA: BLS.

Shaer, Benjamin (2009). German and English left-peripheral elements and the orphan analysis of non-integration. In Shaer, Benjamin, Philippa Cook, Werner Frey and Claudia Maienborn (eds.), *Dislocated elements in discourse – Syntactic, semantic and pragmatic perspectives* (pp. 366-397). New York: Routledge.

Shaer, Benjamin & Werner Frey (2004). Integrated and non-integrated left-peripheral elements in German and English. In Shaer, Benjamin, Werner Frey & Claudia Maienborn (eds.), *Proceedings of the Workshop on dislocated elements* (pp. 465-502). Berlin: Zentrum für Allgemeine Sprachwissenschaft.

Simpson, Jane H. (1983). *Aspects of Warlpiri Morphology and Syntax*. Doctoral dissertation, MIT.

Snider, Neal & Annie Zaenen (2006). Animacy and syntactic structure: Fronted NPs in English. In Butt, Miriam, Mary Dalrymple and Tracy H. King (eds.), *Intelligent linguistic architectures: Variations on themes by Ronald M. Kaplan*. Stanford: CSLI Publications.

Spyropoulos, Vassilios (2007). Finiteness and control in Greek. In Davies, William D. and Stanley Dubinsky (eds.), *New horizons in the analysis of control and raising* (pp. 159-183). Dordrecht: Springer.

Subirats-Rüggeberg, Carlos (1990). The relation between infinitival and sentential complements in traditional grammar. *Lingvisticæ Investigationes* 14: 81-93.

Stowell, Tim (1981). Origins of phrase structure. Doctoral dissertation, MIT.

Stowell, Tim (1982). The tense of infinitives. *Linguistic Inquiry* 13: 561–570.

Surányi Balázs (2006a). Mechanisms of wh-saturation and interpretation in multiple wh-movement. In Cheng, Lisa and Norbert Corver (eds.), *Wh-movement: Moving On* (pp. 289-318). Cambridge, Massachusetts: MIT Press.

Surányi Balázs (2006b). Scrambling in Hungarian. *Acta Linguistica Hungarica* 53(4): 393-432.

Szabolcsi Anna (2009). Overt Nominative Subjects in Infinitival Complements: Data, Diagnostics, and Preliminary Analyses. In Irwin, Patricia and Violeta Vasquéz Rojas Maldonado (eds.), *NYU Working Papers in Linguistics, References 277 Vol. 2: Papers in Syntax*. New York University.

Szűcs Péter (2010). *On Operator Raising in Hungarian*. MA Thesis, University of Debrecen.

Titov, Elena (2013). Do contrastive topics exist? *Journal of Linguistics* 49(2): 413-454.

Toivonen, Ida (2013). English benefactive NPs. In Butt, Miriam and Tracy H. King (eds.), *Proceedings of the LFG13 Conference* (pp. 503–523). Stanford: CSLI Publications.

Tortora, Christina M. (1997). *The syntax and semantics of the weak locative*. Doctoral dissertation, University of Delaware.

Tóth Ildikó (2000). *Inflected Infinitives in Hungarian*. Tilburg: TILDIL Dissertation Series.

Vallduví, Enrique (1992). *The informational component*. New York: Garland.

Vat, Jan (1981). Left dislocation, connectedness, and reconstruction. *Groninger Arbeiten zur Germanistischen Linguistik* 20: 80-103.

von Heusinger, Klaus (2011). Specificity. In von Heusinger, Klaus, Claudia Maienborn and Paul Portner (eds.), Semantics: An International Handbook of Natural Language Meaning (Volume 2) (pp. 1025-1058). Berlin/Boston: De Gruyter Mouton.

Ward, Gregory (1988). *The semantics and pragmatics of preposing*. New York/London: Garland Publishing.

Wedgewood, Daniel (2009). Variation in Focus. In Riester, Arndt and Edgar Onea (eds.), *Focus at the Syntax-Semantics Interface* (Working Papers of the SFB 732: Incremental Specification in Context), volume 3 (pp. 101-120). University of Stuttgart.

Vilkuna, Maria (1995). Discourse configurationality in Finnish. In É. Kiss Katalin (ed.), *Discourse Configurational Languages* (pp. 244–268). Oxford, New York: Oxford University Press.

Yoon, James H-S. (2007). Raising of Major Arguments in Korean and Japanese. In Davies, William D. & Stanley Dubinsky (eds.), *New horizons in the analysis of control and raising* (pp. 71-107). Dordrecht: Springer.

Zaenen, Annie & Elisabet Engdahl (1994). Descriptive and theoretical syntax in the lexicon. In Atkins, Beryl T. S. and Antonio Zampolli (eds.), *Computational Approaches to the Lexicon: Automating the Lexicon II* (pp. 181-212). Oxford: Oxford University Press.

Zec, Draga (1987). On Obligatory Control in Clausal Complements. In Masayo, Iida, Steven Wechsler and Draga Zec (eds.), *Working Papers in Grammatical Theory and Discourse Structure* (pp.139-168). Stanford, CA: CSLI Publications.

Zeller, Jochen (2006). Raising out of finite CP in Nguni: the case of fanele. *Southern African Linguistics and Applied Language Studies* 24(3): 255-275.

Zimmermann, Malte (2008). Contrastive focus and emphasis. *Acta Linguistica Hungarica*, 55: 347-360.