

**THE DISCOURSE STRUCTURING POTENTIAL OF  
INDEFINITE NOUN PHRASES.**

**SPECIAL MARKERS IN ENGLISH, GERMAN AND  
ROMANIAN**

Von der Philosophisch-Historischen Fakultät der Universität Stuttgart  
zur Erlangung der Würde eines Doktors der Philosophie (Dr. phil.)  
genehmigte Abhandlung

vorgelegt von

**Sofiana-Iulia Chiriacescu**  
**aus Braşov, Rumänien**

Hauptberichter: Prof. Dr. Klaus von Heusinger

Mitberichterin: Assistant Prof. Elsi Kaiser

Tag der mündlichen Prüfung: 18.07.2011

Institut für Linguistik / Germanistik der Universität Stuttgart



## Table of Contents

<b>ENGLISH ABSTRACT .....</b>	<b>i</b>
<b>GERMAN ABSTRACT .....</b>	<b>iii</b>
<b>1. Introduction .....</b>	<b>1</b>
1.1 Introducing the phenomena .....	5
1.2 Goals of the dissertation and proposed analysis .....	8
1.3 Limitations of the investigation .....	10
1.4 Outline of the Dissertation .....	11
<b>2. Past work on reference.....</b>	<b>14</b>
2.1 Semantic-pragmatic accounts on (in)definiteness .....	16
2.2 Accessibility as activation .....	22
2.2.1 The Accessibility Hierarchy .....	23
2.2.2 Factors contributing to the accessibility of referents .....	27
2.3 Accessibility as topicality .....	36
2.3.1 Givón's distance model .....	36
2.3.2 Centering Theory.....	39
2.4 Accessibility as Expectancy .....	44
2.4.1 Expectancy of subsequent mention and likelihood of pronominalization .....	44
2.4.2 The discourse structuring potential .....	48
2.4.3 The discourse analysis method.....	50
2.5 Summary.....	58
<b>3. The discourse contribution of English indefinite this .....</b>	<b>59</b>
3.1 Forms and uses of demonstratives .....	61
3.1.1 Definite uses of demonstrative this .....	62
3.1.2 The indefinite use of demonstrative this .....	65
3.2 Referential and discourse-pragmatic properties of indefinite this .....	69
3.2.1 Referential properties of indefinite this.....	69
3.2.2 Discourse-pragmatic properties of indefinite this .....	74
3.3 An experimental investigation on the discourse effects of indefinite this.....	81
3.3.1 The experimental setup .....	83

3.3.2	Procedure and data analysis .....	85
3.3.3	Results and discussions .....	85
3.4	Demonstratives and focus of attention .....	94
3.5	Summary .....	101
<b>4.</b>	<b>German indefinite so'n.....</b>	<b>102</b>
4.1	Demonstrative uses of so'n.....	104
4.1.1	Deictic and anaphoric uses of so .....	106
4.1.2	'Out-of-the-blue' uses of so'n .....	108
4.1.3	So'n as a marker of referential structure .....	112
4.1.4	Summary .....	115
4.2	Referential properties of indefinite so'n .....	116
4.2.1	Different types of readings of indefinite so'n .....	116
4.2.2	Experimental evidence on the referential readings of indefinite so'n.....	121
4.2.3	Summary .....	127
4.3	Discourse properties of indefinite so'n.....	127
4.3.1	Indefinite so'n and immediacy of reference.....	128
4.3.2	The discourse structuring potential of indefinite so'n.....	130
4.4	Summary .....	141
<b>5.</b>	<b>Pe-marking of indefinite noun phrases in Romanian .....</b>	<b>143</b>
5.1	Criteria for differential object marking in Romanian .....	145
5.2	The distribution of pe-marking with definite expressions .....	152
5.2.1	DOM with definite modified noun phrases.....	152
5.2.2	DOM with definite unmodified noun phrases.....	154
5.3	Pe-marking with indefinite noun phrases .....	160
5.3.1	Specificity.....	161
5.3.2	Topicality and pe-marking with indefinite noun phrases.....	170
5.3.3	Summary .....	174
5.4	The discourse structuring potential of pe-marking in Romanian .....	175
5.4.1	The sentence-continuation experiment on pe-marking .....	176
5.5	Summary .....	188
<b>6.</b>	<b>Referential properties and discourse structuring.....</b>	<b>189</b>
6.1	Referential properties.....	191
6.2	The discourse structuring potential.....	193
6.3	The relation between referential properties and discourse structuring.....	201

6.3.1 Grammaticalization of the indefinite article cross-linguistically .....	206
6.3.2 DOM in Romanian as a marker of discourse prominence .....	208
6.3.3 A candidate for discourse prominence: the proximal demonstrative .....	210
6.4 Summary .....	213
<b>7. Conclusions .....</b>	<b>214</b>
<b>BIBLIOGRAPHY.....</b>	<b>219</b>

## ACKNOWLEDGEMENTS

I would like to thank Klaus von Heusinger, who fostered my decision to become a linguist throughout the years of my studies at the University of Stuttgart. I am grateful for the fruitful and inspiring discussions we had, as they have continuously given me an impulse to write this dissertation.

I am dearly indebted to Elsi Kaiser, who patiently assisted me and provided insightful advice especially with the experimental part of this dissertation. During my stay at the Linguistics Department at the University of Southern California, I was particularly impressed by her gift to inspire her students (including myself) and increase their appetite and knowledge in psycholinguistics.

I would also like to thank my colleagues at the Department of Foreign Languages at the Transylvania University of Brasov, Romania for introducing me into the world of teaching, especially to Paula Armasar, Andrei Barna, Andrei Bodi, Mihaela Gheorghe, Carmen Puchianu, Daria Stiuca and Daniela Turcu.

Thank you to Prof. dr. Nicolae Saramandu and the collective from the Institute of Linguistics “Iorgu Iordan-Alexandru Rosetti” of the Romanian Academy of Sciences in Bucharest for fruitful advice and suggestions during the conferences they organized.

Many thanks to my colleagues and friends at the Department of Linguistics at the University of Stuttgart, namely Regine Brandtner, Jin Cui, Annika Deichsel, Cornelia Ebert, Ljudmila Geist, Dolgor Guntsetseg, Daniel Hole, Elena Karagjsova, Udo Klein, Elisabeth Löbel, Edgar Onea and Johannes Wespel. I am grateful to have been given the chance to work in such a friendly, warm and international environment.

I am thankful for the financial support of the collaborative research center SFB 732 “Incremental Specification in Context” (Project C2, “Case and referential context” and Project C4, “Diskurspartikeln. Inkrementelle Interpretation im Kontext”) funded by the German Science Foundation (DFG).

My family and friends provided advice and encouragement at all hours and in all matters from lesser or greater distance. Thank you for your unconditioned support.

*To my family*



## ENGLISH ABSTRACT

### THE DISCOURSE STRUCTURING

#### THE DISCOURSE STRUCTURING POTENTIAL OF INDEFINITE NOUN PHRASES. SPECIAL MARKERS IN ENGLISH, GERMAN AND ROMANIAN

JUNE 2011

SOFIANA-IULIA CHIRIACESCU

Dissertation, UNIVERSITY OF STUTTGART

Professor Klaus von Heusinger, Chair

In this dissertation I investigate different types of indefinite noun phrases to deepen our understanding of the nature of the form–function mapping during the process of referring. More concretely, I explore the extent to which formal linguistic elements (i.e. different markers of indefiniteness) affect discourse participants’ predictions about the subsequent discourse. The centerpieces of this dissertation are indefinite noun phrases marked by indefinite *this* in English, indefinite *so’n* in German and *pe* in Romanian. I argue that these indefinite noun phrases share several common characteristics at the semantic and discourse-pragmatic level, which distinguish them from noun phrases headed by the simple indefinite article *a(n)*. First, at the semantic level, the three types of indefinites show a strong and robust affinity for particular referential properties (i.e. fixed reference, wide scope, epistemic specificity). Second, I argue that any account of indefinite noun phrases that investigates only their sentence-level properties fails to capture their discourse-pragmatic contribution. By analysing larger discourse units (i.e. containing more than two adjacent sentences) in which such indefinite noun phrases occur, I will show that compared to their simple indefinite correlates, referents associated with these specially marked indefinites in English, German and Romanian are more prone to: (i) be frequently

mentioned in the ensuing discourse and to (ii) shift the discourse topic. At first sight, these two discourse effects call for an analysis in terms of accessibility or activation. Most interestingly, however, the above-mentioned effects appear with a latency of one or two subsequent sentences after the indefinites were mentioned for the first time. Furthermore, another argument that cannot be easily accommodated within existing models based on accessibility or activation is the observation that referents introduced in this way show low rates of subsequent pronominalization. I propose that indefinite noun phrases function as discourse structuring devices. Their main function is to establish coherence at a more global level, by pre-activating particular referents. The pre-activating function is a signal to the hearer that the referent introduced in this way will be mentioned next and will eventually become a topical referent.

These results can be best accommodated in an expectancy-based or forward-looking model of language processing, which assumes that language is predictive as speakers use different signals to indicate to their hearers what to expect next. However, compared to previous accounts on language expectation, in this dissertation I distinguish between two types of expectations, which are triggered by different factors at different levels of representation. The first type of expectation, which is generally discussed in the literature on pronoun resolution in terms of accessibility, can be derived from structural factors such as a particular syntactic position and a certain argument structure. Based on such factors, hearers make probabilistic expectations about those referents that are more probable to be mentioned next and more prone to be pronominalized. The second type of expectation, which is important for the present analysis, comes about by discourse structuring devices such as the three types of indefinite noun phrases discussed in this dissertation. Such linguistic devices target larger discourse units in that they indicate which referent will be mentioned over larger discourse segments. The main suggestion resulting from the presented evidence is that expectation operates in parallel at different levels of representation and that languages use different linguistic means to mark it.

## GERMAN ABSTRACT

Eine der wichtigsten Fragen der Sprachtheorie beschäftigt sich mit Referenz-Tracking. Darunter versteht man die Art und Weise, wie Personen, Dinge, Ereignisse und Sachverhalte (allgemein Referenten genannt) in den Diskurs eingeführt werden und wieder aufgenommen werden. Die bisherige (psycho-) linguistische Forschung zeigt, dass unterschiedliche Faktoren die Häufigkeit der Wiederaufnahme eines Referenten, sowie die Art des dafür verwendeten referenziellen Ausdrucks beeinflussen. Die vorliegende Dissertation untersucht die semantischen und diskurs-pragmatischen Merkmale und Diskursfunktionen von Indefinita. Das Hauptanliegen dieser Dissertation ist, den Zusammenhang zwischen den folgenden zwei Faktoren zu untersuchen: (i) die referenzielle Form (markierte indefinite Nominalphrasen vs. unmarkierte indefinite Nominalphrasen), und (ii) den diskursstrukturierenden Potential indefiniter Nominalphrasen. Durch eine sprachvergleichende Analyse mit Schwerpunkt auf Englisch, Deutsch und Rumänisch, wird die Analyse der Funktionen unterschiedlicher indefiniten Nominalphrasen empirisch gestützt und weiter entwickelt. Damit wird die Arbeit zu einem besseren Verständnis von Textstruktur, den kognitiven Grundlagen von Textproduktion und Textverständnis, sowie den allgemeinen Prinzipien der Kommunikation zwischen Gesprächsteilnehmern beitragen.

Unter Indefinita werden in dieser Arbeit indefinite nominale Ausdrücke verstanden, die über einen deskriptiven Teil und einem bestimmten Determinierer verfügen- insofern unterschiedliche Sprachen einen entwickelt haben. Einige Beispiele aus dem Deutschen wären *eine Frau*, *einige Frauen*, Partitive, usw. Die Frage nach dem Alleinstellungsmerkmal von indefiniten Ausdrücken wurde in der semantischen Literatur generell durch die Annahme beantwortet, dass in ihre Bedeutungsanalyse *existentielle Quantifikation* tritt (Fodor and Sag 1982, Farkas 1994, 2002, Kamp and Bende-Farkas 2006, eingereicht). Anders ausgedrückt, sollte ein Sprecher eine indefinite Nominalphrase benutzen, so macht er mindestens eine existentielle Behauptung. Ein zweites Hauptmerkmal von indefiniten Nominalphrasen, welches die Informationsstruktur betrifft, ist ihre Eigenschaft, neue, nicht-familiäre Referenten im Diskurs einzufügen (Kamp 1981, Heim 1982).

Linguistische und psycholinguistische Studien (Givón 1983, Ariel 1988, Gordon et al. 1993, Gundel et al. 1993, Arnold 2008) haben gezeigt, dass verschiedene Faktoren, wie zum Beispiel prominente grammatische oder thematische Rollen, oder bestimmte Kohärenzrelationen, einen großen Einfluss auf die Frequenz der Wiederaufnahme von Referenten und die Wahrscheinlichkeit der Pronominalisierung dieser Referenten haben. Das heißt, dass Referenten, die in Subjektposition erwähnt wurden, oder Referenten, die als Stimulus in einem transitiven Ereignis auftreten, mit einer viel höheren Wahrscheinlichkeit wieder aufgegriffen werden, als solche Referenten, die in Objektposition, oder als Experiencer eingeführt wurden. Ähnliche Ergebnisse werden in einer Reihe von sprachübergreifenden Studien erzielt, die in dieser Dissertation präsentiert werden, die den Einfluss verschiedener Typen von Referenzausdrücken auf den nachfolgenden Diskurs getestet haben. Die Ergebnisse dieser Studien zeigen, dass das Verhalten indefiniter Nominalphrasen (wie z.B. *so'n* im Deutschen, indefinit *this* im Englischen oder *pe-Markierung* im Rumänischen) am besten in Bezug auf ihren Beitrag zum Diskurs zu erklären ist. Die Resultate der Studien und Experimente zeigen, dass bestimmte Typen von indefiniten Nominalphrasen dem Hörer signalisieren, dass die eingeführten Diskursreferenten im nachfolgenden Diskurs mehrmals aufgegriffen werden und das Potential aufweisen, das Topik zu verändern. Diese Untersuchungen werden zeigen, dass indefinite Nominalphrasen nicht nur Diskursreferenten einführen (wie das üblicherweise in der Literatur angenommen wird), sondern dass sie auch ein Potential aufweisen, das die referenzielle Textstruktur des kommenden Diskurses strukturiert.

Der indefinite Demonstrativ *this* im Englischen, der Determinator *so'n* im Deutschen (in informellen Varianten des Deutschen), oder der differentielle Objektmarker *pe* im Rumänischen sind gute Beispiele dafür, um die doppelte Funktion indefiniter Nominalphrasen zu illustrieren. Zum Beispiel lässt sich die Distribution des indefiniten *this* im Englischen damit erklären, dass die so eingeführten Diskursreferenten häufiger wieder aufgegriffen werden (vgl. (1a) mit (1b)). Im Vergleich dazu, werden Diskursreferenten, die durch den einfachen indefiniten Artikel (*ein(e)*) eingeführt werden, seltener wieder aufgenommen.

- (1) (a) Becky wrote some thank-you notes using {a/ #this} purple pen; then she mailed the notes to her friends.

- (b) Becky wrote some thank-you notes using {a/ this} purple pen, which suddenly exploded, spilling purple ink all over Becky's clothes and furniture!

Diese diskursstrukturierende Eigenschaft indefiniter Nominalphrasen wird *discourse structuring potential* genannt. Auf Grund der Überlegungen im 2. Kapitel werden drei Diskursparameter definiert, die in dem empirischen Teil der Arbeit zur Anwendung kommen, nämlich: (i) die *referenzielle Persistenz*, (ii) das *Topikveränderungspotential* und (iii) die *Aktivierung* des Diskursreferenten. Unter *referenzieller Persistenz* (*referential persistence*, vgl. Givón 1983) wird die Diskurseigenschaft eines Ausdrucks aufgefasst, die mit der Häufigkeit seiner anaphorischen Wiederaufnahmen im folgenden Text zu tun hat. Mit dem *Topikveränderungspotential* (*topic change potential*, Chiriacescu & von Heusinger 2010) wird hingegen das Potential erfasst, im folgenden Text Topik zu bleiben (vgl. Givóns 1983 *topic continuity* Methode), oder Topik zu werden. Schließlich wird mit *Aktivierung* die Eigenschaft benannt, die einen Diskursreferenten bezüglich seines deskriptiven Gehalts salient oder zugänglich macht. Die Aktivierung oder Zugänglichkeit eines Referenten zeigt sich unter anderem in dem Typ des anaphorischen Ausdrucks, mit dem der Diskursreferent zum ersten Mal wieder aufgegriffen wird (vgl. Gundel et al. 1993, Arnold 1998, Chiriacescu & von Heusinger 2010). Aus den theoretischen Überlegungen des zweiten Kapitels resultiert die Erwartung, dass natürliche Sprachen eine Differenzierung zwischen diskursprominenten Referenten und weniger prominenten Referenten im morpho-syntaktischen Aufbau der Nominalphrase reflektieren.

In den Kapiteln 3, 4 und 5 werden indefinite Nominalphrasen analysiert, die spezielle morphologische Mittel zur Markierung von indefiniten Nominalphrasen aufzeigen, die zugleich die Diskursprominenz ihrer Referenten reflektieren. Um die theoretischen Beobachtungen hinsichtlich des Diskursverhaltens dieser indefiniten Nominalphrasen zu unterstreichen, wurden drei experimentelle Studien durchgeführt in denen die Diskurseigenschaften der markierten Indefinita im Vergleich zu solchen indefiniten Nominalphrasen, denen der einfache indefinite Artikel vorangestellt werden, getestet. Kapitel 6 fasst zunächst die Ergebnisse der Experimente zusammen. Sprecher nutzen verschiedene linguistische Hinweise, um Erwartungen über den nachfolgenden Diskurs zu generieren. Die vorliegenden Studien unterstreichen die Tatsache, dass sich die von Indefinita eingeführten Referenten im Hinblick auf ihr Potential den nachfolgenden

Diskurs zu strukturieren, unterscheiden lassen. Dieses unterschiedliche Potential den Diskurs zu strukturieren, ist eine Widerspiegelung der Relation, die zwischen der Art der indefiniten Nominalphrase und der Funktion, die ihr Referent im nachfolgenden Diskurs spielen wird, besteht.

Schlussendlich werden die Auswirkungen der gewonnenen Erkenntnisse über die Distribution und Diskursfunktion von indefiniten Nominalphrasen im Hinblick auf die Notwendigkeit von Sprachen solche Unterscheidungen zu machen, problematisiert. Im Ganzen zeigt diese Doktorarbeit, dass verschiedene Variationen in der Distribution von indefiniten Nominalphrasen im theoretischen Rahmen der Schnittstelle von Semantik und Diskurspragmatik behandelt werden können.

# 1. Introduction

This dissertation is about particular types of indefinite noun phrases that give structure to the discourse in terms of predictive expectations. Indefinite noun phrases are nominal expressions with a descriptive part and a special determiner form in those languages that have developed one. Typical examples of English indefinite noun phrases include, but are not limited to: *a man*, *this man*, *several men*, *a certain man*, *some men*, cardinals, the zero of bare plurals, partitives, and, according to some analyses, noun phrases headed by the morpheme *any* (Horn 1999, among others).

In the discussions starting with early work in the philosophical oriented tradition (Frege 1892, Russel 1905) and continuing with more recent, linguistic oriented proposals (Fodor and Sag 1982, Farkas 1994, 2002, Kamp and Bende-Farkas 2006, submitted), the question about the properties, uses, and function(s) of indefinite noun phrases has been answered in several ways. From a semantic point of view, one of the most basic properties that indefinite noun phrases share is existential quantification, which can be paraphrased as “there exists” or “there is at least one”. That means that the predicate within the scope of existential quantifiers is true for at least one value of the predicate variable. Consider the clause in (1), which describes a simple transitive event and in which the indefinite noun phrases *a dog* and *some dog* have existential force. This entails that the sentence is true if at least one entity is a dog and is in the garden.

(1) A dog/ some dog is in the garden.

In addition to asserting existential generalization, indefinite noun phrases display another characteristic that pertains to structuring the information in a paragraph, rather than to the truth conditions of a clause. The so-called Novelty Condition captures the property of indefinite noun phrases to introduce new and unfamiliar entities in the discourse (Kamp 1981, Heim 1982, among others). With respect to this property, indefinite noun phrases behave differently compared to other types of referring expressions (e.g. definite noun phrases), which are used to refer back to entities that have been previously introduced in the discourse and refer to discourse-old entities.

Besides existential quantification and novelty of reference, the notion of specificity has played a crucial role in the analysis of indefinite noun phrases as well (Fillmore 1967, Karttunen 1969, Partee 1970, Kripke 1977, Ludlow and Neale 1991, Kamp 1981, Farkas 2002, von Heusinger 2002, 2011). In pre-theoretical terms, a speaker uses a specific indefinite noun phrase as an expression of his intention to refer to a particular referent he “has in mind”. Specificity does not describe particular aspects that apply to all indefinite noun phrases in a unitary way, but rather distinguishes between a variety of uses and readings of indefinite noun phrases. Given the notoriety of the notion of specificity in the discussions involving indefinite noun phrases, it is not surprising that it is employed in investigations of different phenomena cross-linguistically (Enç 1991, Farkas 2002, von Heusinger 2011).

Another line of research has analysed different types of referring expressions in terms of activation and accessibility. The starting point for such investigations is the observation that human communication is not an arbitrary exchange between different speakers in the sense that interlocutors do not randomly switch between different types of referring expressions during the process of referring. That is, referring expressions are used to introduce and continue reference to particular entities with certain referential properties, but also with particular discourse properties. According to this view, each type of referring expression is associated with one particular referent, which is selected from a list of discourse referents that are ranked with respect to their activation or accessibility. The general assumption is that there is an inverse relation between the explicitness of the anaphoric expression (in terms of descriptive, lexical, and phonological material) and the accessibility, salience, or activation of the antecedent expression. That means that reduced types of referring expression most often correlate with accessible referents, while the opposite holds for more explicit types of referring expressions. For example, it was shown that a referring expression is likely to be associated with an accessible antecedent, if it is realized as a phonologically reduced form. The personal pronoun *er* (‘he’) in the second sentence of (2a) is more readily interpreted as coreferring with the subject *ein Student* (‘a student’) than to the object *einen Professor* (‘a professor’), even though both interpretations are felicitous. In contrast, the demonstrative pronoun *der* (‘this’) in (2b) cannot be linked to the subject or topic of the sentence, but, instead, can only be co-indexed with the second prominent referent, *einen Professor* (‘a professor’) (Bosch, Katz and Umbach 2007, Kaiser and Trueswell 2008).

- (2) (a) Ein Student<sub>i</sub> begrüßt einen Professor<sub>j</sub>. Er<sub>i>j</sub> sagt...  
a student greets a professor. He says...
- (b) Ein Student<sub>i</sub> begrüßt einen Professor<sub>j</sub>. Der\*<sub>i,j</sub> sagt...  
a student greets a professor. DemPron says...  
'A student greets a professor. He says [...].'

The use of a particular type of referring expression depends upon the accessibility of its associated referent, which is in turn computed based on the status retained by this referent in the immediately preceding discourse. As indefinite noun phrases generally do not have a discourse history as they do not refer back to an antecedent, they have been either left unaccounted for, or they have been associated with referents ranked lowest in accessibility or activation (Gundel et al. 1993, Ariel 2001).

This dissertation takes a different perspective on the analysis of indefinite noun phrases by investigating their functions within larger discourse units. I argue that despite being associated with locally non-accessible referents, indefinite noun phrases fulfil different functions that become visible in the subsequent discourse. In a communication model that assumes cooperative discourse participants, speakers use particular indefinite noun phrases to indicate their referential intention and to instruct hearers to establish a permanent discourse representation for the referents associated with those indefinite noun phrases. In addition to these requirements, the newly introduced referents differ with respect to their potential to give structure the subsequent discourse. The discourse structuring potential captures the relation between the type of indefinite expression used and the function the referent associated with this indefinite noun phrase *will* play in the upcoming discourse. This dissertation provides empirical evidence from English, German and Romanian to sustain these claims. Moreover, I will propose three measures for the discourse structuring potential of indefinite noun phrases, which can be observed in the discourse. The three measures are: the likelihood of subsequent mention and the topic shift potential of a referent and the type of referring expression used to pick up the referent for the first time in the subsequent discourse.

An account of indefinite noun phrases that integrates the role played by their associated referents in the ensuing discourse was advanced for English indefinite noun phrases headed by the simple indefinite article *a(n)* and indefinite *this*, as illustrated in (3). The difference

between the two indefinite noun phrase forms is that the latter not only introduces discourse and hearer-new referents, but also indicates the referential intention of the speaker to elaborate upon this referent (Perlman 1969, Prince 1981, Maclaran 1982, Wald 1983, Wright and Givón 1987, Gernsbacher and Shroyer 1989). As the expectation about the elaboration of the referent introduced by indefinite *this* remains unfulfilled in (3a), compared to (3b), the sentence containing this expression is rendered infelicitous. On the contrary, the simple indefinite article *a(n)* is neutral with respect to this expectation (i.e. it does not require its referent to be further elaborated upon), which is mirrored in its acceptance in both sentences (3a) and (3b).

- (3) (a) Becky wrote some thank-you notes using {a/ #this} purple pen; then she mailed the notes to her friends.
- (b) Becky wrote some thank-you notes using {a/ this} purple pen, which suddenly exploded, spilling purple ink all over Becky's clothes and furniture! (Ionin 2006: 181)

English indefinite *this* has been considered an exception found in English and few other languages, but broadening our empirical perspective and looking beyond this language will bring new insights. The present dissertation brings empirical evidence for the observation that English does not represent a singular case in which different types of indefinite noun phrases vary with respect to their preference for particular semantic readings and simultaneously with respect to their discourse structuring potential. Based on data from Romanian and German, I will show that these two languages have adopted or adapted different means to indicate that referents associated with particular types of indefinite noun phrases retain a preferential status in the subsequent discourse. The purpose of this introductory part is to present the phenomena central to the analyses in the next Chapters and to sketch out the central aspects of the claims I wish to make. The conclusion of this Chapter contains an overview of the dissertation.

## 1.1 Introducing the phenomena

This dissertation contains a case-based proposal as to how it could be accounted for the discourse structuring potential of different types of indefinite noun phrases. Besides English indefinite *this*, the two types of indefinite noun phrases central to the present investigation are indefinites headed by *so'n* in German and *pe*-marked indefinites in Romanian. In the following, I introduce the relevant contrasts in the two languages in turn.

German developed two determiners that are similar in function to English indefinite *this*, namely the indefinite demonstrative *diese(r)* (Wald 1983, Lyons 1999) and the complex determiner *so'n* (Wespel 2007, Chiriacescu 2010, von Heusinger 2011). In this dissertation, I focus solely on the latter determiner form, as it comes closer to its English cognate in terms of distribution in colloquial language. Similar to English indefinite *this*, indefinite noun phrases headed by *so'n* show a tendency for referential readings, as illustrated in (4).

- (4) (a) Anna will so'n Buch von Mircea Eliade lesen.  
 Anna wants so-a book by Mircea Eliade read  
 'Anna wants to read so-a book by Mircea Eliade.'
- (b) Anna will ein Buch von Mircea Eliade lesen.  
 Anna wants a book by Mircea Eliade read  
 'Anna wants to read a book by Mircea Eliade.'

While the indefinite noun phrase *ein Buch* ('a book') in (4b) is ambiguous between a referentially specific and a non-specific reading, the indefinite noun phrase headed by *so'n* in (4a) bears only a specific interpretation as it refers to a particular book *Anna* wants to read.

A similar contrast is found in Romanian, a language in which direct objects are realized distinctively, as only some indefinite noun phrases are preceded by the morpheme *pe*, whereas the other class of indefinites remain unmarked (Niculescu 1965, Pană-Dindelegan 1997, Cornilescu 2001, Chiriacescu 2007, von Heusinger and Onea 2008). Although this variation between a *pe*-marked direct object and a non-*pe*-marked direct object applies to

definite unmodified noun phrases in equal manner (Chiriacescu 2007, von Heusinger and Chiriacescu 2009, 2011), in this dissertation I will focus only on cases that involve postverbal indefinite noun phrases, as illustrated in (5). Note that direct object arguments in Romanian are often accompanied by clitic doubling

(5) (a) Petru l -a vizitat pe un actor.  
 Petru CL Aux. visited PE an actor  
 ‘Petru visited an actor’.

(b) Petru a vizitat un actor.  
 Petru Aux. visited an actor  
 ‘Petru visited an actor’.

Only in (5a) is the indefinite noun phrase *un actor* (‘an actor’) accompanied by the free morpheme *pe*, in (5b), the indefinite noun phrase remains unmarked. The distribution of *pe*-marking with indefinite noun phrases has been either left unaccounted for in the literature so far, or has been explained in terms of different types of specificity (Farkas 1978, Dobrovie-Sorin 1994, von Heusinger and Onea 2008). According to the latter proposal, the referent associated with the direct object in (5a) is interpreted as referring to a particular actor, the identity of which is “known” to the speaker, whereas the referent in (5b) is analysed as being neutral with respect to this issue and thus allowing for both, a specific and a non-specific reading of the referent.

Despite their preference for particular referential readings, the distinction between a specific and a non-specific reading of the two types of indefinite noun phrases in German and Romanian is not that straightforward as it might seem at first sight, though. Consider example (6), in which both the indefinite headed by *so’n* as well as the indefinite headed by *einen* (‘*a(n)*’) are compatible with a continuation sentence in which the identity of the referents they are associated with is denied. In other words, both indefinite noun phrases are get a non-specific reading.

(6) Maria will so’nen/ einen Prinz auf einen weißen Ross heiraten. Aber sie hat noch keinen kennengelernt.  
 ‘Mary wants to marry so-a/ a prince on a white horse. But she has not met one yet.’

In this dissertation I argue that the tendency for particular referential readings (i.e. wide scope, fixed reference, epistemic specificity, etc.) stems from the speaker's referential intention to introduce a referent that will give structure to the discourse in terms of referential persistence or continuation and potential topic shift.

To investigate the discourse structuring potential of indefinite noun phrases in English, German and Romanian, I analyse production and comprehension processes while referring. To this end, I will restrict the experimental studies by (i) using a constrained form of discourse, i.e. multi-sentence discourses and by (ii) focussing on the establishment of referential chains. The dependent variable will be the way writers refer subsequently to the referents introduced in the initial stories. For a brief illustration, consider the two mini-discourses given in Table 1, which are taken from the story continuation experiment reported in Chapter 5.

Table 1: Sample experimental item from the experiment on *pe*-marking (Chapter 5)

	<i>pe</i> -marked direct object	non- <i>pe</i> -marked direct object
S0	<i>Anul trecut când am fost la mare am cunoscut un salvamar<sub>1</sub>. (Pro)<sub>1</sub> era tot timpul activ. La sfârșitul sejurului meu, (pro)<sub>1</sub> a salvat-o pe o fată<sub>2</sub> de la înec.</i>  'Last year when I was at the seaside I met a lifeguard <sub>1</sub> . He <sub>1</sub> was very active all day long. At the end of my stay there, he <sub>1</sub> saved <b>PE-a girl</b> <sub>2</sub> from drowning.'	<i>Anul trecut când am fost la mare am cunoscut un salvamar<sub>1</sub>. (Pro)<sub>1</sub> era tot timpul activ. La sfârșitul sejurului meu, (pro)<sub>1</sub> a salvat o fată<sub>2</sub> de la înec.</i>  'Last year when I was at the seaside I met a lifeguard <sub>1</sub> . He <sub>1</sub> was very active all day long. At the end of my stay there, he <sub>1</sub> saved a <b>girl</b> <sub>2</sub> from drowning.'
S1	<i>M-a mirat ca (pro)<sub>1</sub> a reusit sa o<sub>2</sub> salveze, pentru ca (pro)<sub>1</sub> era un burtoș.</i>  'I was surprised that he <sub>1</sub> could save <b>her</b> <sub>2</sub> , because he <sub>1</sub> was a big-bellied man.'	<i>(pro)<sub>1</sub> Era un tip solid, mustacios si increzut.</i>  'He <sub>1</sub> was a strong man, moustachioed and conceited.'
S2	<i>Dar cum (pro)<sub>1</sub> a vazut-o pe fata<sub>2</sub> ca (pro)<sub>2</sub> striga dupa ajutor, (pro)<sub>1</sub> nu a stat pe ganduri si (pro)<sub>1</sub> s-a dus dupa ea<sub>2</sub>.</i>  'But as soon as he <sub>1</sub> saw <b>PE-girl</b> <sub>2</sub> that <b>she</b> <sub>2</sub> was screaming for help, he <sub>1</sub> did not hesitate and went after <b>her</b> <sub>2</sub> .'	<i>Desi (pro)<sub>1</sub> era enervant, multe fete il<sub>1</sub> placeau.</i>  'Despite the fact that he <sub>1</sub> was annoying, many girls liked him <sub>1</sub> .'
S3	<i>Fata<sub>2</sub> era inconstienta cand (pro)<sub>2</sub> a fost adusa la tarm.</i>  ' <b>The girl</b> <sub>2</sub> was unconscious as <b>she</b> <sub>2</sub> was brought to shore.'	<i>Dupa ce (pro)<sub>1</sub> a salvat-o pe fata<sub>2</sub> de la inec, (pro)<sub>1</sub> a devenit atractia plajii.</i>  'After (pro) <sub>1</sub> saved <b>PE girl</b> <sub>2</sub> from drowning, (pro) <sub>1</sub> became the attraction of the beach.'
S4	<i>(pro)<sub>2</sub> nu avea mai mult de 10 ani.</i>  ' <b>She</b> <sub>2</sub> was not older than 10.'	<i>Toata lumea vroia sa se fotografieze cu acest barbat curajos<sub>1</sub>.</i>  'Everyone wanted to be photographed with this brave man <sub>1</sub> .'

Note that the indefinite noun phrases found in direct object position in the initial sentence (i.e. S0) constitute minimal pairs, in the sense that the descriptive material of the noun phrases and the verbs selecting these objects are almost the same. I only manipulated the type of indefinite noun phrases used in direct object position in the two contexts (i.e. a *pe*-marked direct object in the left column of vs. a non-*pe*-marked direct object in the right column of Table 1). The two direct objects in Table 1 differ with respect to their discourse structuring potential. The referent of the direct object preceded by *pe* is introduced in sequence S0 by means of the indefinite noun phrase *pe o fata* ('PE a girl'). Within the next sentences – (S1) through (S4) – the referent of the indefinite noun phrase is mentioned six times. On the contrary, the referent of the non-*pe*-marked direct object *o fata* ('a girl'), on the contrary is mentioned in the discourse following its first mention in S0 only once, in continuation sentence three (i.e. S3). Based on larger discourse contexts like the ones illustrated in Table 1, I show that the functions of the three indefinite markers (*this*, *so'n* and *pe*) discussed in this Section are best accounted for in terms of their discourse structuring potential. The basic claims I wish to make are outlined in the following Section.

## 1.2 Goals of the dissertation and proposed analysis

The main goal of this dissertation is to account for the distribution of the three types of indefinite noun phrases introduced in the previous Section. I will argue that English indefinite *this*, German *so'n* and Romanian *pe*-marking share several common characteristics at the semantic and discourse-pragmatic level, which distinguish them from noun phrases headed by the simple indefinite article *a(n)*. First, at the semantic level, the three types of indefinites show a strong and robust affinity for particular referential properties (i.e. fixed reference, wide scope, epistemic specificity). Second, I argue that any account of indefinite noun phrases that investigates only their sentence-level properties fails to capture their discourse-pragmatic contribution. By analysing larger discourse units (i.e. containing more than two adjacent sentences) in which such indefinite noun phrases occur, I will show that compared to their simple indefinite correlates, referents associated with indefinite *this* in English, indefinite *so'n* in German and *pe*-marked indefinites in Romanian are more prone to: (i) be frequently mentioned in the ensuing discourse and to (ii) shift the discourse topic. At first sight, these two discourse effects call for an analysis in terms of accessibility or activation. Most interestingly, however, the above-mentioned

effects appear with a latency of one or two subsequent sentences after the indefinites were mentioned for the first time. Furthermore, another argument that cannot be easily accommodated within existing models based on accessibility or activation is the observation that referents introduced in this way show low rates of subsequent pronominalization. Despite being referentially highly persistent and showing a high topic shift potential, these referents are picked up by more explicit types of referring expressions (e.g. definite noun phrases). I propose that indefinite noun phrases function as discourse structuring devices. Their main function is to establish coherence at a more global level, by pre-activating particular referents. The pre-activating function is a signal to the hearer that the referent introduced in this way will be mentioned next and will eventually become a topical referent.

These results can be best accommodated in an expectancy-based or forward-looking model of language processing like the one put forth by Arnold (1998), which assumes that language processing is predictive as speakers use different signals to indicate to their hearers what to expect next. However, compared to previous accounts on language expectation, the model introduced in Chapter 2 makes an essential distinction between two main types of expectations, which are triggered by different factors at different discourse levels. The first type of expectation, which is generally discussed in the literature on pronoun resolution in terms of accessibility, can be derived from structural factors such as a particular syntactic position and a certain argument structure. Based on such factors, hearers make probabilistic expectations about those referents that are more probable to be mentioned next and more prone to be pronominalized. The second type of expectation, which is important for the present analysis, comes about by discourse structuring devices such as the three types of indefinite noun phrases discussed in this dissertation. Such linguistic devices target larger discourse units in that they indicate which referent will be mentioned over larger discourse segments. It is important to note that the discourse structuring potential is not to be confounded with accessibility or activation, as different types of referring expressions that are associated with a high discourse structuring potential are not necessarily more activated or accessible compared to referring expressions associated that are not associated with a discourse structuring potential. Moreover, I argue that the expectancy of continuation (i.e. the forward pointing potential) is a property defining indefinite noun phrases, but that at the same time it is a fine-grained feature that differentiates between several sub-types of such expressions. In other words, a referent

mentioned for the first time by means of an indefinite noun phrase is more or less prone to be mentioned again in the following discourse. Besides showing referential persistence, indefinite noun phrases associated with a high discourse structuring potential additionally are more prone to become the topic in the subsequent discourse. Moreover, I will show that the three languages under investigation do not differ with respect to the expectancy types they employ, as both sentence-by-sentence effects and discourse structuring effects are realized. What is different is the formal marking employed by each language for each function and the relative weighting that languages assign to each of them. The main suggestion resulting from the presented evidence is that expectancy operates in parallel at different levels of representation.

### **1.3 Limitations of the investigation**

Before presenting the outline of this dissertation, I will enumerate several limitations of this investigation. First, this dissertation is focused on the semantic and discourse-pragmatic functions of indefinite noun phrases. While I will include brief discussions on other types of referring expressions especially when dealing with notions such as activation, accessibility and topicality, future research will show to what extent the analysis put forth in this dissertation is applicable to other types of referring expressions as well. More recent investigations on the discourse-pragmatic effects of *pe*-marking as differential object marking in Romanian bring promising evidence for an analysis of definite noun phrases in terms of discourse structuring devices.

Second, I restrict my analysis to indefinite noun phrases found in direct object position and do not discuss the semantic and discourse-pragmatic contribution of indefinites in subject position. One of the reasons that contributed to this decision is the intention to compare the observations and experimental results from the three languages relevant for this dissertation. As Romanian *pe*-marking only occurs with direct objects, I limited the analysis of indefinite *this* in English and *so'n* in German to their distribution in direct object position. Another reason for concentrating on the direct object position is due to the well-attested bias of subject referents to be associated with accessible referents and of direct object referents to be associated with less accessible referent. The literature in the

field has left the investigation of referents associated with less accessible referents mostly unaddressed.

## 1.4 Outline of the Dissertation

In the remainder of this introductory part, I present the outline of this dissertation. Chapter 2 reviews research on the discourse properties of different types of referring expressions, which will be important for the investigation of the indefinite noun phrases discussed in Chapters 3-5. The accounts presented in this Chapter differ in many respects, but one of their common denominators is the notion of accessibility. The proponents of the first account I discuss (Chafe 1975, Ariel 1988 and Gundel et al. 1993) analyse accessibility in terms of *attention* or *cognitive activation* while the second account to accessibility investigates it in terms of *topicality* (Givón 1983, Grosz, Joshi and Scott 1995, Walker, Joshi and Prince 1998, Arnold 2010). In this book, I will adopt an *expectancy-driven* account to accessibility (Arnold 1998, 2010, Kehler, Kertz, Rohde and Elman 2008), as this model distinguishes between two types of expectancies: (i) expectancies with respect to the referent who will be mentioned next and (ii) expectancies that determine the type of referring expression used for a particular referent. Most crucial for the present investigation is the fact that an expectancy-driven account allows for an analysis of the referents of indefinite noun phrases in terms of discourse structuring devices.

Despite the fact that most of the above theories focus on pronoun resolution (i.e. on the factors that license the use of a pronoun), I summarize those parts of their accounts, which deal with the import of indefinite noun phrases in structuring the following discourse. By the end of the Chapter, I will propose two textual characteristics to measure the discourse structuring potential of indefinite noun phrases, namely (i) referential persistence, and (ii) topic shift. These factors will be used in the story-continuation experiments presented in the following three Chapters to test the discourse structuring potential of indefinite *this* in English, German indefinite *so'n* and Romanian *pe*.

In Chapter 3 I extended former approaches and argue for an analysis of indefinite *this* in English as indicating the speaker's referential intention to introduce a discourse prominent referent in the following text. First, in line with other findings (Prince 1981, Ionin 2006), I

confirm the preference of indefinite *this* for particular referential readings. Second, the experimental results reported in this Chapter indicate that indefinite *this* functions as a discourse-structuring device. This function is reflected by the high referential persistence and by the topic shift potential of the referents marked in this way compared to their simple indefinite counterparts.

In Chapter 4, I investigate the behaviour of indefinite noun phrases headed by the indefinite determiner *so'n*. I show that *so'n* indefinites display a strong preference for referential readings, as rigid reference, epistemic specificity. This theoretical consideration receives further support from a web-based sentence interpretation task (Experiment 1), which shows that indefinite noun phrases headed by *so'n* tend to presuppose the existence of their referents in contexts that contain intensional operators and negation. Furthermore, indefinites headed by *so'n* introduce referents with particular discourse-pragmatic properties that become visible in contexts that do not contain operators at sentence level. In light of the findings of a sentence-continuation study (Experiment 2), I conclude that referents headed by *so'n* show similar effects upon the subsequent discourse as English indefinite *this* as they are referentially highly continuous and more prone to become shift the topic in a subsequent matrix clause.

Chapter 5 is centred on the distribution of *pe*-marking with indefinite noun phrases as an instance of Differential Object Marking (DOM) in Romanian. I show that *pe*-marking with indefinite noun phrases cannot be thoroughly accounted for, unless the broader discourse contexts in which such noun phrases are used is investigated. More precisely, contrary to previous approaches on DOM in Romanian (e.g. Dobrovie-Sorin 1994), I argue that (referential, scopal or epistemic) specificity is not sufficient to account for the distribution of the *pe*-marker in the domain of indefinite noun phrases and that *pe*-marked direct objects are also sensitive to the discourse context they appear in, changing the discourse properties of their referents. Based on the findings of an off-line story-continuation experiment, I show that *pe*-marked referents are devices for structuring the discourse being interpreted by hearers as signals of high referential continuity. At the same time, *pe*-marked referents are more prone to shift the topic of the discourse. These findings receive further support from the analysis of definite unmodified noun phrases in direct object position that are characterized by the same discourse structuring potential.

Chapter 6 reviews the theoretical observations and the experimental evidence described in this dissertation, which pointed out that different types of indefinite noun phrases give structure to the subsequent discourse in different ways. First, I will argue that the indefinite noun phrases discussed in this dissertation achieve stability of reference (Farkas and Brasoveanu 2010) via prominence at the discourse level in terms of discourse structuring potential. Discourse prominence prevents these noun phrases to scope under certain operators. Thus, the indefinite noun phrases discussed here show a preference for referential interpretation because they are discourse prominent, but they are not discourse prominent because they are specific. Second, I will discuss why languages use different devices to mark referents that will play a preferential role within a discourse. This proposal will build on several investigations on the grammaticalization path of the indefinite article cross-linguistically (Heine 1997, Givón 1981, Stark 2002).

The final Chapter sums up the main conclusions and discusses discourse theories against the findings brought to light in the more empirical Chapters. It is argued that any theory concerned with the status of a referent in the discourse has to distinguish between (at least) two properties of this referent, namely its (local) accessibility or prominence status and its discourse structuring potential. This Chapter also discusses some loose ends that suggest directions for further work.

## 2. Past work on reference

As noted in the introductory Chapter, this dissertation is concerned with the use of a particular class of indefinite noun phrases in English, German and Romanian. In a plain context, speakers of the aforementioned languages could use (seemingly interchangeably) two types of indefinite noun phrases to refer to the same referent. In each of the three languages, one indefinite noun phrase form is headed by a special marker (i.e. indefinite *this* in English, *so'n* in German and *pe*-marking in Romanian,), whereas the other form remains unmarked (i.e. is headed by the simple indefinite determiner *a(n)*). While the propositional contribution of the two types of indefinite noun phrases seems to be the same, the two types of indefinite expressions differ with respect to their contribution at the discourse level. A question that arises at this point is, on what grounds do speakers choose one type of referring expression over the other, so that their hearers can successfully identify the referent in question and generate expectations with respect to the subsequent discourse?

The three markers central to the present analysis, namely English indefinite *this*, German indefinite *so'n* and *pe*-marking in Romanian are used to introduce new referents in the discourse. Moreover, their sentence semantic contribution can be partly explained by appealing to their particular referential properties. In Chapters 3-5, I will discuss in turn the referential behaviour of each of the three types of indefinite noun phrases in sentences containing operators and in plain and transparent sentences and show that while they display an affinity for referential readings, this correlation is imperfect. In other words, while all three types of indefinite noun phrases tend to bear specific readings (compared to their simple indefinite counterparts), we still find sentences in which the indefinite noun phrases are non-specific. Thus, the use of these indefinite noun phrases cannot be (solely) accounted for in terms of their sentence-level contribution. An analysis of the noun phrases headed by English indefinite *this*, German *so'n* and *pe*-marking in Romanian in terms of their discourse function seems straightforward, but it simultaneously raises several questions, which I aim to answer in the remainder of this Chapter. First, it is important to note that the centrepieces of previous studies have been, with some exceptions, definite noun phrases, while the analysis of indefinite noun phrases has been often left unaddressed. The latter are either considered not to be referential expressions at all, or not to contribute

in a significant way to the overall discourse structure. Thus, the first question that arises is, what the discourse contribution of indefinite noun phrases is and how it can be tested? More specifically, what are the factors that indicate their function at the discourse level? Second, most accounts investigating the contribution of different types of referring expressions limit their analyse to short discourse segments, consisting of two adjacent sentences. Another question that arises is, whether the impact of indefinite noun phrases can be observed non-locally, within larger discours (i.e. consisting of more than two adjacent matrix sentences)? And finally, a more general question is, whether the observed effects can be traced back to accessibility alone, or whether we need other mechanisms to account for the observed effects?

This Chapter is structured as follows: In the first Section, I discuss semantic-pragmatic accounts on definite and indefinite noun phrases, which provide an explanation for the choice for a particular type of referring expression on the basis of their contribution at sentence level. The next Sections are dedicated to several approaches that motivate the use of different types of referring expressions in terms of the accessibility of their associated referents. First, in Section 2.2.1 I discuss accessibility in terms of *cognitive activation* (Ariel 1988, 1990), which considers accessible referents as being more activated in the minds the conversation participants. A particular level of activation licenses the use of a type of referring expression over another. The more activated a referent is, the less explicit the type of referring expression used to refer to this referent will be chosen. And conversely, the less activated a referent is, the more lexical material is needed to refer back to it. Another line of research showed that different semantic-pragmatic factors determine the proper use and interpretation of a given type of referring expression. In Section 2.2.2 I discuss five of these factors that have been shown to contribute to the accessibility or prominence of referents, namely givenness, syntactic prominence, semantic prominence, implicit causality and coherence relations. In Section 2.3, I then introduce an alternative model of accessibility in terms of *topicality* (Givón 1981, 1983, Grosz, Joshi and Scott 1995; Walker, Joshi and Prince 1998, among others), which integrates some of the semantic-pragmatic factors discussed in the previous Section in a more complex model of referential management. Section 2.4 is focused on the notion of accessibility in terms of expectancy (Arnold 2001). An accessible referent is one that is likely to be mentioned in the following discourse. The three special indefinite markers (i.e. indefinite *this*, *so'n* and *pe*) will be analysed in terms of the expectancies they trigger on comprehenders' side. In

Section 2.4.2 I introduce the discourse analysis method that will be used in the Chapters to follow to determine the discourse structuring potential (in terms of discourse prominence) of indefinites headed by *this*, *so 'n* and *pe*. The last Section concludes the present Chapter.

## 2.1 Semantic-pragmatic accounts on (in)definiteness

One possible way to answer the question on what grounds speakers choose between different types of referring expressions at a particular stage in the discourse is to look at the semantic-pragmatic contribution of referring expression. Semantic theories account for a speaker's referential choices by distinguishing between two main types of expressions, namely definite and indefinite noun phrases. Languages vary with respect to the lexical or morphological means they employ to mark noun phrases. In English, for example, definiteness can be realized by: the simple definite article (*the*), demonstrative articles (e.g. *this/ that*), personal pronouns (e.g. *I/ you*), (unmodified) proper nouns (e.g. *Mary, Germany*), possessive adjectives (e.g. *my house*), and quantifiers (e.g. *every, all*). Indefiniteness marking can be realized by the simple indefinite article (*a(n)*), and by different quantifiers (e.g. *some, any, one*). Research in the field has generally focused on the prototypes of definite and indefinite expressions, namely on singular nouns headed by the definite article *the* and the indefinite article *a(n)* (Russel 1905, Christophersen 1939, Strawson 1950, Heim 1982, among others). These two types of referring expressions are thought of in terms of one particular property that applies in equal strength to all members of one of the two categories, but not to the other. In contrast to proper names and pronouns, the sentence semantic contribution of definite and indefinite noun phrases is the enrichment of the proposition with descriptive material. In distinguishing the defining properties of definite and indefinite noun phrases, uniqueness, familiarity and salience have been shown to be important factors. In classical semantics (Frege, 1892; Russell, 1905; see Heim 1991, 2011 for an overview), definite noun phrases contribute a uniqueness condition to the sentence semantics and denote a uniquely identifiable referent. Indefinite noun phrases, on the contrary, are treated as quantifiers that assert the existence of their descriptive content but are neutral towards uniqueness, as they do not imply non-uniqueness. Under the uniqueness account, the definite determiner in example (1a) signals that there is only one vase in the discourse domain to which the description *the vase* can be applied. The unique status of the entity can come about by mechanisms as different as the explicit mention of the associated referent in the previous discourse (the anaphoric use), the demonstrative use

of the referring expression, to name just a few (see Hawkins 1978 or Lyons 1999 for a detailed discussion on the different uses of definite noun phrases). The indefinite noun phrase *a vase* in (1b), on the contrary, is not specified for uniqueness. In other words, the use of the indefinite noun phrase is felicitous in a context in which we can find at least one entity in the discourse domain that fits the description of the noun phrase.

- (1) (a) Telia broke the vase.  
(b) Telia broke a vase.

Within the framework of “dynamic semantics” (Karttunen, 1976; Kamp, 1981; Heim, 1982), in which the developing context of interpretation plays an important role in mediating between linguistic form and reality, both definite and indefinite noun phrases are analysed as two types of referring expressions that contribute a discourse referent to a discourse representation. The difference between the two types of noun phrases is that definites (usually) refer back to *familiar* (i.e. already established) discourse referents, whereas indefinites canonically introduce *new* (novel) and non-familiar discourse referents into the common ground. For example, the definite noun phrase *the doll* in sentence (2a) is analysed as referring to one unique individual in the discourse whose existence is presupposed and with which both the speaker and the hearer are familiar. The indefinite noun phrase *a doll* in sentence (2b) is rather analysed as introducing a new entity in the discourse (Heim 1982, Prince 1992), and referring to an unfamiliar and non-presupposed entity. The fact that the speaker chooses an indefinite noun phrase to refer to the doll in sentence (2b) when he could have opted for a definite noun phrase indicates an asymmetry in the amount of shared knowledge between the discourse participants.

- (2) (a) Telia bought the doll on Monday.  
(b) Telia bought a doll on Monday.

In the realm of indefinite noun phrases, which constitute the focus of this dissertation, an additional distinction was made in terms of specificity (Fillmore 1967, Karttunen 1969, Partee 1970, Kripke 1977, Ludlow and Neale 1991, von Heusinger 2002). Indefinite noun phrases have been shown to display an ambiguity between specific and non-specific readings. The notion of specificity is generally understood in terms of the referential intention of a speaker. For example, the indefinite noun phrase *a Croatian* in (3) is

interpreted depending on whether the speaker intends to refer to a particular Croatian, or not. The continuation given in (3a) disambiguates the indefinite noun phrase towards the specific reading, in which the speaker intends to refer to a particular Croatian *Paula* wants to marry. The referent is fixed, independently from the interpretation of the clause. On the contrary, the continuation in (3b) disambiguates the indefinite towards a non-specific reading, in which the Croatian is non-identifiable to the speaker or to the matrix subject, *Paula*. Moreover, the observation that only (3a) allows the insertion of “a certain” to modify the indefinite noun phrase, underlines once more the different readings in terms of specificity given in example (3).

- (3) (a) Paula wants to marry a Croatian. He is her neighbour.  
 (b) Paula wants to marry a Croatian. She has to meet one first.

Based on the findings from a cross-linguistic survey, Haspelmath (1997) points out that the distinction between specific and non-specific is so important that certain languages use different types of morphological or lexical markers to encode this distinction. As researchers focussed on one or more properties of indefinite noun phrases, different views on specificity emerged. This being so, it is no surprise that the concept of specificity has been used as a general term for several partially overlapping phenomena such as: scopal specificity, epistemic specificity, partitive specificity and specificity as noteworthiness. In the literature on this topic we sometimes find the term referentiality being used as a synonymous expression for specificity (Fodor and Sag 1982, Givón 1984). However, in line with von Stechow (2002), I consider that specificity and referentiality should be teased apart and be analysed as two distinct phenomena. In this dissertation, referentiality is taken to be an indicator for the existence and accessibility in the discourse and specificity is understood in terms of the referential intention of the speaker or of another salient agent in the discourse. Contexts containing modals and non-factive propositional attitude verbs (i.e. operators, Heim 1981) make the distinction between a referential and a non-referential reading of an indefinite noun phrase clear. The indefinite in (3a), for example, scopes over the operator *want* and allows for an existential assertion of its associated referent. This allows for an interpretation in which the pronoun *he* in the second conjunct is bound by the indefinite noun phrase in the preceding sentence. The indefinite in (3a) has a referential reading and is specific as well, as it refers to an individual the speaker

has in mind<sup>1</sup>. The indefinite in (3b), on the contrary, is interpreted inside the scope of the operator.

Over the past forty years, a wealth of research has focused on the analysis of larger discourse segments. It is generally accepted that discourses are structured in such a way as to achieve coherence. In order to successfully process a coherent discourse, discourse participants must construct a mental representation of the events being described. Definite and indefinite noun phrases represent one way to achieve such representations and to keep track of the referents introduced in the discourse. It has been argued that the position of an entity in a particular discourse structure determines the use of a particular type of referring expression in the subsequent mention of that referent. Prince (1988: 1) argues that when “a speaker evokes an entity in the discourse s/he first hypothesizes the information status of that entity in the hearer’s mind, with respect to both familiarity and saliency”. In the process of referring, speakers do not randomly opt for a particular type of referring expression; rather they choose the appropriate type of referring expression based on the referential status of the referent. In other words, the use of a certain type of referring expression is negotiated at each point in the discourse between speakers and hearers. Thus, whether speakers use an indefinite noun phrase, a definite noun phrases, a pronoun, and so forth, to refer to a particular referent, depends on the information status of that entity. Prince (1988, 1992) captures the controversial notions of givenness and familiarity discussed above in a more intuitive classificational system of referential status, which impacts the type of referring expression used. She introduces two fundamental referential statuses to account for the use of different types of referring expressions, namely the discourse status and the hearer status, as illustrated in Table 2.1. The discourse status reflects the discourse history of a referent and indicates whether a referent has already been evoked in the discourse (i.e. discourse-old), or not (i.e. discourse-new). The hearer status captures another dimension of the referent that pertains to the shared knowledge between discourse participants. According to this dimension, information is shared between the discourse participants if it was explicitly introduced in the discourse, or if it represents unused, old information for the hearer, otherwise the information is hearer-new.

---

<sup>1</sup> The relation between referential and specific readings of noun phrases will be discussed in Chapters 3-5 in more detail. At this point, it suffices to note that specific noun phrases are always referential, whereas non-specific noun phrases allow for both a non-referential and a referential reading.

Table 2.1: Taxonomy of referential status (Prince 1992)

	Discourse-new	Discourse-old
Hearer-new	brand-new	-
Hearer-old	unused	textually evoked

The discourse status and the hearer status represent two dimensions of an entity and are not independent of each other. For example, if a referent is discourse-old (i.e. textually evoked), it must be hearer-old as well as it simultaneously represents shared information between the discourse participants. However, when a referent is discourse-new, it can represent both hearer-old (i.e. unused) and hearer-new (i.e. brand-new) entities. Prince (1992: 303) notes that the hearer-status is formally marked in English by definiteness, with hearer-new entities being typically realized as indefinite noun phrases and hearer-old entities being realized as definite noun phrases. The discourse status of a referent, however, remains most often unmarked.

Gundel, Hedberg and Zacharski's (1993) propose an alternative discourse-pragmatic model to account for the proper use and felicitous processing of different types of referring expressions. The starting point of their analysis is the observation that the descriptive content of a certain type of referring expression restricts the set of possible referents to which the referring expression might apply to those referents that have (at least) the designated attention and memory status for the hearer. In other words, each type of referring expression is linked to a precise memory node or cognitive status, which is in turn sufficient and necessary so that this referring expression can be used. Gundel et al. (1993) delimitate six cognitive statuses or mental representations, which are understood as processing instructions to the hearer as to which referent suits the referring expression best. The six cognitive statuses are arranged on a scale, ranging from "most restrictive" (i.e. the "in focus" status) to "least restrictive" (i.e. the "type identifiable" status). Table 2.2 illustrates the Givenness Hierarchy with English examples for each cognitive status.

Table 2.2: The Givenness Hierarchy (Gundel, Hedberg and Zacharski 1993)

In focus>	Activated>	Familiar>	Uniquely identifiable>	Referential>	Type identifiable
<i>it</i>	<i>that, this</i>	<i>that N, this N</i>	<i>the N</i>	<i>indefinite this N</i>	<i>a N</i>

Because cognitive statuses are considered parts of mental representations, the Givenness Hierarchy does not give instructions about the way in which a certain type of referring expression acquires a certain status. For example, whether a referent is discourse-old or discourse-new in Prince's (1992) terms, does not impact the cognitive status this referent is associated with. Examples (4a) and (4b) illustrate the relation between a type of referring expression and the mental activation of the referent each designates (Gundel et al. 1993).

- (4) (a) I couldn't sleep last night. A dog (next door) kept me awake.  
 (b) I couldn't sleep last night. It kept me awake.

According to this view, the hearer of sentence (4a) only has to know what the word *dog* means to understand the least restrictive simple indefinite noun phrase *a dog*. However, the hearer of sentence (4b) cannot understand the most restrictive form *it* unless s/he has a unique mental representation of the dog, which is also in his/ her focus of attention. The advantage of the Givenness Hierarchy lies in its implicational nature, which means that a referential expression is individually and separately marked for the degree of accessibility its referent codes. The statuses are not mutually exclusive and correlate with degrees of specification of some property and do not correspond to absolute degrees of the property. In other words, for any cognitive status on the Givenness Hierarchy, the associated lexical item codes the status indicated and simultaneously entails all lower statuses (the least restrictive ones, on its right). This prediction explains, for example, why a referent associated with the highest cognitive status "in focus" can be realized by referring expressions associated to all other lower statuses, i.e. "activated", "familiar", "uniquely identifiable", "referential" and "type identifiable". The indefinite determiner *this* in example (5), which codes the higher status "referential" can be replaced by the indefinite article *a(n)*, which is associated to the lowest status "type identifiable", given the unidirectional entailment relations of the Givenness Hierarchy.

- (5) I saw  $\sqrt{\text{this/}}$   $\sqrt{\text{a}}$  man in my dream, dressed as Santa Klaus<sup>2</sup>.

The cognitive statuses introduced by Gundel et al. (1993) could be thought of as being in part encoded in the grammar, as a set of constraints on certain lexical items, and in part as

---

<sup>2</sup> <http://Section.thisman.org/dreams.htm> [visited: October 2010]

information processing mechanisms for the benefit of the hearer (e.g. the use of demonstratives to signal topic-shift). One interesting question with respect to the use of a particular type of referring expression in a given discourse is reference continuation once the referent has been introduced. That is, given a discourse with several (competing) referents, how do speakers introduce and continue talking about them? The next Sections are dedicated to the discussion of different models of referential management in terms activation, topicality and expectancy. The common denominator of these models on reference tracking is that a referent's history (i.e. its discourse status according to Prince (1992)) in terms of a set of multiple factors associated with it, affects its current discourse status.

## 2.2 Accessibility as activation

Assuming that the process of referring is dynamic, constantly changing, and that referents are more or less accessible as the discourse unfolds, different theories of accessibility or salience emerged, as an attempt to capture the correlation between a particular type of referring expression and the discourse features the referent associated with it encodes. In verbal communication, participants keep mental representations of the entities introduced in the discourse (Bransford et al. 1972, Clark and Haviland 1977). These mental representations help discourse participants organize the information at hand by keeping track of the information that was already mentioned and linking the new information to the already established discourse model. Important to note is the fact that it is information that is more or less accessible at a given point in time and not the referring expression per se. In such frameworks, where *accessibility* is regarded as a gradient category rather than a categorical one, a discourse referent can be more or less accessible. The basic idea behind these theories is that referring expressions are used to introduce and continue reference to entities with particular referential properties, but also with discourse properties such as an activation level that impacts their accessibility for reference with subsequent anaphoric expressions. Returning now to the difference between definite and indefinite noun phrases, the aforementioned theories predict that a salient, accessible or activated referent will be introduced by means of a definite noun phrase, whereas non-accessible, non-salient and non-activated referents will be introduced with an indefinite noun phrase.

The term *accessibility* is used differently in Ariel's (1990), Lambrecht's (1994) and Chafe's (1976) approaches. Ariel uses the term in a more general way as cognitive availability, while for Chafe (1976) and Lambrecht (1994) the term denotes semi-active referents only. Here and throughout this dissertation I use the term accessibility for information more easily retrievable. The approaches to accessibility discussed here propose different (but not necessarily contradictory) explanations for why speakers employ different types of referring expressions. The main aim of this Chapter is to filter out several quantifiable measures of accessibility to be used Chapters 3 to 5 for measuring the discourse behaviour and prominence of referents headed by *this*, *so'n* and *pe*. While I will describe the main approaches to accessibility to delimit the textual characteristics that contribute to the accessibility of different types of indefinite noun phrases, I will keep an eye on the data to be discussed in more detail in the following Chapters. I will start by discussing the approach to referential management in terms of activation and then present the views on accessibility as topicality and expectancy.

### 2.2.1 The Accessibility Hierarchy

Starting from the premise that human communication is not an arbitrary exchange between language users in that interlocutors do not switch between different types of referring expressions at random, Ariel (1988, 1990) analyses the systematic patterns behind referring. Cornish (1999) and other linguists proposed a cognitively motivated account of different types of referring expressions, according to which determiners are tied to different cognitive states and/or statuses (i.e. memory locations and attention states). At the heart of the Accessibility Hierarchy (Ariel 1988, 1990) is the interplay between human memory and the types of referring expressions that are responsive to it. Departing from Chafe's (1976) threefold distinction of activating states (i.e. activated, semi-active and inactive) which, in her view, cannot account for the whole range of data found cross-linguistically, Ariel (1990) advocates in favour of the idea that different types of referring expressions mark *different* degrees of accessibility of their associated referents. According to this analysis, speakers choose different types of referring expressions to indicate to their addressees the degree of assumed accessibility of their associated referents. In other words, the speaker uses a referring expression to instruct the addressee to retrieve a piece of information from his memory by indicating to him how accessible this piece of information

is to him. For Ariel, all types of referring expressions code a different and precise degree of mental accessibility, functioning as markers of accessibility. The prediction made by her theory is that, when an entity is less accessible, the type of referring expression used by the speaker will be more elaborated (e.g. a definite (possibly modified) nominal phrase). And conversely, the higher the accessibility of a referent, the less explicit the type of referring expression chosen by the speaker will be (e.g. a pronoun). What distinguishes Ariel's approach from Gundel et al.'s (1993) is observation that for the former, accessibility is affected at any time in the discourse by four important factors, namely: the *distance* between the anaphor and the antecedent, the *competition* for saliency between the referents present in an utterance, *topichood* (i.e. subjecthood) and *unity*, or whether the antecedent and the anaphor are in the same discourse segment. The Accessibility Theory advanced by Ariel (1988) predicts on the basis of these four factors, the degree of accessibility of a referent at a particular point in the discourse. In turn, the accessibility of a referent will determine the type of referring expression used to pick up that referent in the subsequent discourse. Let us consider example (6), which introduces two referents of different gender, *Paul* and *Ana*. According to the Accessibility Theory, the referent associated with the proper name *Paul* is more accessible, as he is mentioned in the immediately following sentence in a topic position (i.e. grammatical subject position). In this case, a pronoun referring back to this referent seems more natural than a full definite noun phrase. The second referent, *Ana*, is taken up in the immediately following sentence as well. However, as this referent was not introduced in a topical position (i.e. it is realized as a direct object rather than as a subject), the use of a noun phrase to pick this referent up seems more appropriate than referring to it by means of a pronoun.

- (6) Paul<sub>1</sub> met Ana<sub>2</sub> in the park. He<sub>1</sub>/ the boy<sub>1</sub> was enthusiastic to meet the girl<sub>2</sub>/ her<sub>2</sub> again.

Based on a vast corpus analysis of English and Hebrew texts, Ariel (1988, 1990) investigated the distribution of different types of referring expressions, which were eventually ranked on a scale of accessibility, as illustrated in (7). The types of referring expressions important for Ariel are ranked on a continuum, ranging from highest accessibility markers at one end, which are the briefest and least informative forms (e.g. zero anaphors, pronouns), to lowest accessibility markers at the other end, which are the most informative.

## (7) The Accessibility Marking Scale by Ariel (1990: 73), with my own examples

<b>Marking Scale</b>	<b>Examples</b>
<i>Low Accessibility</i>	
Full name + modifier	Ionel Haiduc, the president
Full ('namy') name	Ionel Haiduc
Long definite description	The nice and consequent president
Short definite description	The president
Last name	Haiduc
First name	Ionel
Distal demonstrative + modifier	that car we sold last week
Proximal demonstrative + modifier	this car we sold last week
Distal demonstrative + NP	that car
Proximate demonstrative +NP	this car
Distal demonstrative	that
Proximate demonstrative	this
Stressed pronoun + gesture	SHE (with gesture)
Stressed pronoun	SHE
Unstressed pronoun	she
Cliticized pronoun	(no examples in English)
Reflexive	herself
▼ Gaps	zero, traces and agreement
<i>High Accessibility</i>	

Despite the fact that quantified and indefinite nominal phrases represent one of the most common means to introduce new referents in a given discourse, Ariel (1990) does not include this type of referring expressions on her scale to signal an even lower accessibility than full proper names and definite descriptions. A first reason for their exclusion might be the fact that indefinite descriptions represent new information and thus lack an antecedent. When establishing the accessibility of a referent, its discourse history is of utmost importance for the Ariel. More concretely, the two textual characteristics “distance” and “unity” of any anaphoric expression can be determined only in connection with an antecedent. As indefinite noun phrases do not have antecedents, no prediction can be made about their accessibility in terms of distance and unity. A second possible reason for the

exclusion of indefinite noun phrases from the Accessibility Marking Scale is the non-referential status sometimes attributed to this type of noun phrases (Russell 1919, Evans 1977, 1980). However, as Lewis (1979: 180) puts it, indefinite noun phrases may nevertheless raise the salience of particular individuals “to pave the way for referring expressions that follow”. In sum, even though Ariel’s theory (2001) cannot make any predictions with respect to the accessibility of a referent introduced by an indefinite noun phrase, we will nevertheless keep in mind that accessibility is a graded phenomenon and that different factors seem to be at work in determining the accessibility of a particular referent.

Up to this point it was noted that one of the fundamental questions within theories of language processing concerns the online-tracking of referents. This includes questions about what referents of a sentence speakers (re-)use in the subsequent discourse and what type of referring expression they choose for this referent. The majority of studies presented so far have adopted a *backward-looking* perspective, focusing on the relation between an anaphoric expression and its antecedent and the different factors that determine the use of a particular anaphoric expression at a particular point in the discourse (Givón 1983, Gernsbacher 1989, Arnold 1998, Ariel 2001, Chiriacescu 2011a, Kehler et al., 2008 Kaiser 2011). It is generally assumed that each anaphoric expression is linked to one referent that is selected from a list of discourse referents that are ranked with respect to their accessibility. The general assumption is that there is an inverse relation between the explicitness (in terms of descriptive, lexical, and phonological material) of the anaphoric expression and the accessibility of its antecedent expression. In other words, a less explicit type of referring expression often correlates with an accessible referent; the opposite holds for more explicit types of referring expressions (e.g. definite noun phrases). As pronouns were shown to be associated with highly accessible referents, most research has investigated the factors that contribute to the proper use and interpretation of pronouns in English. In what follows, I will briefly discuss some of the more important factors that have been shown to contribute to the likelihood of subsequent pronominalization.

## 2.2.2 Factors contributing to the accessibility of referents

The purpose of the next Sections is to introduce some of the major factors that have been shown to influence the accessibility of referents, namely: recency, givenness, syntactic prominence, semantic prominence, implicit causality and coherence relations. I will discuss each factor in turn.

### 2.2.2.1 Recency

Linguistic and psycholinguistic research (Ariel 1990, Brennan 1995, Chafe 1976, Givón 1983, among others) observed that recency of mention, or the distance between antecedent and anaphor impacts the likelihood of pronominalization. Recency can be measured by taking into account the last 20 clauses or more to the left of the referent (Givón 1983, du Bois 1987) or only several clauses back (Clark and Sengul 1979, Brennan et al. 1987, Ariel 1988). However, many studies acknowledge the importance of the immediately previous clause in reference resolution in comparison to referents mentioned two or three clauses back. Let us consider example (8), which introduces two referents, *Paul* and *a girl*. Recency of mention affects the type of referring form used to refer back to these referents. In sentence (8a), the referent of the *girl* is picked up in the immediately following sentence, thus, the use of a pronoun is felicitous. In (8b), a sentence intervenes between the two referring expressions associated with the referent of the *girl* and a full lexical noun phrase is preferred over a pronoun.

- (8) (a) Paul<sub>1</sub> met a girl<sub>2</sub> in the park. He<sub>1</sub> said to her<sub>2</sub> that it would be nice to go and watch a movie.
- (b) Pau<sub>1</sub> met a girl<sub>2</sub> in the park. He<sub>1</sub> was enthralled that it was so warm outside. He<sub>1</sub> said to ?her<sub>2</sub>/ the girl<sub>2</sub> that it would be nice to go and watch a movie.

Example (8) seems to indicate that if several sentences intervene between two mentions of the same referent, then the mental representation of this referent decreases in activation and referring back to this referent by means of a pronoun is less likely. Furthermore, if an entity has not been recently mentioned, other referents that have been mentioned in the intervening sentences might compete for the same accessibility status. Besides recency of mention, givenness has been also associated with accessible information.

### 2.2.2.2 Givenness or old information

Many linguistic and psycholinguistic investigations (Chafe 1976, Prince 1981, Greene et al. 1994) showed that more elaborated types of referring expressions are used for new entities, whereas less elaborated forms are usually chosen for old or given information. The pronoun in the second sentence in (9) refers back to the subject referent, *Ana*, which represents the amount of old or given information. The second referent, *a boy*, is introduced in the next sentence by means of an indefinite noun phrase, as he was not mentioned in the discourse.

(9) *Ana*<sub>1</sub> went to the seaside this morning. *She*<sub>1</sub> was happy to meet a *boy*<sub>2</sub> there.

The contrast between given and new information usually differentiates between the amount of information which has been explicitly evoked in the discourse (in contrast to implicitly introduced or inferred information). Givenness as a characteristic of discourse referents offers us a superficial insight into the reasons why speakers opt for a particular type of referring expression. Whether a discourse entity was explicitly mentioned in the previous discourse or not is neither sufficient nor necessary to describe the contrast between accessible and less accessible information. In example (10), both *Mary* and *Paula* are introduced in the first sentence by means of proper names. However, even if both referents represent given information the pronoun in the second sentence is used to refer back to *Mary*, while the second referent, *Paula*, is picked up by a more explicit type of referring expression, i.e. a definite noun phrase.

(10) *Mary*<sub>1</sub> went to the seaside and saw *Paula*<sub>2</sub>. *She*<sub>1</sub> was very happy to see her friend<sub>2</sub> there.

It thus seems that some referents are more prone to be mentioned with reduced types of referring expressions than others. Furthermore, consider a context in which my addressee knows that I have a dog and that this dog suffers of cirrhosis of the liver. It seems very natural for my interlocutor to ask a sentence like the one illustrated in (11) the moment we meet, as a conversation starter, even if we did not mention the dog during the present conversation.

(11) How is he today?

On the basis of shared knowledge or common ground (Stalnaker 1974) between discourse participants, the hearer of sentence (11) is able to identify the intended referent associated with the pronoun *he*. Shorter types of referring expressions depend not only on explicitly introduced referents, but are sensitive to other types of knowledge (i.e. world knowledge, situational knowledge, discourse knowledge, among others) as well.

Summing up, the dichotomy between given and new information is not sufficient to predict which referent is more accessible at a given time in the discourse. Consequently, this distinction cannot predict which referent will be mentioned again in the discourse, or which type of referring expression will be used for this purpose.

### 2.2.2.3 Syntactic prominence bias

Structural factors, as for example the syntactic function of a referent, have been shown to impact the accessibility of a referent as well. According to various studies, referents realized as the grammatical subject are more likely to be subsequently pronominalized and picked up compared to direct object or oblique arguments (Givón 1983, Brennan et al. 1987, Gundel 1988, Crawley et al. 1990, Gordon et al. 1993, Lambrecht 1994, Arnold 1998, Walker et al. 1998, Ariel 2001). An initial sentence like the one in example (12a) is more likely to be continued with sentences (12b) and (12d), in which the initial subject is pronominalized. Note that many speakers consider a continuation as illustrated in (12c) ungrammatical, if the pronoun refers back to the direct object and not the subject referent

- (12) (a) Alex<sub>1</sub> called Tom<sub>2</sub> to share in the great news.  
(b) He<sub>1</sub> told Tom<sub>2</sub> (that he has to take a deep breath first).  
(c) ? He<sub>2</sub> told Alex<sub>1</sub> (that he has to take a deep breath first).  
(d) Tom<sub>2</sub> told him<sub>1</sub> (that he has to take a deep breath first).

Supporting evidence for a subject preference in pronoun interpretation is provided by psycholinguistic literature as well. Clark and Card (1969) showed that in a sentence memory task, grammatical subjects are recalled with more accuracy than referents realized in other syntactic positions. Givón (1983) showed that subjects are referentially more

persistent in the subsequent discourse than non-subjects and Crawley et al. (1990) reported that participants took longer to read sentences with object referents compared to sentences with subject referents. Other syntactic constructions that have been shown to impact referring choice are the topic position in languages like Japanese (Walker et al. 1998, among others) and the focus of clefts (Arnold 1998, Almor 1999, Kaiser, in press).

The preferred referent for a pronoun has been furthermore shown to correspond to the referent in the previous clause, which occupies a parallel syntactic position. Thus, the interpretation of an anaphoric pronoun is facilitated if its coreferential antecedent is found in the same syntactic position in the previous clause (Sheldon 1974, Prince 1992, Gordon et al. 1993, Smyth 1994, Arnold 1998, Ariel 2001, Kehler 2002, among others). For example, the subject of sentence (12a), *Alex*, is mentioned in the continuation sentence (12b) in the same grammatical position (i.e. subject), thus a pronoun referent is chosen to refer back to it. Taking up the direct object referent *Tom* in sentence (12a) with a pronoun in subject position, as in (12c) is less natural and unexpected, as the two mentions do not occur in grammatical parallel positions.

Research in the field (MacWhinney 1977, Gernsbacher and Shroyer 1989, Gordon et al. 1993, McDonald and MacWhinney 1995) has argued that it is not only the (parallel) subject position, which influences the accessibility of referents in terms of likelihood of subsequent mention, but that the linear order of mention matters. According to this view, the first mentioned referent in a sentence occupies a privileged slot within the sentence, independently of its grammatical function. In many languages, among others in English, German and Romanian, the first-mentioned referent is most often realized in grammatical subject position. Despite this correlation, research has shown that in cases in which these two positions can be kept apart, both factors influence reference resolution (Kaiser and Trueswell 2008). Consequently, first mentioned referents tend to be picked up by less explicit types of referring expressions than other referents.

#### **2.2.2.4 Semantic prominence bias**

The semantic or thematic role bias has often been linked to the accessibility of a referent (Stevenson, Crawley and Kleinman 1994, McDonald and MacWhinney 1995, Arnold 1998, 2001, Kehler 2002, Kehler et al. 2008). Stevenson et al. (1994) ran several written

story completion stories and showed that referents bearing certain thematic roles are more likely to be mentioned in the subsequent discourse and to show higher rates of pronominalization. For example, in a sentence describing a transitive event, in which the thematic roles Source and Goal are realized, discourse participants often chose pronouns to pick up the Goal referent (given syntactically controlled constructions). In Stevenson et al.'s study (1994), participants were given story fragment passages like in (13a) and (13b) and were asked to add a continuation sentence to the initial story. Note that the second sentence of each test item ended in an ambiguous pronoun prompt, which could refer back to both female referents introduced in the first sentence.

- (13) (a) Sarah<sub>GOAL</sub> took the cat from Rebecca<sub>SOURCE</sub>. She \_\_\_\_\_  
 (b) Sarah<sub>SOURCE</sub> passed the salt to Rebecca<sub>GOAL</sub>. She \_\_\_\_\_

The experimental findings reported in Stevenson et al.'s study showed that the stories were continued with the pronoun *she* referring to *Sarah* rather than to *Rebecca* in (13a), as *Sarah* was not only the subject referent, but was also realized as the Goal referent. In a sentence describing a transitive event in which the semantic roles are reversed, such that the Source corresponds to the subject position and the Goal to the object position as illustrated in (13b), the participants in Stevenson et al.'s (1994) experiment did not show a pronoun assignment preference towards one of the two referents. In other words, half of the participants resolved the pronoun in the second sentence to the Source=subject referent and the other half of the participants resolved it to the Goal=object referent. Stevenson et al. interpreted these results by assuming the existence of two preferences or biases being at work in contexts like in (13), namely a subject bias and a semantic or thematic role bias. While the two preferences overlap in the Goal=subject condition in (13a), resulting in more pronoun interpretations towards the referent realized in this way, the preferences compete in the Goal=object condition in (13b). As a result, an otherwise low in prominence object referent becomes more prone to be subsequently pronominalized.

#### 2.2.2.5 Implicit causality

Implicit causality has been shown to impact the likelihood of pronominalization of particular referents as well. Previous investigations have noted that besides their actual meaning, certain verbs encode additional aspects about the events they describe (Chafe 1972). These additional aspects include objects, causes and consequences that are typically

associated with the event evoked by the verb. For example, when talking about a stealing event, people activate different roles that are typically linked to this event, as “the burglar”, “the loss”, “the insurance” and so forth. Such associations licence the use of the definite article for typical objects, even though they do not refer back to an explicitly introduced antecedent. Not only typical objects are being activated by verb meanings, but also different other events that represent natural continuations of the initial event. In the given example, the cause or the consequence is activated by the verb meaning as well. This observation was taken as an argument that the verb *to steal* has an inherent feature of causation. Garvey and Caramazza (1974) argue that implicit causality is one of possibly many aspects of “world knowledge” that speakers and hearers use in communication. Implicit causality enables hearers to correctly assign an ambiguous pronoun towards the object referent in (14a) and towards the subject referent in (14b).

- (14) (a) Jane<sub>1</sub> hit Mary<sub>2</sub> because she<sub>2</sub> had stolen a tennis racket.  
 (b) Jane<sub>1</sub> angered Mary<sub>2</sub> because she<sub>1</sub> had stolen a tennis racket.

Caramazza et al. (1977)

To identify the biases in pronoun assignment, hearers have to correctly compute the causality relations between two events. For example, in (14a) they have to understand that the event described in the first sentence of (14a) was a reaction towards what *Mary* did, and in (14b), hearers have to understand that *Mary* is angered, because *Jane* stole a tennis racket. Garvey and Caramazza (1974) divided verbs into three classes depending on the direction of pronoun resolution, i.e. whether a pronoun refers back to the first noun phrase of the preceding clause (NP1-bias), whether it refers to the second noun phrase (NP2-bias), or whether it is neutral towards the preceding noun phrases (neutral). Typical verbs that are biased to the first noun phrase in the preceding context are *anger*, *frighten*, *delight*; verbs that are biased to the second referent are *hit*, *scold*, *admire*; and verbs that are neutral with respect to a particular referent are *see*, *babysit*, *notice*.

Moreover, Au (1986) showed that the way in which the first sentence is continued has an impact upon the referent that will be subsequently mentioned. For example, a full stop after the first sentence and the omission of the subordinating conjunction *because* has the effect of neutralising or even reversing a particular bias to a noun phrase, as illustrated in (15a). On the contrary, participants who were asked to continue a sentence like the one in example (15b) showed a systematic preference to pick up the first referent in the

subsequent discourse, and the pronoun in the second conjunct was resolved to this subject referent.

- (15) (a) Paula infuriated Mary. Paula/ Mary \_\_\_\_\_.  
 (b) Paula infuriated Mary because she<sub>PAULA</sub> \_\_\_\_\_.

Stevenson et al. (1994) accounted for these observations by arguing that subordinating conjunctions like *because* have a direct impact on the accessibility of referents, making these more susceptible for pronominalization. In sum, implicit causality was shown to be an important factor in pronoun resolution, which is generally resolved at clause boundaries (Koorneef and van Berkum 2006).

### 2.2.2.6 Intersentential coherence relations

More recent approaches to anaphora resolution emphasize the crucial role played by intersentential semantic coherence relations in correct pronoun assignment. Following the work of Hobbs (1979), Kehler (2002), Rohde (2008), Kehler et al. (2008), Kaiser (2010) argued that coherence relations, which hold between two adjacent clauses can be of various types, as for example causal relations (e.g. *as a result*), temporal relations (e.g. *next*), to name just a few. Moreover, discourse participants are said to build forward-looking expectations about the ensuing discourse based on the coherence relations they assume to hold between sentences. To illustrate how the coherence relations contribute to pronoun interpretation, consider examples (16a) and (16b), adapted from Kehler et al. (2008).

- (16) (a) Bush narrowly defeated Kerry, and as a result he<sub>KERRY</sub> took some days off.  
 (b) Bush narrowly defeated Kerry, and then he<sub>BUSH</sub> took some days off.

Given the *result*-interpretation in (26a), the pronoun *he* in the subordinate clause is preferentially interpreted to refer back to the direct object referent, *Kerry*, than to the subject of the sentence (i.e. *Bush*). Under a reading in which the events described in the matrix and subordinate clause represent an enumeration of the events that happened, as in (16b), the subject pronoun *he* in the second conjunct is preferentially interpreted to have *Bush* as an antecedent. An analysis of pronoun resolution in terms of coherence relations was shown to mediate between the conflicting subject and thematic roles biases, which

often yielded contradictory results (see the discussion in Kehler et al. 2008). Under carefully controlled conditions, coherence-based analyses were shown to override other semantic or structural approaches to pronoun resolution, and pronoun use was interpreted as an epiphenomenon of a more general preference to establish particular coherence relations at discourse level (Hobbs 1979, Kehler et al. 2008, Kaiser 2010).

The features discussed above, namely recency, givenness, syntactic and semantic prominence, implicit causality and coherence relations were shown to impact the accessibility of a referent. These characteristics, together with some of the predictions or tendencies usually associated with them are summarized in Table 2.3.

Table 2.3: Factors contributing to pronoun use and interpretation

<b>Bias</b>	<b>Predictions</b>
Recency	recently mentioned > not recently mentioned
Givenness	old information > new information
Syntactic prominence	subject > object > other
Parallelism	parallel position > non-parallel position
Semantic prominence	e.g. Goal > Source in a transitive event representing these two thematic roles
Implicit Causality	different types of verbs show different NP-biases
Coherence relations	e.g. Elaboration relation- preference for subject continuation

The factors presented in Table 2.3 predict that a referent which represents given information, which is mentioned as a subject, bearing the semantic Goal role (vs. Source), which is mentioned in the immediately following clause tends to be picked up by a less explicit type of referring expression. On the contrary, a referent mentioned in object position, which conveys new information, and which is not mentioned in the immediately following sentence is less accessible and will need a more explicit type of referring expression to be mentioned in a subsequent discourse. The biases presented above have been shown to impact the accessibility of referents, however, as Arnold (1998, 1999) and Kaiser (2010) have pointed out, a single-factor system cannot explain accessibility as a whole. Rather, the factors differ with respect to how influential they are relative to one another at a particular point in the discourse.

The bundle of characteristics discussed so far make predictions about the likelihood of subsequent pronominalization, but not about the likelihood of (more or less frequent) subsequent mention or about the type of referring expression used next besides pronoun use. In this dissertation I will present cross-linguistic evidence to show that the type of referring expression used (i.e. marked vs. unmarked indefinite noun phrases) can be used as a heuristic to predict the discourse structuring potential of referents introduced in this way. For example, the two referents, *Paul* and *a girl*, are introduced in the first sentence conjunct in (17). The first-mentioned referent, *Paul* is the grammatical subject and the semantic Agent of the sentence. This referent is mentioned in the immediately following sentence by a zero anaphor in a parallel subject position. These characteristics taken together indicate that *Paul* is the most accessible referent in the mini-discourses presented in (17a) and (17b).

- (17) (a) Paul<sub>1</sub> a      întâlnit      o      fată<sub>2</sub>. A      fost      impresionat      de  
 Paul Aux.      met      a      girl. Aux.      been      impressed      by  
 parul ei<sub>2</sub>      negru.  
 hair      her      black.  
 ‘Paul met a girl. He was impressed by her dark hair.’
- (b) Paul<sub>1</sub> a      întâlnit-o pe o fată<sub>2</sub>. A      fost impresionat de părul ei<sub>2</sub> negru.  
 Paul<sub>1</sub> Aux.      met      CL PE a girl<sub>2</sub>. Aux. been impressed by hair her<sub>2</sub> black.  
 ‘Paul met a girl. He was impressed by her long dark hair.’

The second referent is introduced in the first sentence of (17a) as a direct object realized as the indefinite noun phrase *o fată* (‘a girl’) and by means of the indefinite noun phrase *pe o fată* (‘PE a girl’) in (17b). The referents of the direct objects are taken up in the next sentence by a pronoun and in syntactically parallel positions (i.e. as direct objects). Thus, the second referent, *pe a girl*, in (17b) seems to be less accessible than the first-mentioned referent *Paul*, as it conveys new information and is mentioned in direct object position. However, speakers can choose between two possible constructions to refer to the same referent, namely one in which the indefinite noun phrase remains unmarked, as illustrated in (17a) and one in which the indefinite noun phrases receives the marker *pe*, as in (17b).

In the following section I discuss another view of accessibility as topicality, which focuses on entities that are going to be topical in the upcoming discourse. The two models that are discussed in what follows compute the accessibility of a referent by combining the status retained by an entity in the previous discourse with the *forward-looking* potential of this entity. Before I introduce the notion I will work with to explain the alternations found with indefinite noun phrases in English, German and Romanian, I will introduce another account to accessibility in terms of topicality.

### **2.3 Accessibility as topicality**

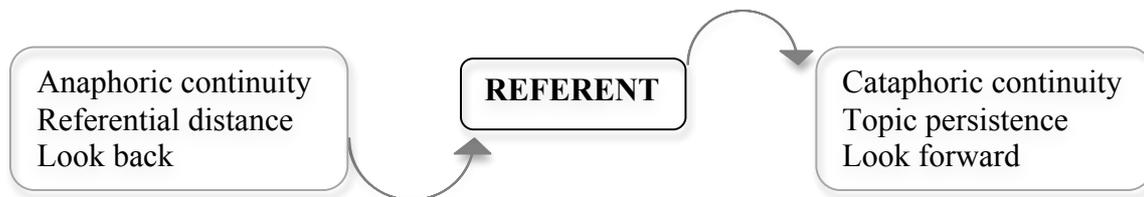
The notion of topic or topicality was in the attention of linguists and psycholinguists for a long time. For Chafe (1976), topics represent the amount of old information given in a discourse, while Reinhart (1982) describes topics in terms of “aboutness”, as discourses seem to be more about certain referents than about others. The traditional distinction between subjects and predicates made exactly this type of difference: between an entity the sentence was about (the subject referent) and another part that predicated on it (the predicate). Givón (1983) and the proponents of Centering Theory (Passoneau 1994, Grosz, Joshi and Weinstein 1995, Walker, Joshi and Prince 1998) offer two distinct but related referential management models in terms of topicality. Contrary to their forerunners’ view (e.g. Reinhart 1982), topicality is not paired with an atomic element within each clause, but is conceived as a graded notion. In comparison to the approaches to accessibility in terms of (cognitive) activation discussed in the previous Sections, which are mainly concerned with the amount of attention allocated to a particular referent in the preceding utterance, the proponents of the view of accessibility as topicality focus on entities that are going to be topical in the upcoming discourse.

#### **2.3.1 Givón’s distance model**

For Givón (1983), the amount of importance allocated by speakers to different referents is decisive. He assumes that all referents are more or less topical from the production perspective that (i) thematically more important (topical) referents tend to be picked up more often in the discourse, and that (ii) the more recurrent a referent is, the less descriptive material is needed to pick it up. In other words, less explicit types of referring

expressions will be used for recurrent or frequently mentioned referents. Bearing these two assumptions in mind, Givón (1983) proposes two measures of importance or topicality, which is not measured directly, but by referential continuity in two opposite directions, namely in the preceding and in the upcoming discourse, as illustrated in (18).

(18) Measuring accessibility as topicality in two directions (Givón 1983)



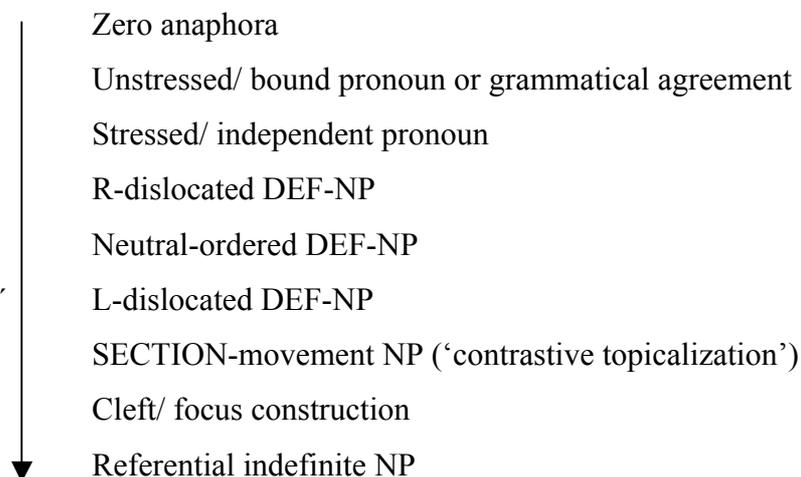
The first measure method of topicality is anaphoric continuity or referential distance, which investigates a referent's history by exploring whether the referent has an antecedent, and if so, in which sentence in the preceding discourse it occurs. The smaller the distance between antecedent and anaphor, the more topical the referent will be. This measure corresponds directly to Ariel's (1988, 1990) textual factor "distance", which was also concerned with the distance between anaphor and antecedent in terms of number of intervening sentences. The second measure method for topicality proposed by Givón (1983) is cataphoric continuity, which measures the persistence of a (topical) referent in the subsequent discourse. The prediction is that the more often a referent is picked up in the subsequent discourse, the more topical it is, thus reflecting the speaker's intentions about the role that entity will play in the continuing discourse. The advantage of Givón's (1983) model of topic continuity is that it combines forward-looking and backward-looking methods to measure the topicality of a referent in a given discourse. This model can be applied to explore the accessibility of referents introduced by different types of referring expressions, including indefinite noun phrases, which generally lack an antecedent and thus have no anaphoric force. Givón (1983) furthermore observes that *potential interference*, which refers to the copresence of more referents, which are similar in animacy or semantic and syntactic prominence, could dampen the topicality of a given referent. The prediction is that when many competing referents are present in a text, a pronoun referring back to one of them is (generally) dispreferred, to avoid ambiguity.

In a cross-linguistic analysis of texts, Givón (1983) crossed the two factors determining the topicality of referents (i.e. referential distance and topic persistence) with different

syntactic, semantic and phonologic coding devices to determine the topical constituent. The results of his findings are presented in form of a gradable scale, as illustrated in (19).

(19) Topic accessibility scale (Givón 1983)

*Most continuous/ accessible topic*



*Most discontinuous/ inaccessible topic*

The orientative scale in (19) shows in the same way as the Accessibility Hierarchy (Ariel 1988) that an entity realized as a zero anaphor, or an unstressed pronoun is an accessible topic and is most continuous in both the previous and the subsequent discourse, while the most inaccessible and discontinuous referents are realized by means of referential indefinite noun phrases.

Givón's topic accessibility scale comprises not only different types of referring expressions, as full noun phrases or pronouns, but other subscales as well, as for example: (i) the scale of phonological size (e.g. stressed pronoun > unstressed pronoun), (ii) the word order scale (e.g. R-dislocation > L-dislocation), (iii) scale of syntactic roles (e.g. subject > direct object), (iv) the passive/active scale (e.g. active > passive), (v) the scale of different indefinite grammatical markers (e.g. indefinite *this* > indefinite *a(n)* in English); the scale of (vi) definite grammatical markers (e.g. the *ga-* vs. *wa* distinction in Japanese). By combining different syntactic coding devices, Givón (1983) and his colleagues offer a detailed account for the importance of *multiple* textual aspects in determining the topical status of referents cross-linguistically. The next Section is dedicated to the discussion of Centring Theory, a framework that integrated the two measures of topicality proposed by Givón in a formal model of reference processing.

### 2.3.2 Centering Theory

The proponents of Centering Theory integrated Givón's (1983) findings in a more formal model on topicality to account for referent tracking in discourse. This theory has its starting point in the work on attention focus in discourse and was developed at the beginning of the 1980's within the framework of artificial intelligence. Firstly formulated by Grosz, Joshi and Weinstein (1995), Centering Theory was constantly adjusted and extended, most prominently by Walker and Prince (1996), Grosz and Sidner (1986) and was experimentally validated in the works of Brennan (1995) and Gordon and Chan (1995), among others. Assuming that within a discourse attention is focused or centred, the independent notion of "center" was introduced<sup>3</sup>. The "center of attention" of a discourse segment is accounted for in terms of focus of attention, choice of referring expressions and the perceived coherence between two adjacent sentences (Walker et al. 1998: 1). This relationship mirrors the formulations presented in Givón (1981, 1983) as that between the "set of potential antecedents" and the "anaphor" and integrates both referential distance and topic continuity as measure methods for topicality. Except for few works (Passoneau 1994, Grosz et al. 1995, Walker, Joshi and Prince 1998), the main concern of Centering Theory is the local or sentence-level relation (i.e. between two adjacent sentences) that holds between the centres of attention in a given discourse segment.

Proponents of this model suggest that speakers form referential predictions on a clause-by-clause level. The model introduces two sets of focused discourse referents (i.e. 'centers'): the *backward-looking center* and the *forward-looking center*. The backward-looking center (Cb) is the most prominent discourse referent in both the current and the previous utterance. In English, the Cb position is generally associated with the subject of the current utterance. Intuitively, the Cb is the discourse entity that is the center of attention of the current utterance, being psychologically the most salient discourse referent for both speaker and hearer. The Cb is the *local* discourse topic and is identical to the preferred center of the immediately following sentence in an ideal discourse. In order to achieve a coherent discourse, the identity of the Cb has to be determined for each utterance. The only exception is the first sentence of a discourse segment, which does not contain a Cb, as discourse-initial utterances are interpreted as context-creating devices and not context

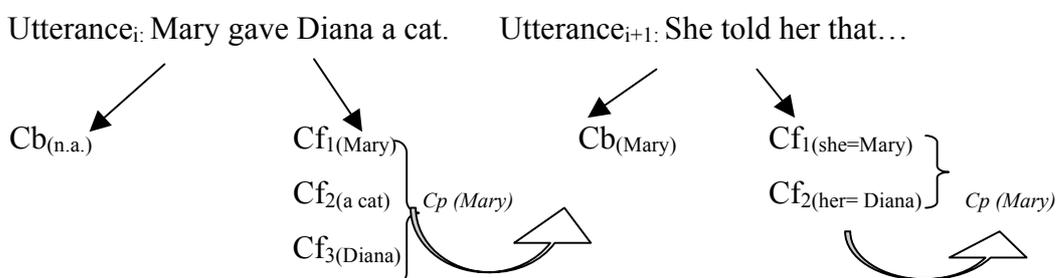
---

<sup>3</sup> This notion should be kept apart from "focus" (which might be ambiguous between a psychologically salient discourse referent and important and new information within the clause).

elaborating ones (see Lambrecht 1994: 129 for a similar argumentation).

The second type of center is the forward-looking center (Cf), which is a list of discourse referents in the current utterance that are ranked according to several factors. Centering Theory assumes that, for example, in German and English the referent in subject position is ranked higher than the referent in direct object position, which, in turn, is ranked higher than the referent in indirect object position. Thus, a referent mentioned in subject position enters the list of Cfs as the highest ranked, *preferred center* (Cp), which is the most expected Cb for the next utterance. Different instantiations of Centring Theory propose various other factors (e.g. animacy, clausal subordination and the lexical semantics of the verb) for the ranking of centers on the Cf list. Walker et al. (1998) note that the complete set of factors influencing the ranking of the Cfs has yet to be determined. The interplay between the centers of two adjacent utterances is illustrated in (20) for the sentences *Mary gave Diana a cat. She told her that [...]*.

- (20) Relationship between the forward-looking centers (Cf) and the backward-looking center (Cb) between two adjacent utterances.



In Centring Theory, transitions from an initial utterance  $U_i$  to a second utterance  $U_{i+1}$  are categorized according to the established relations between the centers. If the backward-looking center remains the same across utterances, the model outputs *continue* for the transition from  $U_i$  to  $U_{i+1}$ . If the backward-looking center remains the same but is no longer the preferred center in  $U_{i+1}$ , the model outputs the transition relation *retain*. Finally, if a new backward-looking center is introduced in  $U_{i+1}$ , which is different from the old one in  $U_i$ , the model outputs a *smooth or rough shift*, depending on whether the Cp remains the same or not. Center continuations, or transitions between utterances in which the backward-looking center is continued, are preferred in a coherent discourse. In contrast, a shift of the backward-looking center is dispreferred.

It is important to note that one of the differences between Ariel's (1990) Accessibility Marking Scale and Centring Theory is that the two approaches make different predictions. For example, while Ariel's model predicts, on the basis of four textual factors (i.e. distance, unity, competition and saliency), the type of referring expression that will be used for a particular referent, Centring Theory is primarily concerned with the referent that is most likely to be mentioned next as the most important or topical constituent. In other words, Centring Theory is interested in the factors that contribute to the likelihood of subsequent mention of a referent. The type of referring expression used to pick up a referent in the subsequent discourse is considered a by-product of a referent's expectancy to be mentioned next. The theory predicts that a Cb must be realized as a pronoun if any member of the Cf is realized as a pronoun, or if the Cb in the previous clause is realized as the present Cb as well.

Centering Theory has been widely adopted in the fields of computational linguistics and psycholinguistics; however, various aspects still need to be integrated in the existing framework. One challenge for Centring Theory is to model the referential management within longer discourse segments. Furthermore, Centering Theory could be extended to include a discussion about lower-situated entities on the Cf-list, which make good candidates for Cbs two or three sentences after being introduced in the discourse. Table 2.4 presents an example taken from the multi-sentence-continuation study on *pe*-marking in Romanian (see Chapter 5), which highlights important aspects, which cannot be accounted for by the current version of Centering Theory. Several observations can be made on the basis of the discourse segment presented in Table 2.4. In the first utterance ( $U_1$ ) a referent is introduced in subject position, which becomes the Cb in the next sentence ( $U_2$ ). As predicted by Centring Theory, the Cb is realized by means of a reduced type or referring expression in  $U_2$ . As Romanian is a pro-drop language, the re-mentioned Cb in subject position is realized by the least explicit type of referring expression, a null pronoun. In  $U_2$ , a second referent (i.e. 'pe a boy') is introduced in a non-prominent position (i.e. as an indefinite noun phrase realized as a direct object, which is preceded by the optionally occurring differential object marker *pe*) and this referent becomes the second-ranked referent on the Cf list in  $U_2$ . The story continues with the same Cb, *Graur*, mentioned in subject position in  $U_3$  and ranked highest on the Cf list of  $U_3$ . Additionally, a new referent is added in  $U_3$  (i.e. 'a lady') as the second-ranked referent, which was introduced as a *pe*-marked definite noun phrase in direct object position. The interesting observation is that

the lower-ranked Cf *pe a boy* becomes the Cb in U<sub>4</sub> and is picked up by means of a null pronoun in the same sentence. The shift of the Cb between U<sub>3</sub> and U<sub>4</sub> is formally indicated by the proximal demonstrative *this*. The third referent, which was initially introduced by a *pe*-marked definite noun phrase in direct object position, becomes the Cb in U<sub>8</sub>.

Table 2.4. Sample continuation story from the experimental study on *pe*-marking in Romanian (Chapter 5)

	Utterances	Cb	Cf list with preferred center	Transition
U <sub>1</sub>	<i>Graur s-a hotarât sa mearga în oraş.</i> 'Graur decided to go downtown.'	n.a.	[Graur]	n.a.
U <sub>2</sub>	<i>Pe drum l-a văzut pe un băiat intrând într-un magazin.</i> 'On the way, he saw <b>pe a boy</b> entering a store.'	[Graur]	[Graur, boy]	(CONTINUE)
U <sub>3</sub>	<i>Graur o recunoscuse imediat pe doamna de lângă acest băiat suspect.</i> 'Graur immediately recognized <b>the lady</b> next to this suspect boy.'	[Graur]	[Graur, lady, boy]	CONTINUE
U <sub>4</sub>	<i>Purta o şapcă Hello Kitty,</i> 'He was wearing a <b>Hello Kitty cap</b> .'	[boy]	[boy]	SMOOTH SHIFT
U <sub>5</sub>	<i>iar în mâna dreaptă avea o geantă diplomat.</i> 'and in his right hand he had a suitcase.'	[boy]	[boy]	CONTINUE
U <sub>6</sub>	<i>Tiptil, tiptil, s-a furişat după băiat în magazin.</i> 'Slowly, slowly, he followed the boy into the store.'	[Graur]	[Graur, boy]	SMOOTH SHIFT
U <sub>7</sub>	<i>Graur se temea să nu fie observant de cineva.</i> 'Graur feared someone could see him.'	[Graur]	[Graur]	CONTINUE
U <sub>8</sub>	<i>Femeia presimţi ceva.</i> 'The woman suspected something.'	[lady]	[lady]	SMOOTH SHIFT

The main observation from this example is that particular referents have the potential to shift the topic with a certain delay. The relations and restrictions between utterances of a discourse segment are analysed in Centering Theory in a linear way, such that each utterance necessarily connects with the one immediately preceding it, although in a given context, the discourse relations are not always between utterances. As it will become more evident in the discussions from Chapters 3-5, certain referents are more prone to shift the topic (i.e. become Cbs in Centering Theory terms) two or three sentences after being introduced in the discourse. The example in Table 2.4 argues in favour of an adjustment of the ranking criteria of the Cfs in such a way as to encompass referents that have the potential to become Cbs as the discourse progresses. Several other studies showed that under certain conditions, other referents might outrank subjects in salience or accessibility. Smyth and Chambers (1996), for example, found out that repeated-name penalties arise for subject and object referents alike. Moreover, they showed that reading times for pronoun

referents are faster when they refer back to a parallel object referent than to a non-parallel subject or indirect object referent. Thus, besides the grammatical roles of the referent, syntactic parallelism between Cp and Cb seems to play a role as well (Smyth 1994). Recall the discussion in Section 2.2.2, in which it was shown that several factors impact the likelihood of subsequent mention and the likelihood of pronominalization.

The theories discussed so far consider accessibility a graded and dynamic phenomenon that is subject to constant change as the discourse unfolds. While various factors have been shown to affect prominence, a constant indicator of a referent's prominence status at a particular point in the discourse is the type of referring expression used to refer back to the antecedent of this referent. Highly accessible or topical referents tend to be coded by means of attenuated types of referring expressions (e.g. zeros, pronouns). On the contrary, less accessible referents tend to be picked up by more elaborated types of referring expressions (e.g. definite (unmodified or modified) noun phrases). Moreover, the preceding discourse (i.e. the discourse in which the referent was introduced for the first time) is considered of utmost importance in determining the status of a referent in a given discourse. In other words, the way in which the antecedent was mentioned in the previous sentence(s) decisively impacts the prominence of its anaphor. Thus, establishing the prominence of a referent is a circular method, as it requires looking back at the properties of the antecedent in order to define the accessibility of the anaphoric expression. Nevertheless, this analysis has at least some advantages, as it provides us with a means to calculate the prominence of referents and make predictions about an antecedent and the correlation between this antecedent and its subsequent coreferential expressions.

The theories on accessibility presented so far focus on the preceding discourse to determine the accessibility of a given referent. Few studies focused on the behaviour of referents that are realized in structural or semantic non-prominent positions. Such non-prominent referents were considered at best 'secondary topics' (Givón 1983) or, non-preferred backward-looking centers in Centring Theory's terms. In what follows, I discuss a more recent expectancy-based account that reconciliates between the two discrepant views on accessibility as activation and topicality. This expectancy-motivated approach will be used to explain the contribution of the indefinite noun phrases discussed in this dissertation.

## 2.4 Accessibility as Expectancy

Recent psycholinguistic research has convincingly shown that statistical regularities can be observed at different levels of linguistic output, as for example at the phonetic level (Saffran, Aslin and Newport 1996), the syntactic (Elman 1993, Levy 2008) and semantic level (Tabossi 1988, Bicknell, Elman, Hare and Kutas, 2008). All these studies bring favourable evidence that hearers identify frequency patterns to predict what is likely to occur in the ensuing discourse. Bearing such regularities in mind, proponents of the expectancy-based view on language comprehension and production have investigated how speakers use regularities at discourse level and how addressees form predictions about the way in which the discourse will unfold. Many studies have shown that referents mentioned as subjects, in a syntactic parallel position, bearing prominent thematic roles display a higher probability to be mentioned again in the subsequent discourse than other entities mentioned in less prominent positions. Furthermore, the more likely it is (and thus expectable) that a referent will be mentioned in the subsequent discourse, the more accessible that referent will become for the addressee. During conversation, speakers indicate to hearers the relative probability that a particular referent will be accessible in the following discourse. Moreover, as expectable referents are high in prominence, speakers do not need much lexical material to pick them up. This explains, for example, why referents mentioned as subjects, as the thematic Goal in a transitive event, are often referred to by pronouns. Before introducing the model I will work with to explain the distribution of two different types of indefinite noun phrases in English, German and Romanian, I will present evidence from previous research, which convincingly shows that the likelihood of subsequent mention of a particular referent does not necessarily have to correspond to its likelihood of pronominalization. The dissociation between these two types of expectancies with respect to a referent's status in the subsequent discourse cast for a revision of the notion of accessibility.

### 2.4.1 Expectancy of subsequent mention and likelihood of pronominalization

Arnold (2001) put forth a model of referential management, which she coined the "Expectancy Hypothesis". In her view, the accessibility of a referent can be computed on the basis of how *expected* that referent is. Accessibility as expectancy is not in conflict with other theories on accessibility, but rather extends previous results in several ways. First,

exactly like Centring Theory, the idea of accessibility as expectancy is integrated in a *forward-looking* model, in which the likelihood of subsequent mention is computed on the basis of the discourse history of a referent (i.e. the relation between antecedent and anaphor) and on other factors as well. That is, comprehenders generate forward-looking expectations about the way in which the discourse is likely to unfold and it is these expectations that give rise to predictions about who is going to be mentioned next. For example, when encountering an ambiguous pronoun, comprehender disambiguate its associated referent based on their expectations about the subsequent discourse. In comparison to Centring Theory, the view of accessibility as expectancy is more flexible and integrates more (surface-level) factors that might impact the likelihood of subsequent mention of a referent. The factors discussed in Section 2.2.2 (i.e. recency, givenness, syntactic prominence and semantic prominence, implicit causality and coherence relations) are regarded as by-products of a more general discourse principle to establish coherence in discourse. Second, accessibility as expectancy challenges an important observation of previous accounts on accessibility, which considered the likelihood of subsequent mention and the likelihood of pronominalization equally reliable indicators of accessibility. Several investigations emphasized the fact that these two types of expectancies must be dissociated because they do not always point in the same direction. For example, Arnold's (2010) Source-Goal experiment yielded seemingly contradictory results. She found out that most references to the grammatical subjects were pronominalized (81%), whereas only 21% of the references to the grammatical objects were picked up by a pronoun. Rather unexpected was the observation that it was the Goal referent (i.e. which was realized as the grammatical object of the sentence), was more prone to be picked up in the subsequent discourse, and not the Source referent (i.e. the subject). These results showed that the Goal referent associated with the direct object became a competitor for the subject referent in terms of likelihood of subsequent mention, but not in terms of likelihood of subsequent pronominalization. Stevenson, Crawley and Kleinmann (1994) reported on similarly intriguing results when investigating the effects of different thematic roles on the subsequent discourse in terms of pronoun use and next mention bias. Investigating the consequences of alternative focus and grammatical function on the production and interpretation of referring expressions, Kaiser (2010) showed that in a sentence with a focused subject referent, participants were most likely to remention the alternative to the focused subject in the subsequent discourse. However, despite its strong probability of remention, a focused subject was most likely to be realized by a full noun phrase in the

upcoming discourse, rather than by a pronoun. Fukumura and van Gompel (2010) reported on complementary results from a story continuation study on implicit causality verbs with subordinating conjunction prompts. In one of their experiments, they used verbs biased to the subject referent and manipulated for the prompt (i.e. *Mary scared Anna after the long discussion ended in a row. This was because/so...*), as other studies showed (Stevenson et al. 1994, among others) that subordinating conjunctions change biases from the subject to the object referent. While varying the conjunctions impacted the next mention bias, this manipulation had no impact on the subject bias. That is, participants showed a preference to continue talking about the object referent, while the subject referent was still more likely to be pronominalized.

At first blush, it seems that the findings concerning referential continuity and the type of the first anaphoric expression contrast with the general opinion, which considered both factors as equal signals of accessibility. To account for such conflicting results, Kehler et al. (2008), Rohde (2008), Chiriacescu and von Heusinger (2010) and Kaiser (2010) considered that hearers develop expectations about both: the referents more likely to be picked up in the following discourse and also about the type of referring expression that will be used to mention a referent again. This analysis distinguishes between two sets of factors: those that impact the choice of next mention and another set that affects the choice of pronominalization. Kehler et al. (2002) propose a Bayesian way to formalize the relations between the different factors involved in referential processing, as illustrated in (21).

$$(21) \quad p(Ref | Pro) = \frac{p(Pro | Ref)p(Ref)}{p(Pro)}$$

According to this model, the probability ( $p$ ) that a pronoun refers back to a particular referent,  $p(Ref | Pro)$ , is dependent on two factors: the probability that this referent will be picked up in the subsequent discourse,  $p(Ref)$ , and the probability that this referent will be pronominalized  $p(Pro | Ref)$ . Furthermore, the overall probability of pronominalization  $p(Pro)$  can affect pronoun resolution and should be determined as well.

Besides distinguishing between two types of expectancies that pertain to the subsequent discourse, the accessibility as expectancy approach makes predictions in a forward-looking

manner about the status a particular referent will occupy in the subsequent discourse. As referents differ in accessibility they are more or less prone to be picked up in the subsequent discourse. Crucial for the analysis of indefinite noun phrases in direct object position is that this account makes predictions about the status a referent will enjoy in the subsequent discourse, even when it lacks an antecedent. In this sense, it resembles the forward-looking centers integrated in Centering Theory and Givón's (1983) topic continuity measure of referential persistence. While other theories exclude indefinite descriptions from their analyses, the Expectancy Hypothesis accounts for the use of indefinite noun phrases as well, and even shows that new information can sometimes outrank old information in terms of accessibility. Evidence for this observation comes from different experimental studies. For example, Arnold (2001) showed that listeners are, under certain conditions, inclined to expect new referents that are not accessible in terms of the textual characteristics discussed in Section 2.2.2. A common testing ground represent non-pathological disfluencies, as for example the use of fillers like *uhm*, *uh*, *er*, *aaah*. Arnold and Tanenhaus (2007) analysed data from a production experiment in which the stimulus items were new (i.e. unmentioned) or old (i.e. already introduced in the discourse) and were introduced by short (1-2 words) or long (3 and more words) noun phrases. They analysed each type of referring expression and found out that disfluency is paired with reference to new, unmentioned referents, while fluent mentions mostly correspond to old referents. This tendency was reported for both long and short types of referring expressions, demonstrating that disfluency does not come about due to the length of the noun phrase. In another study, Arnold (2009) showed that pronouns were scarcely used in disfluent utterances, compared to fluent ones, though the referents were highly prominent (i.e. realized in subject position). Furthermore, new entities (i.e. lacking an antecedent) were not taken up by pronouns, regardless of their fluent or less fluent character. Arnold (2009) concludes that disfluency has an impact upon the discourse structure, lowering the overall use of pronouns, especially in such cases in which the referent is given (already known to the discourse participants). Thus, disfluency, rather than being just white noise, seems to represent a mechanism to link the hearer's attention to a new and unfamiliar entity and thus (unintentionally) draws attention to and facilitates the comprehension of that referent. The Expectancy Hypothesis (Arnold 2010) suggests that accessibility is influenced by structural factors (i.e. the way in which a referent was introduced) as well as by non-linguistics ones (e.g. disfluency). This evidence challenges studies that consider old information more accessible than new information. In other words, both new and old

information can be accessible, when expected. In the following Section I will propose the extension of previous accounts on accessibility by introducing the notion of discourse structuring potential (DSP) to account for the discourse behaviour of certain referents introduced by means of indefinite noun phrases.

### **2.4.2 The discourse structuring potential**

Up to this point I discussed several approaches that accounted for the successful use of different types of referring expressions. Throughout the discussion, two main lines of research have been highlighted. According to the first, the use of particular types of referring expressions is determined by different heuristics that hold between adjacent sentences. Most investigations showed that multiple factors, as for example recency, givenness, and syntactic prominence determine the felicitous use of pronouns in the immediately following sentence. The second line of research argued that the use of a certain type of referring expression (i.e. a pronoun) is only a by-product of a more general strategy to maintain coherence in a discourse segment. According to Centring Theory, for example, an ideal discourse is one in which the discourse topic does not shift abruptly from one sentence to the other and in which pronouns will be preferably used for the continued topic constituent. The topic constituent of a given discourse is determined based on the factors that predict the rate of pronominalization in the former model (e.g. subject prominence, semantic prominence). In sum, different structural factors were shown to impact the rate of subsequent pronominalization and the likelihood of subsequent mention. Importantly, however, pronominalization and the next-mention bias do not overlap at all times in the discourse. That means that not all referents, which are likely to be picked up in the subsequent discourse will be pronominalized.

In the remainder of this dissertation I explore the impact upon the subsequent discourse of referents realized in less-prominent positions. More precisely, I will focus on referents associated with a low degree of accessibility, namely indefinite noun phrases introduced in direct object position, encoded as the Patient argument of a transitive event, rather than on subject referents realized as the semantic Agent. Based on the results of the three experimental studies that will be introduced in Chapters 3-5, I will show that the type of referring expression used (i.e. marked indefinite vs. unmarked indefinite) gives structure to the subsequent discourse on a non-local level. More concretely, marked indefinite noun

phrases will be shown to be more prone to shift the discourse topic and to show higher rates of subsequent mention compared to their unmarked counterparts. Furthermore, I will show that these discourse structuring effects can be observed in non-local contexts (i.e. between non-adjacent sentences). That is, the factor *type of referring expression* does not impact the following discourse in a clause-by-clause manner (i.e. between two adjacent sentences) like the accessibility factors discussed above, but rather shows *non-local effects* on larger discourse units, two or more sentences after being introduced in the discourse. I will argue that marked indefinite noun phrases are used as discourse structuring devices, rather than as a means to reflect the accessibility of their associated referents. In other words, the accessibility and the discourse structuring potential of a linguistic item are two factors that should be kept apart.

The view on accessibility advocated for in this dissertation combines backward-looking and forward-looking perspectives as well, but nevertheless differs from Centering Theory and the Expectancy Hypothesis in at least three main aspects: First, while Centering Theory computes the next-mention probability of a referent based on local, clause-by-clause biases (i.e. between two adjacent sentences), the model adopted here investigates comprehenders' predictive expectancies not only inter-sententially, but non-locally, over more sentences as well. Second, while previous accounts mostly focused on prominent referents (i.e. the "backward-looking" and the "preferred center" in Centering Theory terminology), the present analysis investigates comprehenders' expectations about less prominent referents as well (i.e., referents mentioned in non-prominent syntactic or semantic positions). And, finally, the present study investigates multiple effects of non-prominent referents on the subsequent discourse from a cross-linguistics point of view by investigating data from English, German and Romanian.

In sum, we can make an essential distinction between two main types of expectations, which are triggered by different factors and at different levels of representation. The first type of expectation, which is generally discussed in the literature on pronoun resolution in terms of accessibility, can be derived from structural factors such as a particular syntactic position and a certain argument structure. Based on such factors, hearers make probabilistic expectations about those referents that are more probable to be mentioned next and more prone to be pronominalized. The second type of expectation, which is important for the present analysis, comes about by discourse structuring devices such as the three types of

indefinite noun phrases discussed in this dissertation. Such linguistic devices target larger discourse units in that they indicate which referent will be mentioned over larger discourse segments. It is important to note that the discourse structuring potential is not to be confounded with accessibility or activation, as different types of referring expressions that are associated with a high discourse structuring potential are not necessarily more activated or accessible compared to referring expressions associated that are not associated with a discourse structuring potential. Moreover, I argue that the discourse structuring potential is a property defining indefinite noun phrases, but that at the same time it is a fine-grained feature that differentiates between several sub-types of such expressions. For example, a referent mentioned for the first time by means of an indefinite noun phrase is more or less prone to be mentioned again in the following discourse. I will show that the three languages under investigation do not differ with respect to the expectancy types they employ, as both sentence-by-sentence effects and discourse structuring effects are realized. What is different is the formal marking employed by each language for each function and the relative weighting that languages assign to each of them. The main suggestion resulting from the presented evidence is that expectancy operates in parallel at different levels of representation. In what follows, I will briefly describe the experimental methodology used in the next three Chapters to test the discourse structuring potential of the indefinites central to this dissertation.

### **2.4.3 The discourse analysis method**

In Section 2.2.2, I discussed several structural factors that contribute to the overall accessibility or discourse prominence of a given referent. Moreover, most studies insist upon the discourse history of a referent in order to determine its status in the actual discourse segment. My aim is to analyse the way in which indefinite noun phrases, which lack an antecedent and which introduce a new referent in the discourse, give structure to the subsequent discourse structure.

To investigate the effects on the subsequent discourse of indefinite noun phrases in English, German and Romanian, I conducted a discourse analysis for each of the three languages. As it is complicated to find examples in which the marked and unmarked forms appear in parallel sentences (i.e. sentences in which all factors are controlled for, except for

the type of referring expression used), I will use a sentence-continuation experiment to test the discourse effects of the indefinite noun phrases marked by indefinite *this* in English, *so'n* in German and *pe* in Romanian. The written discourses were elicited from adult participants, who were presented stories consisting of two or three sentences for which they were asked to provide five meaningful sentence-continuations. Investigating written discourses over spontaneous speech situations has the advantage of excluding other factors, such as gestures, eye gaze or disfluencies, which have been shown to alter the effects of linguistic information (e.g. Arnold and Tanenhaus 2007). This off-line story continuation task was adopted in several experimental studies (Givón 1981, Garrod and Sanford 1988, Gernsbacher and Shroyer 1989, Arnold 1998, among many others), as it has the advantage of combining both comprehension and production processes involved in the process of referring. Participants' task was to produce five new continuation sentences, but it required them to comprehend the stories before providing a continuation. Of particular importance is that participants' responses were made on the basis of the mental representations they developed while reading the story. Accordingly, their responses reflect their predictive expectations with respect to the initial multi-sentence stories. This experimental method permitted me to explore multiple aspects. First, participants were not restricted to the type of continuation they added. Note that compared to other studies using a sentence-continuation task, participants' continuations in the studies reported in the next chapters were not influenced by a pronoun or connective prompt at the end of the experimental items. This freedom meant that their responses provided information about the natural way in which they took up the referent again, which referent they mentioned and how frequently. Second, participants' responses were more natural, as they were asked to provide five sentence-continuations for each initial story instead of only one. This task was particularly important, as the main aim of the studies was to determine the discourse effects of referents realized in less prominent positions, which are expected to become evident only at a later stage in the discourse (i.e. after the first sentence-continuation on, due to the attested strong subject/first mention-bias).

The experimental designs and the analyses in each language were similar. I will first discuss the experimental design and then describe the analysis procedures. As already stated, each experimental item consisted of a story of two or three sentences. I only manipulated the type of indefinite noun phrase, which resulted in two conditions. In one condition the target items are headed by the indefinite marker (i.e. indefinite *this* in

English, *so'n* in German and *pe* in Romanian) and in the other condition the target items remain unmarked (i.e. they are realized by means of simple indefinite noun phrases), as illustrated in (22) for Romanian.

- (22) (a) Marked indefinite [*pe*-marked indefinite noun phrase]

*Anul trecut când am fost la mare am cunoscut un salvamar. Era tot timpul activ. La sfârșitul sejurului meu, a salvat-o **pe o fată** de la înec.*

‘Last year when I was at the seaside I met a lifeguard. He was very active all day long. At the end of my stay there, he saved PE-a girl from drowning.’

- (b) Unmarked indefinite [simple indefinite noun phrase]

*Anul trecut când am fost la mare am cunoscut un salvamar. Era tot timpul activ. La sfârșitul sejurului meu, a salvat **o fată** de la înec.*

‘Last year when I was at the seaside I met a lifeguard. He was very active all day long. At the end of my stay there, he saved a girl from drowning.’

Note that the target items in all test sentences are realized in direct object position for several reasons. Firstly, *pe*-marking of indefinite noun phrases in Romanian is possible only for direct objects. Indefinite noun phrases preceded by *so'n* and indefinite *this* in German and English respectively can be sometimes used in subject position as well, even though this grammatical position is dispreferred. In order to compare the findings from the three languages under investigation, I constructed all target referents in direct object position. Second, another advantage for the coding of the target indefinite noun phrases as direct objects is that we can compare their accessibility or prominence to that of the subject referents of the sentence. As subjects are generally considered to be more accessible than referents realized in other syntactic positions, it is interesting to see in what way the marked indefinite noun phrases realized in direct object position will influence the overall discourse structure and accessibility of the referents.

Besides the manipulated form of the indefinite noun phrases occurring in direct object position, the experimental were constructed in a similar way. The first sentence set the scene and introduced a global topic that the story was about (i.e. *un salvamar*, ‘a lifeguard’ in example (22)). If the experimental item consisted of three sentences, then the second sentence picked up the topic of the preceding sentence. The third and last sentence introduced the target referent for the first time, which was realized as a marked or unmarked indefinite noun phrase.

Two independent judges analysed the five continuation sentences (main and subordinate ones), provided by each participant, coding three aspects of the indefinite noun phrases realized in direct object position: (i) referential persistence (i.e. the frequency of remention of the referent), (ii) topic shift potential and (iii) the type of the first anaphoric expression used to pick up the referent. In what follows, I will briefly discuss the coding methods and the predictions for these aspects.

### 2.4.3.1 Referential persistence

The first aspect investigated is the referential persistence of the stories' referents, which comes very close to Givón's (1983) measure of referential continuity. This aspect indicates how frequently an entity will remain in the subsequent discourse after it was introduced for the same time. Recall that most investigations into the accessibility or prominence of referents investigated whether a referent is likely to be mentioned in the immediately following sentence, as accessible referents have been shown to display this preference. While previous studies investigated local next-mention biases, in this study I will additionally test how often a referent introduced in the discourse is likely to occur in the next five sentences after it was introduced. To this end, the total number of occurrences of subject and direct object referents in the five continuation sentences will be determined. In calculating the persistence or continuity of a referent in the subsequent discourse, I will consider all anaphoric mentions of that referent in a following matrix or subordinate clause. It was important to decide beforehand, which noun phrases will be coded as an instance of that referent in the subsequent discourse. This aspect was crucial for the third factor pertaining to the discourse structuring potential of the referents as well, i.e. the type of referring expression used to pick up the referent for the first time in the subsequent discourse. The full set of codings for the different types of noun phrases is given in Table 2.5. Note that non-referentially used noun phrases were excluded from the present analysis. Other noun phrases that were not coded include bare noun phrases, generic noun phrases, bridged or linked noun phrases, appositives and predicate nominatives. Furthermore, the analysis of the English, German and Romanian data was similar, except for some minor differences. Compared to the English and German reference systems, Romanian allows for both pro-dropped subjects and clitics in direct and indirect object position. For the English and German referential system it is still a debate whether the morphologically unrealized subject position in the second conjunct of a sentence like *Paul came in and Ø sat down*

should be considered an anaphoric element, or not. In the present study, I coded the deleted subject referent in such examples as an anaphoric instance of that referent, without adopting a particular syntactic analysis for that position (but see Van Valin (1986) or Brandner and Fanselow (1992) for a discussion of several approaches to subject lacking in finite structures in English).

Table 2.5: Codings for noun phrase types

Coding	Details	Examples
zero	Deleted arguments in a conjoined phrase	Paul came in and $\emptyset$ sat down.
pro	Omission of the subject (verb morphology gives person and number information)	$\emptyset$ merge la universitate. [Romanian] go.3P.SG to the university. '(S)he goes to the university.'
pronoun	Personal pronouns, demonstratives, possessive pronouns, reflexive pronouns	He; she; it; this; that; his; her; himself; herself.
clitic	Weak pronouns with specialized morphological inflectional forms for accusative and dative. Clitics can occur by themselves or with a direct or indirect object.	Paul o [...] vede. [Romanian] Paul cl.3P.SG.ACC.FEM see 'Paul sees her'.
proper name		Paul Smith; Mary
indefinite pronoun	When used as an article word or pronoun, quantifiers.	Both; all; some
definite (modified) NP		The (big) man, the (big) house
indefinite (modified) NP		A (big) man, a (big) house

After discussing the coding methods used for referential persistence, I will now turn to the main predictions for referential persistence, which are assumed to hold for all languages investigated.

### Overall referential persistence

Based on the findings discussed in the previous Sections, the prediction is that referents realized in prominent semantic and syntactic positions will outrank less prominent referents in referential persistence. More concretely, referents realized as definite noun phrases in subject position, bearing the semantic Agent role will outrank in referential

persistence those referents realized as indefinite noun phrases in direct object position bearing the semantic Patient role.

### **Referential persistence of marked indefinite noun phrases**

With respect to the referential persistence of marked indefinites (i.e. indefinite *this* in English, indefinite *so'n* in German and *pe*-marked indefinites in Romanian) these referents will outrank simple indefinite noun phrases in referential persistence, if they are used as discourse-structuring devices. Thus, under the plausible assumption that prominent referents tend to be picked up more often in discourse (Givón 1983), I expect participants to pick up the marked referents more often in the discourse than their non-marked counterparts. Moreover, with respect to the next-mention bias, the prediction is that referents associated with the marked indefinite noun phrases, will be more prone to be immediately re-mentioned after being introduced in the discourse. On the contrary, if non-marked referents will be picked up, they are not expected to be re-mentioned in the immediately following sentence, but rather later on in the subsequent discourse.

#### **2.4.3.2 Topic shift potential**

Topic shift was the second aspect investigated with respect to the discourse structuring potential of the marked indefinite noun phrases. Before I present further details on the predictions with respect to the topic shift potential of marked indefinites, it is important to stress that the notions of topic and topicality are not used in a unitary way among researchers. While the notion of topic has been traditionally understood as a device that pertains to information structure, most psycholinguistic research considered it a structural factor. Recall, for example, that most studies on accessibility or prominence discussed in the previous Sections did not distinguish between subjects and topics and each instance of subjecthood was treated as an instance of topichood. The overlap between subject and topic found in the psycholinguistic literature might be due to the observation that topics are generally realized as the grammatical subject in English. However, there is convincing cross-linguistic evidence that this does not always have to be the case, as not all topics are restricted to occur in subject position and as some sentences lack topics (e.g. event sentences).

At least since Strawson (1964), the notion of topic has played an important role in the linguistic literature, where it is commonly understood as the referent a particular sentence is about (Reinhart 1982). Along these lines, I will use the term ‘topic’ as an information structural device, which stands for the entity the current sentence makes a predication. I assume that each sentence can have at most one aboutness topic. Following Maki et al. (1999) and subsequent work, I furthermore assume that aboutness topics canonically occur in matrix sentences and are rather less typically used in embedded sentences. For the purposes of the studies presented in Chapters 3-5, we applied several tests to establish whether a particular referent associated with a noun phrase is the aboutness topic, or not. For example, a noun phrase X is the aboutness topic of a sentence that contains it, if: (i) the sentence would be a natural-sounding continuation to the statement “Let me tell you what happened about X”, (ii) the sentence would be a plausible answer to the question “What about X”, or (iii) the sentence could be embedded in a structure as “Concerning X, [...]”.

Summarizing, the term topic employed in this dissertation is crucially distinct from its use in the studies discussed at the beginning of this Chapter. For the sake of simplicity, I consider the first instance in which a target referent realized in direct object position was mentioned as a topic in a matrix clause an instance of topic shift. Referents mentioned in subordinate clauses were not coded for topichood. It is important to note that the topics in the stimulus items in the experiments introduced in Chapters 3-5 were established by introducing one character as the referent the sentence was about. Furthermore, the same referent was the grammatical subject and was realized by means of a proper noun and picked up in the experimental item at least once by means of a pronoun or a zero anaphor. Thus, the first introduced referent was realized in a prominent position in every sentence of the mini-discourse in the experimental items. This referent was not only the aboutness topic, but according to Cowles’s (2003) analysis, also the discourse or global topic of each experimental item. On the contrary, the target referents were mentioned only in the last sentence of each experimental item by means of an indefinite noun phrase in direct object position. This manipulation aimed at making the target referent less prominent and thus less expected to show any effects upon the subsequent discourse. However, if the target referents impact the subsequent discourse despite this manipulation, then the observed discourse structuring effects should be even more convincing. Assuming that marked indefinite noun phrases serve as discourse structuring devices, the predictions concerning their topic shift potential are as follows.

### **Overall topic shift potential**

Given the general tendency for subjects to be more accessible than object referents (Gordon et al. 1993, McDonald and McWhinney 1995, Ariel 1988) and given the observations that naturally occurring discourses tend to continue the topic instead of shifting it (Givón 1983), the overall prediction is that the topic will tend not shift in the subsequent discourse from the prominent referent (i.e. the global topic) to the target referent. However, if the marked referents function as discourse structuring devices, then they will: (i) appear in topic positions earlier in the continuation sentences compared to their unmarked counterparts, but not earlier than the second sentence-continuation and (ii) shift the topic more often than their unmarked counterparts.

#### **2.4.3.3 Type of the first anaphoric expression**

The last aspect investigated with respect to the discourse structuring potential of the marked indefinite noun phrases was the type of referring expression used to pick up for the first time the target referent. For each continuation sentence I looked at whether the referent of the direct object (if present) was picked up by means of one of the types of referring expressions enumerated in Table 2.5 above. I did not analyse the anaphoric chains used to refer back to a referent, as it was not possible to code for different factors that could impact the types of referring expressions used for a referent after the first sentence provided by the participants.

Given the observation that highly accessible referents tend to be picked up by less explicit types of referring expressions (e.g. pronouns), the prediction is that the global topic will be more prone to be realized by reduced types of referring expressions. Moreover, the less prominent target referents headed by *this*, *so'n* and *pe* are expected to be picked up in the first sentence after being introduced in the discourse by more explicit types of referring expressions. That is, if pronoun use is an indication of accessibility or prominence, the prediction will be that less prominent referents will not show high rates of subsequent pronominalization. Furthermore, there will be no difference between the types of referring expressions used to pick up the referents of the marked indefinite noun phrases in comparison to their unmarked counterparts (i.e. simple indefinite noun phrases).

## 2.5 Summary

The main goal of this Chapter was to define some viable methods to capture the (future) discourse structuring potential of referents introduced in the discourse by indefinite noun phrases. The three textual factors discussed in this Chapter, namely (i) referential persistence, (ii) topic-shift potential and (iii) the type of referring expression of the first anaphoric item will be used for my investigations of indefinite descriptions in the following Chapters. In the first three Sections I revised different theories on accessibility and several factors that contribute to the accessibility of referents in terms of rates of subsequent pronominalization. In Section 2.4 I argued that an analysis of accessibility in terms of expectancy is to be preferred over other theories, because it can successfully integrate the observation that non-prominent referents can be more predictable in the upcoming discourse. Finally, I presented the text analysis method I will use in the following Chapters together with the predictions concerning the behaviour of referents associated with different types of indefinite noun phrases. With this general theoretical foundation in place, in the remainder of this dissertation I will bring empirical evidence to show that indefinite noun phrases give structure to the subsequent discourse in different ways. The main results of these studies will show that the use of indefinite *this* in English, indefinite *so'n* in German and *pe*-marking in Romanian are correlated with referential persistence and subsequent topic shift, but not with predictions concerning upcoming reference form. These findings will experimentally sustain the claims advanced in the previous Section, namely that these effects are not by-products of a referent's accessibility status, but that they rather indicate the way speakers organize or give structure to larger discourse segments.

### 3. The discourse contribution of English indefinite *this*

Cross-linguistically, the predominant use of demonstratives is as definite expressions. In this use, their main function is to identify an entity in the (physical) discourse situation or in the addressee's mental representations. The main use of English noun phrases headed by demonstrative *this* is to introduce referents in the discourse, which are identifiable in the discourse situation based on a particular gesture accompanying the referring expression, based on the previous discourse, or on the basis of the amount of knowledge shared by the discourse participants. What is remarkable about English *this*, is the fact that it can be used as an indefinite determiner, despite its primary definite use. On the one hand, it functions as a demonstrative determiner alongside *that*, and, on the other hand, it is a determiner for indefinite noun phrases alongside the simple indefinite article *a(n)* (Perlman 1969, Prince 1981, Maclaran 1982, Wald 1983, among others). In this Chapter I focus on the latter use of *this*, which occurs in colloquial, mainly conversational English, as shown in (1).

- (1) Well, I met *this guy* last night that is a year older than me at a pool Party in San Francisco (Well I live in LA but I was visiting family) and I really liked him. Okay, and there was also *this girl* who was really mean [...]. Look at *this book* he gave me [...].<sup>4</sup> (italics mine)

In example (1), only the first two instances of *this* (i.e. *this guy*, *this girl*) are indefinite, whereas the third *this* in *Look at this book* is an instantiation of the stressed, deictic use of *this*. An analysis of English indefinite *this* is revealing for several reasons. First, it represents a relatively new development of the English language, attested since the late 1930's predominantly in the narrative register (Perlman 1969, Wald 1983). Second, indefinite *this*, as opposed to definite *this*, serves to introduce hearer-new and discourse-new referents. In this use, indefinite *this* does not refer to a physically perceptible referent, nor does it refer back to an antecedent. Third, a body of linguistic research (Prince 1981, Lyons 1999, Ionin 2006, among others) has convincingly shown that indefinite noun phrases headed by *this* tend to have referential meanings (i.e. take wide scope with respect to different operators, are epistemically specific, etc.). And fourth, besides being

---

<sup>4</sup> <http://Section.ihav.net/vb/introductions/i-met-him-last-night-358268.html> [viewed in March 2009]

semantically specific, *this* indefinites have generally been associated with accessible or noteworthy referents (Perlman 1969, Prince 1981, Maclaran 1982, Wald 1983, Wright and Givón 1987, Gernsbacher and Shroyer 1989).

The aim of this Chapter is to show that the behaviour of indefinite *this* cannot be thoroughly accounted for unless the broader discourse context in which this lexical item occurs is taken into consideration. To this end, I extend the analysis of indefinite noun phrases headed by *this* by investigating the subsequent discourse in which they occur. Using the story-continuation method introduced at the end of Chapter 2, I will show that referents headed by indefinite *this* are: (i) referentially more continuous in the subsequent discourse compared to indefinite noun phrases preceded by the simple indefinite determiner *a(n)* (an observation which sustains previous observations about indefinite *this*) and (ii) likely to shift the discourse topic. It will become obvious by the end of this Chapter that indefinite *this* is a linguistic device used to give referential structure to the subsequent discourse, rather than signalling the accessibility of its associated referent.

This Chapter is organised as follows. In Section 3.1, I sketch out the main uses of the demonstrative *this* and show that despite its definite origin, it may be used to introduce brand-new referents in the discourse, alongside the simple indefinite article *a(an)*. In Section 3.2, I discuss the referential and discourse-pragmatic properties of indefinites headed by *this*. More specifically, indefinite *this* noun phrases will be shown to differ from the indefinite article *a(n)*, as they display a strong tendency for referential interpretations. Then I review previous accounts on the “forward-looking” potential of referents preceded by indefinite *this*. In the literature so far (Prince 1981, Gernsbacher and Shroyer 1989, Ionin 2006) it was suggested that *this* introduces accessible, important or noteworthy referents in the discourse, as such referents are likely to be subsequently mentioned. I will argue that the discourse behaviour of indefinite *this* can be better accounted for in terms of its potential to give referential structure to the ensuing discourse, which is reflected in the referential persistence and the topic shift potential of the newly introduced referents. In Section 3.3, I will test the nature of discourse prominence of referents headed by *this* (vs. indefinite *a(n)*) by means of the multi-sentence story continuation experiment already introduced at the end of the second Chapter. The results of the experimental study show that referents headed by *this*: (i) are referentially more continuous in the subsequent discourse than referents headed by *a(n)*, and (ii) show a stronger tendency of becoming the

topic in the following five continuation sentences. Section 3.4 focuses on the basic contribution of demonstratives and of indefinite *this* in particular to focus hearer's attention to a certain referent. Finally, Section 3.5 concludes the Chapter and outlines an enriched analysis for indefinites headed by *this* in terms of their discourse structuring potential.

### 3.1 Forms and uses of demonstratives

The literature distinguishes three main classes of demonstratives: nominal demonstratives, local adverbial demonstratives and verbal demonstratives (Dixon 2003). Nominal demonstratives can be used in combination with a noun to form a noun phrase, or they can form a noun phrase on their own (e.g. *this story* vs. *this*). Local adverbial demonstratives modify a verb and can be used on their own (e.g. *bring it there*), or in combination with a noun that is marked for the location (e.g. *bring it (to the man) there*). Verbal demonstratives can be used as the only verb in the predicate or they can occur with a lexical verb. Verbal demonstratives as for example *Make it like that* are accompanied by gestures. As it will be discussed in the next Chapter, this threefold distinction holds for the German demonstrative expression *so* ('such') as well. Throughout this dissertation I am concerned only with the different uses of nominal demonstratives.

The goal of this Section is to show that English demonstrative *this*, which is predominantly used as a definite referring expression has developed an indefinite use that becomes visible in out-of-the-blue contexts (i.e. in contexts in which the expression does neither involve pointing gestures, nor does it refer back to an antecedent expression). The first part of this Section discusses the four main uses of demonstrative *this* as a definite referring expression, namely the deictic use, the anaphoric/cataphoric use, the discourse deictic use and the recognitional use. The second part of this Section is dedicated to the relatively new and often neglected indefinite use of *this*. Compared to their definite counterparts, noun phrases headed by indefinite *this* will be shown to introduce hearer-new and discourse-new referents, similar to indefinite noun phrases preceded by the simple indefinite article *a(n)*. The final part of this Section discusses the semantic and discourse-pragmatic properties of indefinite *this* noun phrases, which are distinct from those of the simple indefinite article headed by indefinite *a(n)*.

### 3.1.1 Definite uses of demonstrative *this*

Throughout this Chapter, I consider demonstrative noun phrases, or demonstratives those types of referring expressions that consist of a demonstrative expression functioning as a determiner, or as a demonstrative pronoun (representing the noun phrase on its own). Demonstratives represent a closed class of linguistic expressions that are universally distributed (Himmelmann 1996, 1997, Diessel 1999, Dixon 2003). Languages differ with respect to the exact number of demonstrative expressions they developed and with respect to their morphological realization. For example, the English demonstrative system distinguishes two demonstrative expressions, namely *this* and *that*<sup>5</sup>, with their corresponding forms inflected for plural. Demonstrative expressions are predominantly used as *definite* referring expressions, and their main function is to identify an entity in the (physical) discourse situation or in the addressee's mental representation. Their four major uses are the deictic use, the anaphoric/cataphoric use, the discourse deictic use and the recognitional use (Haspelmath 1997, Lyons 1999). While the deictic/non-deictic use has been considered the most basic one, there seems to be no consensus about the status and characteristics of other demonstrative uses. The aim of this Section is not to discuss the exact number of demonstrative uses, but rather to illustrate the special status of indefinite *this*. To this end, I will briefly discuss the other uses of the demonstrative and the characteristics that have been put forth to differentiate between them.

When they refer deictically, demonstratives are often accompanied by a physical gesture (e.g. pointing, gazing, moving an object, etc.), which serves to indicate to the addressee the speaker's intended referent (Bühler 1934, Diessel 1999, Enfield 2003, Levinson 2004). In their deictic or exophoric use, demonstratives refer directly to a referent in the perceivable, extra-linguistic context, as illustrated in (2). The referent associated with the definite noun phrase *this man* can be either new or old to the discourse or to the addressee, however, the intended concrete referent is identifiable on the basis of the descriptive material (i.e. *man*) and of the accompanying pointing gesture, which locates the referent in the indicated context (i.e. in the picture on the desk).

---

<sup>5</sup> Several studies (Halliday and Hasan 1976, Strauss 2002, among others) include *it* as a non-specific lexical item, or as a signal of low focus, in addition to the specific or high/medium demonstratives *this* and *that*. However, given that the present study does not aim to investigate the English demonstrative system as a whole, I will not go into the various debates on which morphemes should be considered demonstratives and which not.

- (2) (Pointing to a picture on the desk)

*This man* stole my heart.

The second main use of demonstratives is the anaphoric/cataphoric use. Compared to the deictic use, anaphoric or cataphoric demonstratives represent discourse-old and hearer-old information, as they corefer with a referent in the previous or subsequent discourse, respectively. Accordingly, the interpretation of the endophoric demonstrative expression is dependent on the antecedent expression within the text. In example (3) below, the referent headed by demonstrative *this* refers anaphorically to its antecedent (i.e. *a cat*) introduced in the first sentence.

- (3) I saw a cat<sub>1</sub> in the park two days ago. This cat<sub>1</sub> was still there last night.

The discourse deictic use of demonstratives is a particular case of the anaphoric use. In this role, the demonstrative expression refers back to a coreferring antecedent, which expresses an entire proposition, rather than being associated with an individual referent, like in the anaphoric/cataphoric and deictic use. For example, the demonstrative *this story* in (4) refers back to the entire story introduced in the first sentence, which describes the fact that Paul's parents met 35 years ago. Note, however, that the discourse segment referred to by the demonstrative is in the immediately adjacent sentence.

- (4) Paul told me about the way his parents met 35 years ago on the outskirts of Los Angeles. It was so romantic. After finishing this story, he said he had to go home.

The recognitional use of demonstrative *this* is the most controversial use, as it subsumes different interrelated uses of the demonstrative. Compared to the three other uses of demonstrative *this* presented above, in the recognitional use, the referent associated with the referring expression is identified based on the amount of shared knowledge between the discourse participants and does not rely on a deictic gesture or on a previously introduced antecedent expression or segment (Auer 1981, Himmelmann 1996). In example (5), the intended referent is hearer-old, but it is presumably not part of the hearer's short-term memory representation. In other words, the referent introduced is familiar to the hearer based on the discourse history he shares with the speaker and not based on global, non-interactional, non-specific knowledge. The difference in the type of shared information

distinguishes between the use of the definite article for the latter use and of the demonstrative for the former.

(5) I met this neighbour of yours in the park.

Auer (1981) points out that another central characteristic of the recognitional use is the uncertainty on speaker's side about whether the hearer will be able to properly identify the intended referent based on the descriptive material provided by the noun phrase headed by the demonstrative. The demonstrative indicates to the hearer that the referent could be elaborated upon if needed.

Lakoff (1974), Bowdle and Ward (1995), and Wolter (2006) distinguish an additional use of demonstrative *this* as an emotional marker, signalling the speaker's involvement with the event. In this Chapter, I take the emotional use of demonstratives as a subclass of the recognitional use. Similar to the standard recognitional use, emotionally used demonstratives introduce referents familiar to the hearer based on the specific, personalized information he shares with the speaker. The difference between the two uses rests upon the fact that emotionally used demonstratives express the speaker's positive or negative attitude towards a particular referent. In this use, the core function of *this* is to express emotional closeness between speaker and hearer, "fostering a feeling of shared sentiment and experience" (Lakoff 1974). Emotional *this* evokes speaker-hearer solidarity that strongly correlates with evaluativity, which is found with exclamative constructions and words. The fact that the speaker assumes that the addressee shares the information conveyed by the demonstrative noun phrase is indicated by the accent on the demonstrative.

In sum, the classification of the major uses of demonstratives presented above rests on several characteristics that pertain to the interaction between the discourse participants with respect to: (i) information status (discourse-new vs. discourse-old referents); (ii) hearer status (shared/ hearer-old vs. unshared/hearer-new information); and (iii) discourse function (identifying vs. referent tracking).

### 3.1.2 The indefinite use of demonstrative *this*

In addition to the four main uses of demonstratives mentioned so far, English demonstrative *this* developed an additional use in out-of-the-blue contexts, which represents the focus of this Chapter. This relatively new use of *this* has been discussed in the literature under names as different as: “new-this” (Prince 1981, Wald 1983), “non-phoric use of this” (Halliday and Hasan 1987), “specific indefinite-this” (Ionin 2006), “referential this” (Gundel 1993), “cataphoric this” (Halliday and Hasan 1976, Gernsbacher and Shroyer 1989), among others. I will adopt Diessel’s (1999) term “indefinite *this*” to refer to this new use of the demonstrative. In its indefinite use, the noun phrase headed by demonstrative *this* does not represent shared knowledge between speaker and hearer, as in all other uses of the demonstratives. Rather, the demonstrative expression represents discourse-new and hearer-new information and does not bear main stress. Consider for example sentence (6) in which the noun phrase *this woman* does not corefer with a previously introduced referring expression, it is not accompanied by a pointing gesture and does not contain an evaluative comment on the speaker’s side either.

- (6) There is this woman in my neighbourhood who is over 100 years old.

All studies investigating the use of demonstrative *this* have recognized a striking resemblance between the use of *this* in out-of-the-blue or indefinite contexts and the regular use of the simple indefinite article *a(n)*. Both determiners are used to introduce hearer-new and discourse-new referents. Prince (1981) was one of the first to argue in favour of an analysis of demonstrative *this* in out-of-the-blue contexts in terms of indefiniteness. She proposed several tests to show that *this* is an indefinite description, rather than a definite one. A first test used by Prince is *replaceability* and it illustrates that indefinite *this* can be replaced by the indefinite article *a(an)*, but not by the simple definite article, as illustrated in (7).

- (7) (a) I work in electronic and auto shows. Companies hire me to stay in their booth and talk about products. I have this speech to tell.  
 (b) [...] \*I have the speech to tell.  
 (c) [...] I have a speech to tell. (Prince 1981: 233)

Furthermore, the use of indefinite *this* was shown to be felicitous in existential-*there* sentences, as exemplified in (8a), on a par with the indefinite article *a(n)* (e.g. 8c), and unlike the definite article *the*, as illustrated in (8b).

- (8) (a) When I was a freshman there was this kid named [...].  
 (b) When I was a freshman there was \*the kid named [...].  
 (c) When I was a freshman there was a kid named [...].

Another evidence in favour of an indefinite treatment of *this* comes from the fact that it can precede ‘warmth’-nominals (as defined by Lakoff 1974: 347), as in (9a), while expressions headed by the definite article, as witnessed in (9b) are excluded from such contexts. In example (9a), the presence of *this* seems to make the narrative more palpable for the addressee than the indefinite article in (9c).

- (9) (a) He hugged her with this unbelievable ardour.  
 (b) \*He hugged her with the unbelievable ardour.  
 (c) He hugged her with an unbelievable ardour.

Fourth, the possibility to occur in existential-*have* sentences (Partee 1999) as witnessed in (10a) below, is another argument for an indefinite analysis of *this*.

- (10) (a) Linda has this noisy cousin who lives with her in the same house.  
 (b) \*Linda has the noisy cousin who lives with her in the same house.  
 (c) Linda has a noisy cousin who lives with her in the same house.

Fifth, indefinite *this* cannot combine with noun phrases that represent typical definite descriptions implying uniqueness, as illustrated in (11).

- (11) (a) I talked to \*this mother of my friend Sam, she is really nice.  
 (b) I wanted to meet \*this tallest man at the reunion.

(adapted from Ionin 2006: 192)

Moreover, as noted by Maclaran (1982), stressed *this* cannot receive an indefinite interpretation, as exemplified in (12). In this use, the noun phrase headed by *this* can only be interpreted as a deictic expression.

(12) I met *THIS* girl at the party.

Lastly, demonstrative reinforcement with relative clauses is possible in English. However, restrictive relative clauses are only compatible with an indefinite reading of the demonstrative *this*, as illustrated in (13a). The same observation holds for reduced relative clauses, as seen in (13b), which bear a restrictive reading only if the demonstrative noun phrase they modify is interpreted restrictively. The demonstratives in (13) may appear with nonrestrictive (reduced) relative clauses, if they are used deictically (Perlman 1969, Bernstein 1997).

(13) (a) This guy that I met in Cologne [...].  
(b) This guy organizing the conference [...].

The tests presented above have been used in the literature to underline the indefinite status of *this* in out-of-the-blue contexts (i.e. in contexts in which the noun phrase headed by *this* neither does involve a pointing gesture, nor does it refer back to an antecedent expression).

Summarizing, besides its regular uses as a deictic, anaphoric/cataphoric, discourse deictic and recognitional expression, *this* can be used as an indefinite as well. Table 3.1 summarizes the main uses of demonstrative *this* and contains additional information about the discourse and hearer status tight to each of these uses. In its indefinite use, *this* was shown to introduce discourse-new and hearer-new entities, as it is not used deictically, anaphorically or affectively.

Table 3.1: Main uses and characteristics of English demonstrative *this*

Demonstrative uses	Information status	Hearer status	Further properties
deictic	discourse-new/ discourse-old	hearer- new	Points to a referent in the perceivable situation
anaphoric/ cataphoric	discourse-old	hearer- old	Points back/forward to an antecedent
discourse deictic	discourse-old	hearer- old	Points back to the previous discourse (proposition)
recognitional	discourse-new	hearer- old	Represents shared information between the discourse participants
indefinite	discourse-new	hearer- new	Discourse prominent in the following discourse

We have seen so far that indefinite *this* resembles the behaviour of simple indefinite noun phrases headed by *a(n)*, as it passes several traditional indefiniteness tests. Two main questions arise from this observation: (i) whether indefinite *this* and indefinite *a(n)* differ with respect to their referential properties at sentence level or with respect to their discourse function, and, if so, (ii) what is special about the demonstrative determiner *this* that it fulfils these functions? In the remainder of this Chapter, I will discuss these questions in turn. First, I will focus on the referential properties of indefinite *this* in sentences with and without operators. The picture that emerges by the end of the next two Sections is that, while the simple indefinite article is underspecified with respect to its referential and discourse properties, indefinite *this* shows a strong tendency for referential readings. Then, referents associated with indefinite *this* are more likely to be continued in the subsequent discourse compared to the referents of simple indefinites. I will present the results of a sentence-continuation study investigating the discourse contribution of indefinites headed by indefinite *this*. At the end of this Chapter, I will review several arguments against the popular analysis of demonstratives in terms of (spatial) distance and propose an analysis of indefinite *this* as a means to draw the attention of the hearer towards a particular referent.

### 3.2 Referential and discourse-pragmatic properties of indefinite *this*

In this Section, I discuss several referential properties that distinguish indefinites headed by *this* from those headed by the simple indefinite article *a(n)*. In the first part I show that indefinite *this* displays a strong preference for referential readings (Prince 1981, Ionin 2006, Wespel 2007). In the second part of this Section I will discuss the discourse-pragmatic contribution of indefinite *this* by taking into account the larger discourse context in which this type of referring expression occurs.

#### 3.2.1 Referential properties of indefinite *this*

Previous studies investigating the behaviour of indefinites in opaque contexts (Prince 1981, Ionin 2006) have shown that indefinites headed by *this* are interpreted independently of other operators (i.e. their value does not vary depending on the presence of other scope-taking elements in the sentence), as shown in (14a) and (15a). On the contrary, in examples (14b) and (15b), the indefinites preceded by the simple indefinite article *a(n)* can be interpreted either under or above the same scope-taking element in the sentence. Given the contrast between an existential reading and one that does not entail existentially, the indefinites headed by *this* (14a) and (15a) are compatible with an existential entailment (i.e. Reading<sub>1</sub>), whereas the simple indefinites in (14b) and (15b) are ambiguous between an existential and a non-existential entailment, even though the latter (i.e. Reading<sub>2</sub>) seems to be preferred<sup>6</sup>. In other words, indefinite *this* is not compatible with the cancellation of the existential entailment.

(14) *Intensional operators*

- (a) Eva wants to watch *this movie about Eliade*.

Reading<sub>1</sub>: There is a movie about Eliade. [existential entailment]

- (b) Eva wants to watch *a movie about Eliade*.

Reading<sub>1</sub>: There is a movie about Eliade. [existential entailment]

Reading<sub>2</sub>: There is no such movie about Eliade. [non-existential entailment]

<sup>6</sup> One of the experiments presented in Chapter 4 investigates the readings of *so'n*-indefinites and *ein*-indefinites in German containing two different sentence operators. The findings of that study indicate that instead of showing a true ambiguity between a referentially specific interpretation and a non-specific one, *ein*-indefinites display a clear preference for referentially non-specific interpretations (i.e. non-existential entailment), even when they are further modified by an adjective (for similar considerations, see Ionin 2010).

(15) *Modal operators*

- (a) Eva must recite
- this poem by Eminescu*
- .

Reading<sub>1</sub>: There is such a poem by Eminescu. [existential entailment]

- (b) Eva must recite
- a poem by Eminescu*
- .

Reading<sub>1</sub>: There is such a poem by Eminescu. [existential entailment]Reading<sub>2</sub>: There is no such poem by Eminescu. [non-existential entailment]

Next, *this* indefinites are similar to simple indefinites headed by *a(n)*, as both can escape scope islands such as *that*-clauses and *if*-clauses, receiving long-distance interpretations. Accordingly, sentences (16a) and (16b) are compatible with Reading<sub>1</sub>. The difference between the two sentences is that only the indefinite headed by *this* in (16a) cannot scope within the *that*-clause, which results in dispreference of Reading<sub>2</sub>.

(16) *Scope island escaping*

- (a) Eva thinks that
- this man*
- will fall asleep.

- (b) Eva thinks that
- a man*
- will fall asleep.

Reading<sub>1</sub>: There is someone who is a man such that Eva thinks he will fall asleep. [wide scope reading]Reading<sub>2</sub>: Eva thinks that there is someone who is a man and that he will fall asleep. [narrow scope reading]

Similarly, in a context containing extensional universal quantifiers as (17), the noun phrase *this president* in (17a) refers to a particular president that every girl listened to (i.e. Reading<sub>1</sub>). On the contrary, the referent of the simple indefinite *a(n)* in (17b) is not fixed to a particular entity and may covary with the subject noun phrase *every girl* and is thus compatible with both Reading<sub>1</sub>, in which the president is the same individual for every girl, and Reading<sub>2</sub>, in which each of the girls possibly listened to different presidents (e.g. Maria listened to Chirac, Ellen to Obama, and so on).

(17) *Extensional universal quantifiers*

- (a) Every girl had listened to
- this president*
- .

Reading<sub>1</sub>: There is only one president. [no covariation with the girls]

- (b) Every girl listened to *a president*.  
 Reading<sub>1</sub>: There is only one president. [no covariation with the girls]  
 Reading<sub>2</sub>: There are possibly different presidents.  
 [covariation with the girls]

Much literature (Farkas 1981, Kratzer 1998, among others) has showed that indefinite noun phrases display, besides narrow and wide scope readings, intermediate readings that arise whenever the indefinite scopes out of the scope island, but not out of higher quantifiers. Contrary to the behaviour of indefinites headed by *a(n)*, *this* indefinites seem to block intermediate readings. In (18a) but not (18b), the referent of the noun phrase *this teacher* does not covary with the referent of the students, thus only Reading<sub>2</sub> is possible.

(18) *Intermediate scope readings*

- (a) Every student read every book that this teacher recommended.  
 (b) Every student read every book that a teacher recommended.  
 (adapted from Ionin 2006: 202)

Reading<sub>1</sub>: For every student there is a *possibly different* teacher such that the student read every book the teacher recommended.

Reading<sub>2</sub>: For every student there is possibly the same teacher such that the student read every book the teacher recommended.

Besides plain scopal readings, indefinites headed by the simple indefinite article bear functional readings in certain contexts (i.e. in the context of functional relative clauses (Sharvit 1997), scope phenomena (Schwarz 2001) and *wh*-questions (Chierchia 1993, Krifka 2001)). Consider the sentences in (19), adapted from Hintikka (1986), for which three different continuations (i.e. Reading<sub>1</sub>-Reading<sub>3</sub>) are given. Under Reading<sub>1</sub>, which makes an individual reference, both indefinites in (19a) and (19b) are interpreted as if there was a certain woman (e.g. the Queen) such that every true Englishman loves her. The second possible reading is that in which *a woman* is interpreted as a function of *every true Englishmen* and which could be paraphrased as: “Every true Englishman adores his sister”. This reading holds for the use of *a(n)* heading the noun phrase, as illustrated in (19b). The functional dependency reading (e.g. Reading<sub>2</sub>) becomes unavailable when the simple indefinite article is changed for indefinite *this*.

(19) *Functional readings*

- (a) Every true Englishman adores *this* woman.  
 (b) Every true Englishman adores a woman.  
     Reading<sub>1</sub>: The Queen. [individual reading]  
     Reading<sub>2</sub>: His sister. [functional reading]  
     Reading<sub>3</sub>: John adores his mum, Paul his sister, etc. [pair-list reading]

The same holds for a reading in which pairs of Englishmen-women are listed (i.e. Reading<sub>3</sub>), which is available for the simple indefinite article, but not for the parallel indefinite *this* example given in (19a)<sup>7</sup>.

In transparent and neutral contexts (i.e. in contexts which do not contain any operators), indefinites headed by *this* are interpreted as referring to a particular individual “known” by the speaker. More concretely, an analysis of the noun phrase *this student* in (20a) in terms of epistemic specificity assumes that the speaker of the sentence must have a particular individual in mind in order to be able to introduce the man in the discourse in this way (Kasher and Gabbay 1976, Fodor and Sag 1982, Kratzer 1998, van Rooy 1999).

- (20) (a) This student in the Syntax class cheated on the exam.  
 (b) A student in the Syntax class cheated on the exam.  
     Reading<sub>1</sub>: His name is Smith.  
     Reading<sub>2</sub>: We still have to find out who he is.

On the contrary, the indefinite *a student* in (20b) is ambiguous between a reading which involves speaker’s knowledge about the referent (i.e. Reading<sub>1</sub>) and one in which the identity of the referent is unknown to the speaker (i.e. Reading<sub>2</sub>).

The interpretation of specific indefinites in transparent contexts in terms of “speaker knowledge” alone is insufficient and untenable to explain the exact type of specificity involved in different contexts. Haspelmath (1997: 46) notes that there are many languages which morphologically mark indefinite noun phrases as ‘unknown to the speaker’, but this

<sup>7</sup> But see Hawthorne and Manley (in press), which consider that indefinite *this* is compatible with functional readings.

does not mean that the indefinite in question can be interpreted only in a non-specific way. A case in point represent noun phrases headed by indefinite *this*, which are embedded under verbs of propositional attitude (e.g. *assume, believe, dream, expect, fear, hope, intend, say, suspect and think*, Karttunen 1973b: 1974) and which, unexpectedly, sometimes reflect the knowledge status of an agent present in the sentence, other than the speaker. Prince (1981) shows that in (21), the referent of the indefinite *this* is rather tied to the referent of the matrix sentence (i.e. John) than to the speaker.

(21) John dreamt that he was in this Eskimo restaurant. (Prince 1981: 241)

The notion of specificity anchored to the speaker can be extended to encompass more contexts, not being limited to a function saying ‘what the speaker has in mind.’ The example above can be continued with a sentence like *I don’t know whether the restaurant exists at all*, which underlines the fact that it is not the speaker that has some epistemic knowledge about the Eskimo restaurant, but John, the subject of the matrix sentence. Similarly, consider example (22) in which the speaker refers to a fictional individual, which was part of her friend’s dream world and not part of the speaker’s personal experience.

(22) This didn't happen to me but a close friend. She said that when she was younger, she used to sleep walk. One night she saw this man in her dream, and he was telling her to come outside<sup>8</sup>.

The attitude verb, even though not explicitly expressed in (22) can be deduced from the context. The verb “to say” could be interpreted in the clause in which the indefinite preceded by *this* appears and the clause can be paraphrased as *One night she said she saw this man in her dream*. The noun phrase preceded by *this* is evaluated with respect to the importance the referent associated with it bears for the referent of the matrix subject. In conclusion, in neutral and transparent contexts, indefinite *this* noun phrases can be better accounted for in terms of anchored specificity (in the sense of Kornfilt and von Heusinger 2002), where the anchor can be on the speaker, but also on another referent in the current utterance, or on a non-atomic discourse referent.

---

<sup>8</sup> <http://Section.animepaper.net/source/forum/forum-handler.php?quote=513213> [viewed on 23.09.2010]

To sum up, we have seen so far that indefinite *this* resembles the behaviour of simple indefinite noun phrases headed by *a(n)*, as it passes several traditional indefiniteness tests. However, we cannot account for the distribution of *this* on a par with the distribution of *a(n)*, as the two determiner forms differ with respect to at least two characteristics, namely their referential properties and discourse functions. In contexts which involve sentence operators and in contexts that lack them, noun phrases headed by indefinite *a(n)* take scope freely, while noun phrases preceded by indefinite *this* were shown to display a strong and robust tendency for referential meanings (i.e. specific, wide, non covarying endings)

### 3.2.2 Discourse-pragmatic properties of indefinite *this*

In the previous Section we observed that noun phrases headed by indefinite *this* introduce hearer-new and discourse-new referents that show a strong tendency for referential readings. In addition to these two aspects, indefinite *this* noun phrases have been shown to display additional discourse-pragmatic properties. For example, Fillmore (1977) was one of the first to note that indefinite *this* simultaneously fulfils two functions, which generally do not overlap. First, indefinite *this* introduces hearer-new and discourse-new referents in the discourse establishing a new discourse referent, second, it promotes these referents to a topical or prominent position, as illustrated in (23b).

- (23) (a) I met a friend of yours [*introduction of a new discourse referent*] last night. Well, this guy [*promotion to topic*] told me some pretty interesting things about you. (Fillmore 1997: 105)
- (b) I met this friend of yours [*introduction of a new discourse referent AND promotion to topic*] last night who told me some pretty interesting stories about you.

Please note that the simple indefinite headed by *a(n)* in (23a) serves to introduce discourse-new and hearer-new referents as well, however, the referent is not promoted to a topical position. The prominent position of the referent associated with indefinite *this* is reflected by its immediacy of reference, i.e. by its likelihood of subsequent mention.

Prince (1981) and Gundel (1993) have made similar observations about the definite article, which can be used to introduce a hearer-new and discourse-new referent and simultaneously promote that referent to a prominent position, as (24B) shows.

(24) A: What is wrong with James?

B: The woman he went out with last night was nasty to him.

B': He went out with a woman. The woman was nasty to him.

(adapted from Hawkins 1978: 131)

The noun phrase *the woman* in (24B) can be felicitously used in this context, even if the hearer is not familiar with the identity of the referent (i.e. it is hearer-new and not activated based on shared knowledge). Furthermore, the definite noun phrase in (24B) does not refer to a unique woman (e.g. as the superlative *the tallest woman* would), but is rather interpreted as the contracted version of the two sentences given in (24B'). The fact that the definiteness of the woman represents no semantic requirement on the noun *woman* in context A is underlined by the possible alternative answer given in (25B).

(25) A: What is wrong with James?

B: A woman he went out with last night was nasty to him.

The use of the indefinite article in B's answer in (25) could be interpreted as bearing a partitive reading, implying that *one of the set of women James went out with last night was nasty to him*. Following Gundel et al.'s (1993) line of argumentation, the speaker in (25B) is more informative when using the definite article over the indefinite one in a context in which both determiners are allowed. In (25B) the speaker does not only state that James went out with a woman, but he seems to suggest that the hearer must know the woman in question. Likewise, indefinite *this* is used to signal that its associated referent has particular properties, whereas the simple indefinite *a(n)* is underspecified with respect to this type of information and is more appropriate in a context in which the speaker does not want to elaborate upon the referent in question. Maclaran (1982) notes, however, that the existence of a particular property of a certain referent does not obligatorily trigger the use of indefinite *this*. Consider example (26) in which both *a(n)* and *this* are felicitous, despite the fact that the noun phrases are further modified by a relative clause. The presence of the

simple indefinite article *a(n)* is neutral towards the referential continuation of its associated referent as it does not draw attention to the referent.

(26) I saw *a/ this* cat that reminded me of Bubi.

The referential intention of the speaker to introduce an important or prominent referent in the discourse is highlighted in contexts in which both the definite article and indefinite *this* introduce hearer-new and discourse-new referents, as illustrated in (27) below. The presence of the definite article in (27a) underlines the existence of only one owner of the store, whereas the indefinite *this* in (27b) indicates that there is a particular (one maybe more) property relevant to only one of the possibly more store owners. The crucial observation is that the presence of indefinite *this* does not restrict the set of owners to one individual, as (27a) does, as it allows for (possibly) multiple store owners to which the relevant property may apply.

(27) (a) I need to talk to the owner of the store, Ms. Greene.  
 (b) I need to talk to *this* (one) owner of the store, Ms. Greene. (Ionin 2006: 193)

Ionin (2006) argues that the difference between the use of the definite article in (27a) and the use of *this* in (27b) arises from the different felicity conditions of the two determiners. The presupposition of uniqueness (i.e. a statement that is supposed to hold for both speaker and hearer) forces the definiteness marking in (27a), but since there is no such presupposition in the case of indefinite *this*, as the hearer does not know of the existence of the referent, the noun headed by *this* behaves like an indefinite, despite having a unique instance. Ionin (2006) analyses indefinite *this* by adding a felicity condition on the context, as the speaker considers only his own view of what is important or “noteworthy” about the particular referent. The crucial requirement for this analysis is that the speaker indicates to the hearer the noteworthy property of the referent introduced in this way. The noteworthiness of a referent could be signalled in various ways. For example, many studies have shown (Prince 1981, MacLaran 1982, Ionin 2006, among others) that upon using indefinite *this*, the speaker’s referential intention is mirrored by the subsequent discourse, as the speaker is committed to adding more information about the referent (i.e. attributing properties to the noun phrase introduced by *this*). Interestingly, nouns headed by *this* most often occur in combination with linguistic constructions that serve to immediately give

further information on the referent. Wald (1983) carried out a corpus study into the distribution of indefinite *this* and concluded that these types of referring expressions are often modified by a relative clause, an adnominal modifier, or are left-dislocated, as shown in the examples (28) through (31). It seems that immediacy of reference guarantees the identification of the newly introduced referent headed by indefinite *this*. However, as Ionin (2006) notes, speakers need not use elaborated modifiers so that the hearer can correctly interpret the expression preceded by *this*, but that a conjunction of “noteworthy predicates” (i.e. an unexpected adjective or surprising event) suffices to pick up the referent, as shown in (31).

#### Relative clause

- (28) There is this man who lives upstairs from me who is driving me mad because he jumps rope at 2 a.m. every night. (Ionin 2006: 176)

#### Adnominal modifier

- (29) I read this book about the habits of the underwing moths the other day. (Ionin 2006: 191)

#### Left dislocation

- (30) This fella we had working here, he tried to hide the fact that he was feminine. (Prince 1981: 246)

#### Noteworthy predicates

- (31) John has this weird purple telephone. (Maclaran 1982: 814)

The observation that indefinite *this* introduces hearer-new and discourse-new referents that are more likely to be picked up in the subsequent discourse receives support from findings of naturally occurring discourse analyses and experimental investigations. Prince (1981), for example, found out on the basis of elicited narrations that 242 out of 243 occurrences of the indefinite *this* found in Terkel’s (1974) book *Working* introduced a new referent in the discourse. More importantly, 209 of the 242 brand-new referents (i.e. roughly 86%) headed by indefinite *this* were referred to again in the subsequent text. Wright and Givón (1987) extended Prince’s (1981) by comparing the discourse behaviour of indefinite *this* with that of indefinites headed by *a(n)*. They recorded eight- and nine-year-old children telling one

another stories and found out that children opted for indefinite *this* for referents that they would continue talking about and chose indefinite *a(n)* for referents that they would rather not pick up in the following discourse. The average of referential continuation or persistence was 5.32 for concepts introduced by indefinite *this* and 0.68 for concepts introduced by the simple indefinite article *a(n)*.

Gernsbacher and Shroyer (1989) went a step further and explained the use of indefinite *this* over indefinite *a(n)* in terms of accessibility. They started from the premise that accessible referents should be: (i) mentioned more often in the subsequent discourse (this corresponds to Givón's (1983) textual method referential persistence), (ii) mentioned earlier in the subsequent discourse, and (iii) taken up with more attenuated types of referring expressions (i.e. pronouns) in the subsequent discourse. On the contrary, less accessible referents should be more likely to display low values for all three factors. To test the degree of accessibility associated to referents introduced by indefinite *this* compared to indefinite *a(n)*, Gernsbacher and Shroyer used an auditory sentence-continuation task in which participants were asked to continue 28 narratives within 30 seconds after hearing the last sentence of each target narrative. The last sentence of each target item consisted of a noun phrase which was either headed by indefinite *this* or indefinite *a(n)*. Their findings revealed that on average referents headed by *this*, as opposed to referents headed by *a(n)*, were mentioned more often (i.e. 4.05 vs. 2.76 times) in the immediately following sentences after being introduced in the discourse. Importantly, the difference in average number of mentions of the referents associated with the target noun phrases could not be attributed to an unequal number of continuation sentences provided by the participants for the two conditions. More concretely, participants produced an average of 17.8 continuation sentences per story, irrespectively of the condition (*this* vs. *a(n)*). Another observation was that within the first continuation sentence, 47% of the *this*-referents vs. 34% of the *a(n)*-referents were mentioned again, which confirmed their second prediction with respect to accessible referents being mentioned in the immediately following sentence. Their findings furthermore revealed that indefinites headed by *this*, as opposed to the *a(n)* indefinites, needed less lexical material (i.e. zero anaphors and pronouns) to be mentioned again for the first time in the subsequent discourse. In other words, the rate of subsequent pronominalization was higher for *this* indefinites than for *a(n)* indefinites, which were preferentially picked up by means of definite noun phrases.

On the basis of the above findings, Gernsbacher and Shroyer (1989) concluded that entities introduced by indefinite *this*, as opposed to *a(n)*, are “more accessible (i.e. more focussed, foregrounded or topical) from listeners’ mental representations” and therefore more easily retrievable from short term memory. I will return to Gernsbacher and Shroyer’s experimental results in the last part of this Chapter, when I will discuss the findings of the story continuation experiment conducted by me. I will show that the likelihood of subsequent mention and the type of anaphoric expression used to take up a referent introduced by *this* should be kept apart, as they yield different results and point to different aspects of a referent (see discussion in Chapter 2). At this point, however, it suffices to note that the preferential discourse status occupied by the referents of indefinite *this* is mirrored by their high likelihood of (immediate) subsequent mention.

Besides the high referential persistence that usually characterizes indefinites headed by *this*, it seems that the potential to change the current topic is a manifestation of the discourse structuring potential of *this* as well. Consider the riddle given in example (32) in which *this* heads the new referent introduced in direct object position. After its first mention, the referent associated with indefinite *this* becomes the topic of the immediately following main sentence. In other words, the referent introduced by *this* shifts the current topic in the next sentence and is simultaneously highly persistent in the subsequent discourse.

- (32) I was walking on a bridge and I saw this man. He tipped his hat and drew his can, in this riddle he mentioned his name. What was his name?<sup>9</sup>

In this Section, noun phrases preceded by indefinite *this* were shown to display a strong and robust tendency for referential meanings (i.e. specific, wide, non covarying endings). The picture that emerges by the end of this Section is that, while the simple indefinite article is underspecified with respect to its referential and discourse properties, indefinite *this* shows a robust tendency for referential readings and particular discourse properties. Table 3.2 summarizes the observations made in this Section with respect to the referential behaviour of the two determiners in different contexts.

---

<sup>9</sup> <http://Section.trickyriddles.com/riddle/3904-The-Man/show/answer> [viewed in January 2011]

Table 3.2: Referential properties of indefinite *this* and indefinite *a(n)*

Contexts	referential	island escaping	independent	intermediate	functional	Pair-listing	epistemic
<i>this</i> NP	specific	wide> narrow	no covariation	no	no	no	yes
<i>a(n)</i> NP	specific/ non-specific	narrow> wide	No covariation/ covariation	yes	yes	yes	yes/no

The two types of determiners do not differ only with respect to their referential interpretations in the contexts discussed so far. In this Section, several accounts on the discourse-pragmatic contribution of referents associated with indefinite *this* noun phrases were presented. The common denominator of these studies is the observation that indefinites headed by *this* do not only show particular referential properties at sentence level, but also introduce brand-new referents with special discourse-pragmatic properties. The studies cited above differ, however, with respect to two central aspects. First, there seems to be no consensus about the number and exact nature of the properties associated with the use of indefinite *this*. While almost all studies have observed that indefinites headed by *this* are more likely to be picked up in the immediately following discourse, only a subset of these studies mention other discourse-pragmatic effects of indefinite *this*. For example, the high rate of subsequent pronominalization and the topic shift potential associated with *this* indefinites is attested in few investigations only. Second, the reported discourse-pragmatic effects of *this* indefinites have been considered epiphenomena of the accessibility or noteworthiness of the referents associated with these types of referring expressions. I will discuss these two aspects in turn in the remainder of this Chapter. In Section 3.3 I will report on the results of a sentence-continuation experiment, which tested whether indefinites headed by *this* show effects on the subsequent discourse in terms of (i) immediacy of reference, (ii) shift potential and (iii) likelihood of pronominalization. Then I discuss why an analysis of *this* in terms of a discourse-structuring device is to be preferred over an analysis in terms of importance or accessibility. Section 3.4 offers an answer to the second question formulated at the end of Section 3.1, namely what made the demonstrative *this* particularly suited to be used as a discourse-structuring device? To this end, I will review several arguments against the popular analysis of demonstratives in terms of

distance and propose an alternative analysis of indefinite *this* in terms of a discourse-structuring device.

### **3.3 An experimental investigation on the discourse effects of indefinite *this***

The aims of this Section are twofold. On the basis of a sentence-continuation experiment, I will investigate the nature of the discourse structuring potential of *this* indefinite noun phrases by means of the three textual characteristics already introduced in Chapter 2 (i.e. referential persistence, topic-shift potential and the type of referring expression used to take up the referent for the first time in the subsequent discourse). The current investigation differs from previous experimental studies (e.g. Prince (1981) or Gernsbacher and Shroyer (1989)), as it focuses on the discourse structuring potential of referents headed by indefinite *this*, which are characterized by the semantic feature [+human] and which appear in direct object position. In this way, the findings from this experiment will be comparable to those in Chapters 4 and 5, which investigate the discourse effects of indefinite *so'n* in German and *pe*-marked indefinites in Romanian, respectively. The second aim of this study is to analyse the behaviour of indefinites preceded by the simple indefinite article *a(n)* under the same conditions in order to better filter out the discourse characteristics of the concepts introduced by indefinite *this*. It was already noted that the discourse structuring potential of different linguistic devices is predicted to surface in the participants' continuations in two ways: by referential persistence and topic-shift potential. Alongside these two textual characteristics, I will test the type of anaphoric expression used to refer back to a certain referent as well, as Gernsbacher and Shroyer's (1989) empirical findings showed that referents introduced by indefinite *this* show higher rates of pronominalization compared to simple indefinite noun phrases. In light of the considerations presented in Chapter 2 and in the preceding Section, I will test following three main predictions with respect to the discourse effects of referents introduced by indefinite *this*:

*Prediction 1 (referential immediacy and referential persistence)*

- (a) Referents introduced by noun phrases headed by *this* will be more readily picked up in the subsequent discourse (compared to their simple indefinite counterparts), as indicated by former studies on the discourse-effects on indefinite *this* (see Section 3.2.2).
- (b) In Chapter two, we discussed referential persistence (i.e. the frequency of subsequent mention) as a reflection of a discourse-structuring device. If referents introduced by noun phrases headed by *this* display the potential to give structure to the discourse, then their associated referents will show a high frequency of subsequent mention (referential persistence).

*Prediction 2 (topic-shift potential)*

Referents of noun phrases preceded by indefinite *this* will show the potential to shift the topic in the subsequent discourse (see discussion in Chapter 3.2.2).

*Prediction 3 (referential type of the first anaphoric expression)*

- (a) If the referent associated with indefinite *this* is a marker of accessibility in the sense of Gernsbacher and Shroyer (1989) and other accessibility-based accounts, the likelihood of subsequent mention of this referent by means of a pronoun should be high.
- (b) If, on the contrary, indefinite *this* is a discourse structuring device, then it will not impact the type of the first referring expression used to take up the referent (see discussion in Chapter 2.4.3). In other words, the mere presence of *this* will be neutral towards the first type of anaphoric expression used to refer back to its referent. This prediction derives from recent findings presented by Kehler et al. (2008) and Kaiser (2010), which showed that the expectancy of subsequent mention of a particular referent does not necessarily have to coincide with the type of referring expression used to pick that referent up. For example, the likelihood that a referent will be mentioned in the following discourse does not necessarily mean that this referent is more likely to be realized by a reduced type of referring expression (e.g. a pronoun).

### 3.3.1 The experimental setup

#### *Participants*

Twenty-one speakers of English from the University of Southern California participated in the experiment. It took about 30 minutes to complete a questionnaire.

#### *Design and procedure*

The current experiment takes a comprehension and processing approach to understanding the discourse contribution of indefinite *this*. The methodology used is the multi-sentence-continuation task introduced in Chapter 2. Participants were presented target items of three-sentence stories each and were asked to add five natural sounding, logical continuation sentences to the given sentence fragments. As participants' continuations were not restricted in any way (e.g. by pronouns, sentence connectives, etc.), their responses were made on the basis of the mental representations they developed during reading the stories. As indefinite *this* in English is encountered in conversational, less formal language, the target and filler items were constructed in a colloquial tone.

#### *Materials*

The sentence-continuation experiment included four target stories and six filler stories. Each target story had two versions, one with indefinite *this* and the other with indefinite *a(n)*. The type of determiner (*this* vs. *a(n)*) was manipulated between subjects: each participant either saw all four target items with *this*, or he saw all four target items with *a(n)*. The full list of target items used for each condition is given in Table 3.3. Each story comprised two or three sentences. These sentence fragments set the context of the story and contained an individual reference to a character that was the clearly established topic constituent, as it was mentioned at least once, it was realized in grammatical subject position and was the referent the story was about. In the third and last sentence of each target item, the critical referent was introduced in grammatical direct object position (except for target item 4, TI4) and was realised as an indefinite noun phrase headed by *this* in the *this*-condition, and as an indefinite headed by *a* in the *a(n)*-condition. In order to compare the results from the three experiments reported in this dissertation, both subject

and direct object referents were human<sup>10</sup> and of different genders, except for target item one in which both referents are male.

Table 3.3: List of experimental items (TI) used for both conditions

	<b><i>this</i>-condition</b>	<b><i>a(n)</i>-condition</b>
TI1	<i>Yesterday evening was so warm that James decided to hang out with friends at the local coffee shop. On his way downtown, he saw <b>this kid</b> coming down the street.</i>	<i>Yesterday evening was so warm that James decided to hang out with friends at the local coffee shop. On his way downtown, he saw <b>a kid</b> coming down the street.</i>
TI2	<i>Let me tell you what happened to me some years ago. I met a cool lifeguard at Zuma Beach in Malibu. He was always busy and one day I saw him save <b>this girl</b> from drowning.</i>	<i>Let me tell you what happened to me some years ago. I met a cool lifeguard at Zuma Beach in Malibu. He was always busy and one day I saw him save <b>a girl</b> from drowning.</i>
TI3	<i>It had been only 2 hours but it seemed like the party would never end. Anna was extremely bored. After a while she saw <b>this man</b> sitting alone on a couch next to the fridge.</i>	<i>It had been only 2 hours but it seemed like the party would never end. Anna was extremely bored. After a while she saw <b>a man</b> sitting alone on a couch next to the fridge.</i>
TI4	<i>Alice was rushing around the office yesterday morning, because she had to go to the local high school for a presentation in the afternoon. She shook hands with <b>this man</b> and sat down near the window.</i>	<i>Alice was rushing around the office yesterday morning, because she had to go to the local high school for a presentation in the afternoon. She shook hands with <b>a man</b> and sat down near the window.</i>

Please note that the target items used in this experiment were constructed slightly differently from those in Chapters 4 and 5, as I intended to test further aspects that might impact participants' continuations. Two changes have to be mentioned. First, the target referent of the second experimental item (TI2) was introduced in an embedded clause and not in a matrix clause as the other test items. This change is predicted to have a detrimental impact upon the topic shift potential of the indefinite noun phrases introduced in this way, as several studies have shown that referents introduced in embedded sentences are less likely to make good topics (Maki et al. 1999). This observation parallels Roberts' (2010) analysis of topics as representing "root phenomena", occurring only in matrix sentences. Second, the fourth experimental item (TI4) differs from the rest of the test items, in the way the referents were established in the initial stories. Recall that the stories in TI1-TI3 introduced a global referent in subject position, whereas the target item was introduced in

<sup>10</sup> Please note that *pe*-marking in Romanian, in contrast to the distribution of indefinite-*this* and indefinite *so'n*, is possible with human referents in direct object position only.

direct object position in the second sentence of the stories. In TI4, the first referent was mentioned in subject position as well, but the target item was introduced by means of a prepositional phrase, which is generally associated with less prominent referents. Different studies have showed that the way in which a referent is introduced in the discourse impacts its accessibility. Thus, the critical items are faced with a stronger competitor, the subject referent, which is the clearly established aboutness topic constituent of the sentence fragment.

### 3.3.2 Procedure and data analysis

The first five main clauses (including subordinate ones, if there were any) of each continuation story provided by the participants, were coded by two independent judges for three aspects of the indefinite direct objects: (i) their referential persistence, (ii) their topic shift potential, and (iii) the type of anaphoric referring expressions used to pick up the referent for the first time. The two judges agreed upon 92% of the cases; in the case of disagreement (which were mostly concerned with determining the topic), differences were resolved through discussion. To determine the referential persistence of the critical items, the number of mentions of each referent in the subsequent discourse was coded. This method indicates at what stage in the continuation fragment there are more anaphoric expressions referring to the direct object referent than to the subject referent. Furthermore, the first instance at which a critical item became the aboutness topic in a matrix clause was considered an instance of topic shift. Whether this change was maintained after this point or not, was unimportant for the present analysis. Finally, the last textual characteristic established whether there are significant differences between indefinite-*this* and indefinite-*a(n)* referents with respect to the type of the first anaphoric expression used to refer back to them.

### 3.3.3 Results and discussions

One continuation was excluded from the analysis, because the participant was not a native speaker of English. This left a total of 10 continuations for each condition of the experiment (i.e. *this*-condition and *a(n)*-condition) to be coded and analysed.

### Referential persistence

The findings with respect to the referential persistence of referents reveal several interesting patterns. Figure 3.1 displays the average number of anaphoric references to the direct object referents for each continuation sentence (S1-S5) in both conditions of the experiment up to the last continuation sentence S5. Prediction 1 is confirmed, as the referents of the direct objects in the *this*-condition are referentially more persistent than the direct objects in the *a(n)*-condition. In other words, upon reading the target stories, participants showed a preference to continue talking about the referent of the critical item when it was preceded by *this* rather than when it was introduced by *a(n)*. The presence of *this* has a boosting effect in terms of referential persistence upon the noun phrase it precedes. Thus, the presence of *this* on the critical item makes it a better competitor for the subject referent.

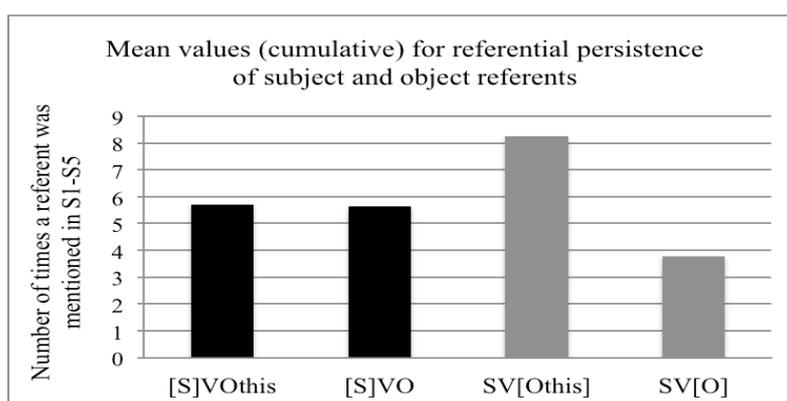


Figure 3.1: Mean values for referential persistence of subject and object referents

Furthermore, Figure 3.1 reveals an interesting insight into the relation between the referents of the critical items in both conditions and the subject referents. The difference in referential persistence between subject and critical referents in direct object position is higher in the *a(n)*-condition. On average, the subject referents are picked up in the subsequent discourse more often than the *a(n)*-referents. This observation does not hold for the *this*-condition (i.e. [SVOthis]), as we do not observe a significant difference between the referential persistence of the subject referent and the critical referent in direct object position. This observation brings up an interesting question, namely whether the presence of *this* on the direct or oblique object (TI4) has an impact upon the referential persistence values of the subject referent as well. Asked differently, is the higher referential persistence of the critical item (realized as a direct object or as a prepositional phrase) headed by *this*

suppressing the referential persistence of the subject referent? Or is it rather the case that the subject referents have a standard referential persistence value which changes only minimally, when other referents are present in the sentence?

To give an answer to this question it is necessary to look at the referential persistence of the referents for each target item separately. More concretely, TI2 showed a somewhat deviant continuity pattern, as: (i) both subject and direct object referents were referred to less in the subsequent discourse, and because (ii) the subject referent in the *this*-condition showed a higher referential persistence than the direct object. Starting from these observations, a question that arises at this point is, whether TI2 is responsible for the overall slightly higher referential persistence values for the subject referents in the *this*-condition?

A closer look at the referential continuity of all experimental items for the *this*-condition, presented in Figure 2, shows that the critical items realized as direct objects are picked up more often in the subsequent discourse compared to the subject referents. The values for referential persistence for the experimental items TI2 and TI4 show the reverse pattern: the subject referents exceed the critical items in referential persistence.

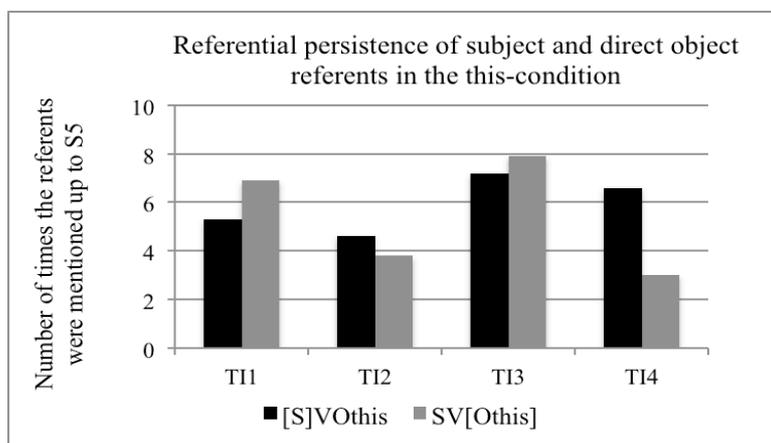


Figure 3.2: Referential persistence of subject and direct object referents in the *this*-condition for each target item

It is interesting to ask whether the overall referential persistence values for TI1 and TI3 are lower in the *a(n)*-condition as well, or not? A positive answer to this question is expected, if the divergent continuity patterns are due to the fact that referents in these conditions

were realized in less prominent positions (i.e. in an embedded clause (TI2) and as an oblique argument (TI4)).

A detailed analysis of the referential persistence values in the *a(n)*-condition (see Figure 3.3) shows that the subject referents are always referred to more often than the referents of the critical items headed by *a(n)*. However, the critical items show lower referential persistence values for TI2 and TI4 compared to TI1 and TI3. This observation patterns with the referential persistence valued in the *this*-condition. Givón (1983), among others has shown that referents mentioned in grammatically less salient positions (e.g. as prepositional phrase arguments) are less prone to be subsequently mentioned. Furthermore, the high topical position of the subject referent in TI2 and TI4 is also responsible for its high referential persistence. Thus, both factors: the high topicality of the subject referents and the weakened position of the critical referents in the two experimental items interact and show effects upon the referential persistence of their referents.

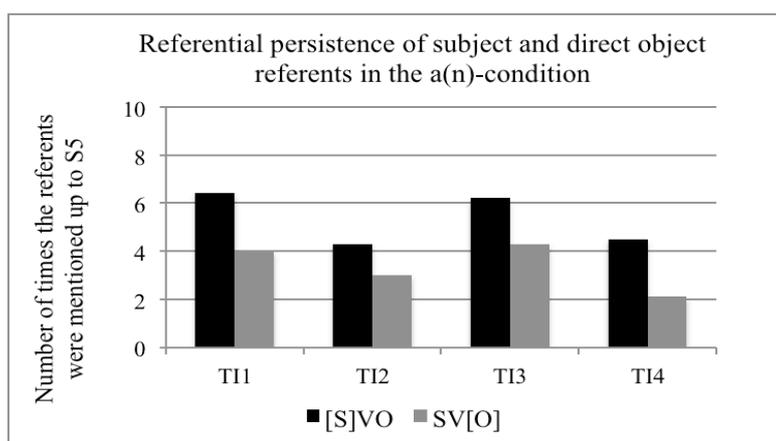


Figure 3.3: Referential persistence of subject and the referents of the critical items for the *a(n)*-condition

### Topic shift potential

The second textual characteristic investigated is the topic shift potential of the referents of the critical items headed by *this* compared that of the referents preceded by the indefinite article *a(n)*. The results in Figure 3.4 reveal two interesting patterns. First, the topic shift potential of indefinites headed by *this* is higher than that of indefinites headed by *a(n)*. Furthermore, referents preceded by *this* are more prone to be mentioned as the aboutness topics in the subsequent discourse compared to their simple indefinite counterparts. Second, for both types of critical items, the third continuation sentence (S3) seems to play

an important role, as the probability that a referent will be mentioned in topic position after this point dramatically drops. Thus, it seems that the effects of *this* in terms of topic shift potential are best seen by S3. After this ‘turning point’, the topic shift potential remains constant (i.e. it is not expected to increase).

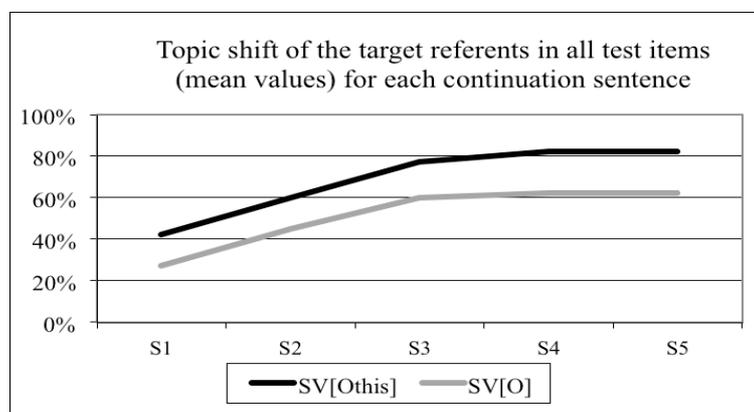


Figure 3.4: Topic shift potential of *this*- and *a(n)*-direct objects for each continuation sentence (S1-S5)

However, the topic shift potential of referents headed by *this* and those headed by the indefinite article does not differ significantly. It seems that the direct object referent in English is more prone to shift the topic after being introduced for the same time in a discourse, regardless of the presence or absence of *this*. If this observation is true, then it would imply that the correlation between topic(hood) and subject is less strong in English. One reason for this correlation might be due to English word order being more rigid. Thus, when a referent is mentioned in subject position (or perhaps in sentence initial position) may not be as informative as in languages with a more flexible word order. Several studies on this topic hinted at this difference (Hemforth 2000), but a suitable evaluation of the tenability of this suggestion awaits future research.

### Type of the first anaphoric expression

The last textual characteristic investigated was the type of referring expression used to refer back to the target items for the first time in the continuation sentences provided by the participants. Tables 3.4 and 3.5 display the absolute and mean values of this characteristic for both indefinite *this* and indefinite *a(n)* referents. More concretely, Table 3.4 lists the types of referring expressions used to refer back to the referent introduced by *this* and also the exact sentence-continuation in which these referents were mentioned for the first time,

whereas Table 3.5 presents the same results for the referents headed by the indefinite article.

Ten different participants provided continuations for the target items, resulting in 40 continuations for the *this*-condition and another 40 for the *a(n)*-condition to be coded and analysed. The data show that for the *this*-condition, four participants (i.e. 10%) chose not to mention the direct object referent again in a matrix clause, while in the *a(n)*-condition, almost 20% of the participants did not return to the direct object referent. This information is mirrored by the two sums in the last row of the last column (i.e. 90% for the *this*-condition and 82.5% for the *a(n)*-condition).

Table 3.4: Type of referring expression used for the first anaphoric item in the *this* condition

<i>Referents for the this-condition</i> (absolute and mean values)						
Ref. Type Sentence	Zero	Pronoun	Name	Def. unmod. NP	Def. mod. NP	Sum
S1	(0)	12.5%(5)	(0)	35%(14)	(0)	47%(19)
S2	(0)	15%(6)	(0)	5%(2)	(0)	20%(8)
S3	(0)	5%(2)	2.5%(1)	7.5%(3)	2.5%(1)	17.5%(7)
S4	(0)	2.5%(1)	(0)	(0)	(0)	2.5%(1)
S5	(0)	(0)	(0)	(0)	2.5%(1)	2.5%(1)
Sum	(0)	35%(14)	2.5%(1)	47.5%(19)	5%(2)	90%(36)

Table 3.5: Type of referring expression used for the first anaphoric item in the *a(n)*-condition

<i>Referents for the a(n)-condition</i> (absolute and mean values)						
Ref. Type Sentence	Zero	Pronoun	Name	Def. unmod. NP	Def. mod. NP	Sum
S1	(0)	12.5%(5)	(0)	30%(12)	2.5%(1)	45%(18)
S2	(0)	2.5%(1)	(0)	10%(4)	5%(2)	17.5%(7)
S3	(0)	5%(2)	2.5%(1)	7.5%(3)	(0)	15%(6)
S4	(0)	2.5%(1)	(0)	(0)	(0)	2.5%(1)
S5	(0)	(0)	(0)	(0)	2.5%(1)	2.5%(1)
Sum	(0)	22.5%(9)	2.5%(1)	47.5%(19)	10%(4)	82.5%(33)

The prediction that indefinites headed by *this* in direct object position do not need more lexical material to be taken up in the subsequent discourse was not confirmed by the above findings. The preferred type for the first anaphoric item used in a matrix clause to pick up both the *a(n)* and *this* referents is a definite unmodified noun phrase (i.e. 47.5% in the fifth column of the last row in both Table 3.4 and 3.5). Furthermore, slightly more participants took up the referent of the direct object by means of a pronoun in the *this*-condition (i.e. 35%), than in the *a(n)*-condition (i.e. 22.5%).

These findings show that the two types of indefinite noun phrases (headed by *this* vs. by the simple indefinite article *a(n)*) do not have a different impact upon the type of referring expression chosen to take up their referents again. In other words, both attenuated and less attenuated types of referring expression were chosen almost in the same proportion for the referents of indefinite *this* and *a(n)*. At first blush, this result seems to contradict Gernsbacher and Shroyer's (1989: 539) findings, which showed that participants "used less explicit anaphors to refer to *this*-introduced nouns and more explicit anaphors to refer to *a(n)*-introduced nouns". More concretely, they noted that out of the total of 376 pronouns used to refer back to both *this* and indefinite *a(n)* referents, 57% picked up the indefinite *this* referents, while only 43% picked up the *a(n)* indefinites. Furthermore, out of 177 definite noun phrases used to refer back to both types of indefinite noun phrases, 54% were found to refer back to the referents of indefinite *a(n)*, while only 46% definite noun phrases were used to take up the referents headed by *this*. Note, however, that the mean values reported in their study indicate the probability of an anaphor (realized as a pronoun or a definite noun phrase) to refer back to the antecedent of indefinite *this* rather than to the antecedent of the *a(n)* indefinite. This is one of the reasons why the results reported in Gernsbacher and Shroyer's (1989) study differs from the findings of this study. The study presented in this Chapter investigated the types of referring expressions employed *more frequently* by speakers to pick up referents headed by *this*, compared to referents headed by indefinite *a(n)*.

In Table 3.6 I recalculated the absolute and mean values for each type of referring expression used as an anaphor to refer back to indefinite *this* and indefinite *a(n)* referents in Gernsbacher and Shroyer's (1989) study. It seems that even in their study, there is no difference between indefinite *this* and indefinite *a(n)* with respect to the preferred type of anaphoric expression used for the first time to remention a referent. From the perspective

of language production, the distribution per type of expression is equal in the two conditions. In other words, Gernsbacher and Shroyer's (1989) study reveals that participants did not use a particular type of anaphoric expression to pick up the indefinite *this* referents and another type of anaphoric expression to pick up the indefinite *a(n)* referents, which casts a different light on the results.

Table 3.6: Mean and absolute values for the type of referring expressions in Gernsbacher and Shroyer's (1989) study

	<i>This</i> -indefinites	<i>A(n)</i> -indefinites
Pronoun	69% (214)	67% (161)
Definite NP	31% (95)	33% (81)
Sum	(309)	(242)

Despite choosing the same type of anaphoric expression to remention the referents in both conditions, participants preferred to continue talking about the referents using pronouns rather than definite noun phrases. In contrast, the results of the story-continuation experiment reported in the present study showed that referents in both conditions needed more lexical material to be mentioned again in the subsequent discourse than one might have expected based on the results of Gernsbacher and Shroyer (1989). It is interesting to ask why participants chose to refer back to both types of critical referents by means of a pronoun in Gernsbacher and Shroyer's (1989) study, and not by means of a definite noun phrase, as in the present investigation? Formulated differently, what is the crucial factor that reversed the preference of anaphoric continuation in the two studies?

A closer look at the experimental items in the two studies reveals a difference with respect to the animacy of the subject and direct object referent. More exactly, the target items introduced in direct object position in Gernsbacher and Shroyer's (1989) study were semantically inanimate (e.g. *a/ this egg*, *a/ this vase*) while the subject referents were semantically human and were generally realized by means of proper names. In the story-continuation experiment presented in this Chapter, the subject and target referents did not vary in animacy and were semantically human. Moreover, there was no match in gender between subject and direct object referents (except for T11). As there was no competition between the referents in terms of gender participants could have pronominalized the critical referents more often, by using the appropriate pronoun forms for each referent. However,

the prediction with respect to pronoun use in the present experiment was not met, as participants chose more elaborated types of referring expressions (i.e. definite noun phrases) to pick up the critical referents in both conditions. Thus, it was not the competition in terms of gender that hindered the participants in the present study to use more pronouns for the target referents. The presence of multiple referents in a sentence describing a transitive event generally reduces the overall rate of pronominalization (Arnold and Griffin 2007). However, as the experimental items used in both experiments introduced two referents, the competition in terms of mere presence of several referents in the sentence cannot be the factor that reversed the likelihood of pronominalization in the present study. I argue that the low pronoun rates are due to the competition of the two referents in terms of animacy<sup>11</sup>. This observation converges with a body of neurolinguistic research, which showed that transitive sentences are easier to process when the arguments are different in terms of their distinctive features such as animacy, for example (McElree 2006).

Summarizing, the results of the present study confirm the predictions made at the beginning of this Section, thus underlining the preferential discourse status of indefinites headed by *this* compared to *a(n)* indefinites. First, it was shown that upon introducing a referent by means of *this* in the discourse, the speaker chooses to refer to this referent more often in the subsequent discourse, i.e. usually up to the third sentence-continuation, after which the referential persistence does not increase dramatically. On the contrary, referents introduced by the indefinite article were less referentially persistent. Second, indefinite *this* referents showed a higher potential to shift the topic of the discourse than *a(n)* indefinites. Third, a more general observation is that the referents associated to *a(n)* indefinites are not excluded from being rementioned in the subsequent discourse and from shifting the topic. In contrast, referents associated with *this* indefinites are less likely to remain unmentioned in the ensuing discourse and not to shift the topic. The former type of indefinites is thus more flexible in occurring in both types of continuations, i.e. in those continuing with their associated referents and in those that do not. *This* indefinites on the contrary tend to be more strongly associated with only one type of continuation (i.e. one in which its associated referent is frequently rementioned and is shifting the topic). This observation

---

<sup>11</sup> In a language like Romanian, which does not differentiate between [+human] and [-human] personal pronouns, lower pronoun rates for multiple referents within one sentence would be expected, irrespectively of their animacy.

provides further evidence for the generalization that unmarked forms (i.e. *a(n)* indefinites) are less restricted in their distribution compared to marked forms, i.e. *this* indefinites (Givón 1981). Fourth, contrary to previous experimental findings, *this* indefinites did not show a preference to be taken up in the first continuation sentence by a more reduced type of referring expression. Participants preferred to pick up the referents introduced by both indefinite *this* and indefinite *a(n)* in the first following matrix clause by means of a definite noun phrase.

Up to this point, it was shown that indefinite introduces referents with particular referential properties and discourse functions. In the last Section of this Chapter, I come back to the second question raised by the end of Section 3.1, namely what is special about the demonstrative determiner *this* that it fulfils these functions?

### 3.4 Demonstratives and focus of attention

In the previous Sections, we distinguished between demonstratives that can be used deictically, anaphorically/cataphorically, discourse deictically and recognitionally. Additionally, English demonstrative *this* was shown to display a fifth use, which becomes visible in indefinite contexts. Despite their different uses, demonstratives have been shown to serve particular, arguably universal functions. In the following I will argue that indefinite *this* can be best accounted for in a model in which the interplay between speaker and addressee is taken into account. I argue that indefinite *this* was developed as a means to draw hearers' attention to new referents, compared to referents that are already part of the discourse participants' mental representations.

The traditional view on demonstratives is that they are used to indicate spatial distance. This view derives from the observation that the most basic or prototypical use of nominal demonstratives is as deictic elements (i.e. expressions used for pointing) and that all other uses are derived from the deictic use (Halliday and Hasan 1976, Lyons 1977, Diessel 1999, Roberts 2002). In this view, the speaker is the primary locus of referential information (i.e. the deictic center) and the relative distance between the speaker and the entity being referred to is responsible for the use of one demonstrative form over the other. Demonstratives differ cross-linguistically with respect to the complexity of spatial

distinctions they make. While, for example, Portuguese and Armenian employ a three-way distinction between proximal, distal and neutral, the English demonstrative system developed a two-way distinction, discerning between the proximals *this* and *these* on the one hand, and the distals *that* and *those* on the other hand.

After investigating the demonstrative system of 80 languages from different geographical areas and linguistic families, Diessel (1999) concluded that the binary distinction between distal and proximal is the most prevalent one, with 55% of the languages in his sample employing this distinction. Under the view that the spatial dimension is relevant in choosing one determiner form over the other, proximal demonstratives denote entities in the (immediate) proximity of speaker, while distal demonstratives denote entities remote from the speaker (Halliday and Hasan 1976, Lyons 1977, Anderson and Keenan 1985, Levinson 1983, Clark and Marshall, 1981, among many others). The use of *this* in (33a) indicates that the woman in question is located somewhere close to the speaker, while the woman mentioned in (33b) is at a certain distance away from the speaker. Demonstratives, then, contribute information beyond their descriptive content.

- (33) (a) This woman is insane.  
(b) That woman is insane.

Maclaran (1982: 91), among others, adopts a similar line of argumentation to account for the definite and indefinite uses of English *this*. He suggests that when a speaker uses (in)definite *this*, “the speaker is being instructed that the referent is to be found in the speaker’s world, not in the shared world of speaker and hearer. *This/these* expresses proximity to the speaker.” However, in the previous Sections we noted that indefinite *this* does not primarily direct hearers’ attention to referents in the proximity of the speaker. Rather, we observed that indefinite *this* signals: (i) the speaker’s intention to refer to a specific referent; (ii) referential continuity and (iii) a high potential for topic shift. Thus, an analysis in terms of proximity to the speaker may not be the best foundation for an account of English indefinite *this*.

Semantic and pragmatic approaches to the use of demonstratives in context made similar observations with respect to the fact that these types of referring expressions have extended functions, which cannot be reduced to the strict perceptual distinction between close and

faraway relative to a referential center. Talmy (1988: 168-169) has argued that the demonstrative expressions *this* and *that* are not strictly spatial, as the distance or space between the referential center and the indicated object cannot be objectively measured. Following this line of argumentation, Klein (1982: 166) has noted that “it is possible to say *Here comes my mother* when she is at a distance of 100 m, but one can also say *There’s my mother* when she is at a distance of 10 m.” So, the referential function of demonstratives is context-dependent and subjective and it does not indicate a particular measurable physical distance. A speaker may use a demonstrative expression to refer to more or less remote objects.

The proximal-distal distinction seems not to be enough to account for all the data found with demonstratives cross-linguistically. Some examples from Dutch and German underline this observation once more. Take for example the following dialogue excerpt cited from Jansen (1993: 768), in which a doctor consults a patient.

- (34) Doctor: Doet het zeer op *deze plek*?  
 Does it hurt on this place?  
 Is this where it hurts?
- Patient: Ja, op *die plek*.  
 Yes, on that place.  
 Yes, that is where it hurts.

Even though the patient suffers pain from a certain body part, it is the doctor who uses the proximal demonstrative *deze*, whereas the patient uses the distal determiner form *die* to refer to his hurting body part. In investigating the use of proximal and distal demonstratives in German, Fuchs (1993) showed that the contribution of demonstratives can not be reduced to a proximal-distal distinction. Rather, the distinctions made by demonstratives are tailored to suit the amount of shared knowledge between the discourse participants.

Another piece of evidence for the insufficiency of the spatial analysis of demonstratives in terms of the distinctions between proximal and distal comes from Enfield (2003), who investigated the use of the demonstrative determiners *nii* and *nan* in Lao (a Tai-Kadai language). He observed that only one determiner gives information about location, and

none of the two determiners encodes information about distance. He furthermore noticed that the determiner *nii* is more general and neutral with respect to distance, whereas *nan* is used for referents that are remote from the discourse participants, or not “here” (Enfield 2003: 115). It seems that the space close to the discourse participants (the “here” space), as argued by Enfield, is negotiated at each point dependent on different contextual factors, including visibility and access. Enfield considers that the semantics of the two determiners in Lao is rudimentary and that the different uses of the determiners are to be attributed to different contextual factors during the process of referring. In sum, it seems that demonstratives encode abstract semantic notions and the use of a demonstrative reflects the speaker’s subjective attitude towards the physical surrounding. Hottenroth (1982) considers that objective distances are variables in the meaning of demonstratives, which become fixed only in the (extra)linguistic context and only when particular pragmatic factors are met.

An analysis of demonstratives in terms of faraway vs. nearby is less promising, as it seemed at first blush, especially if we aim to explain the indefinite use of *this* as an extension of its basic deictic use. While the analyses of demonstratives in terms of distance have focused on the centrality of the speaker, alternative accounts have highlighted the importance of the interaction between speaker and hearer. For example, deixis represents for Fillmore (1982, 1997) and Hanks (1992: 57): “[...] perhaps the clearest linguistic indicator of the interactive transformations involved in producing context”. Kirsner (1979) following García (1975) pursued a similar account of demonstrative expressions in terms of high or low deixis, which is loosely defined as „the force with which the hearer is instructed to find the referent“ (García 1975: 65). Kirsner shows that the Dutch equivalent of English *this* signals high deixis as the speaker aims to draw attention to the referent. The Dutch equivalent of English *that* signals low deixis, as the speaker is less interested to draw attention to the referent headed by this determiner. Demonstratives are treated as instructions directed towards the hearer to identify the speaker’s intended referent (Kirsner 1979, Diver 1984, Strauss 2002). Thus, they represent a mechanism to signal the information status or a referent in terms of the speaker’s knowledge with respect to the amount of information he believes to be shared by the hearer. Drawing on English data, Fillmore (1997) analysed the use of the demonstrative *that* as representing shared information between the discourse participants (i.e. hearer-old information), while *this* was

analysed as an indicator of lack of shared knowledge between speaker and hearer (i.e. hearer-new information).

Extending Halliday and Hasan's (1976) work, Sidner (1983) and Strauss (2002) proposed an analysis of demonstratives in terms of gradient focus, which translates into the degree of "attention a hearer should pay to the referent" (Strauss 2002: 135). In these models, focus is computed based on the amount of shared knowledge between discourse participants and by the relative importance attributed by the speaker to the referent headed by the demonstrative. According to this account, *this* is used whenever the speaker intends to draw attention to a particular referent, which is not in the focus of attention of the hearer. On the contrary, *that* is used when the referent does not want to draw attention to a referent, which represents shared information between the discourse participants. Figure 3.5, which is an adapted version of Halliday and Hasan's 1976 and Strauss' 2002, displays the interaction of the two factors built into the system of demonstratives *this* and *that*.

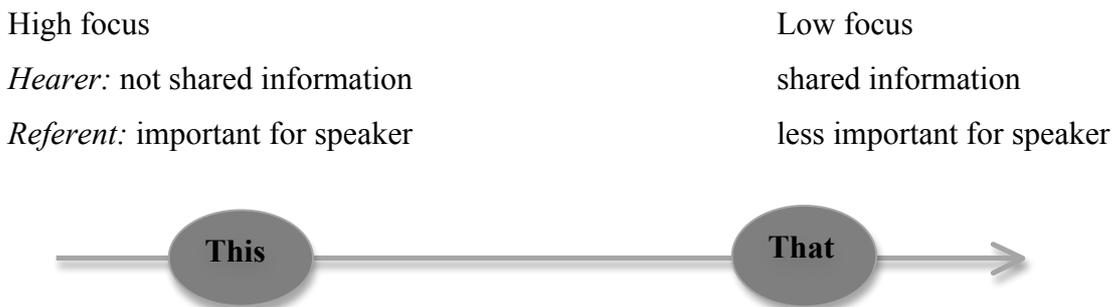


Figure 3.5: Schema for demonstrative reference adapted from Halliday and Hasan (1976) and Strauss (1993)

It is important to note that in these models, the two demonstrative forms do not rigidly express either low or high focus. Rather, given their different contexts of use, each demonstrative form can be interpreted as signalling different degrees of focus. Based on an extensive corpus research of several genres from spoken discourse, Strauss (1993) furthermore notes that:

“[...] subordinate to and implicit in this notion of variable degrees of focus motivating speaker choice forms are at least two additional factors : 1) the relative amount of information that the speaker presumes the hearer to have

with respect to the referent and 2) the relative importance of the referent itself to the speaker. (Strauss 1993: 404)

Subsequent corpus investigations into the use of demonstratives in written discourse (Oh 2001) revealed that writers use demonstrative *that* for peripheral reference. On the contrary, demonstrative *this* is used: (i) to refer to central referents; (ii) to emphasize a new property of the referent (property of “newness”); and (iii) to introduce a referent cataphorically, foreshadowing the centrality of the referent.

I consider that the use of indefinite *this* can be best accounted for in a model in which the interplay between speaker and addressee is taken into account. In line with previous research, I consider that a speaker uses a definite or indefinite demonstrative expression whenever he intends: (i) to refer to a particular referent, and (ii) to direct the attention of the addressee to this particular referent. A referent introduced in this way in the discourse has a strong discourse structuring potential, as it is referentially more persistent and shows a high topic shift potential. Firstly, the referential intention of the speaker is in general straightforward for definite uses of demonstratives, which generally indicate a particular referent. However, as we observed in Section 3.2.1, noun phrases preceded by indefinite *this* display a strong and robust tendency for referential readings (i.e. specific, wide, no covariation) as well. Referents associated with indefinite *this* are specific and show a preference for wide scope readings. The referential properties of indefinites headed by *this* become even more evident, when compared to their simple indefinite counterparts, which are ambiguous between a specific and a non-specific reading. Secondly, different researchers have noted that demonstratives are a means to focus hearer’s attention to a particular referent. In other words, demonstratives function as a means to coordinate the discourse participants’ joint focus of attention (Clark 1996, Diessel 2006). To this end, the speaker directs the hearer’s attention to a particular referent in the discourse. Directing hearer’s attention to a particular referent can vary in intensity (García 1975, Halliday and Hasan 1976, Kirsner 1979, Strauss 2002, Piwek et al. 2008). While on the one hand demonstrative *this* is associated with an instruction for the hearer to find the referent, on the other hand, demonstrative *that* is associated with a less intensive, neutral instruction for the addressee to look for the referent. Demonstrative *this* is thus used for referents that signal intense indicating or high deixis in García’s (1975) terminology, while *that* signals less intensive and low deixis.

In their fundamental deictic use, demonstratives are accompanied by a pointing gesture, which is considered “the quintessential tool for initiating joint attention” (Franco 2005: 139, see also Moore and Dunham 1995, Tomasello 1995, Diessel 2006). The pointing gesture fulfils two basic functions, namely locating the intended referent in the discourse situation relative to the deictic center (i.e. the speaker) and manipulating the hearer’s shared focus of attention (Moore and Dunham 1995, Tomasello 1995, Levinson 2004, Diessel 2006).

Apart from their exophoric use, demonstratives can be used without a pointing gesture, as discussed in the previous Sections. However, the endophoric uses of demonstratives have been shown to constitute extensions of their basic deictic use. Different investigations into the acquisition and diachronic development of demonstratives have shown that deictically used demonstratives are acquired first and that they provide the initial stage for the development of demonstratives into grammatical markers (Brugmann 1904: 7-8; Bühler 1934: 390; Diessel 1999: 109-112). In their anaphoric/cataphoric, discourse deictic and indefinite uses, demonstratives refer textually to a linguistic referent, rather than to a concrete referent in the physical surrounding. However, their discourse function is similar to that of exophoric demonstratives, as they are used to coordinate the hearer’s shared attention focus to the intended referent. In all uses of the demonstratives, hearer’s attention may be manipulated in such a way as to establish a referent as the new focus of attention, or to change the focus of attention to a new referent, which was not in the current focus. In other words, demonstratives can be use to introduce new topics, as illustrated in (35a) in the discourse, or to initiate a topic shift, as illustrated in example (35b).

(35) (a) Look at this book.

(b) Mary saw Paul. This man was wearing a blue jacket.

This development is not a particularity of English. As Küntay and Özyürek (2002) have shown, Turkish uses two different demonstrative forms to differentiate between new referents and referents that contrast two previously established referents. Speakers use *şu* to draw hearers’ attention to a new referent, while *bu* and *o* differentiate between two or more referents that have been already introduced in the discourse, depending on their location relative to the speaker. I argue that indefinite *this* was developed as a means to draw hearers’ attention to new referents, compared to referents that are already part of the

discourse participants' mental representations. I do not assume two lexical entries for the definite and indefinite uses of *this* (as for example Prince 1981, Lyons 1999, Ionin 2006 do), but analyse indefinite *this* as an extension of its regular, definite deictic use. Furthermore, as the results of the experimental study reported on in this Chapter showed, indefinite noun phrases headed by *this* are referentially more specific and their referents show particular discourse properties. On the one hand, the referential properties of the referents associated with the indefinite *this* noun phrases such as specific, wide scope readings derive from the intention of the speaker to refer to a particular referent. On the other hand, the particular discourse properties of indefinite *this* referents are not a reflex of the degree of importance attributed to the referent by the speaker (contrary to previous accounts). In other words, a referent need not be intrinsically important in the discourse in order to be introduced by indefinite *this*. Rather, when speakers introduce a referent by means of indefinite *this*, they intend to draw hearers' attention to a particular referent newly introduced in the discourse. This intention to draw attention to a particular referent is reflected in the text by the referential persistence of the referent introduced in this way and by its high topic shift potential, which can be subsumed under the notion of discourse structuring potential.

### 3.5 Summary

The results of the experimental study presented in this Chapter underline the function of English indefinite *this* as a means to introduce brand-new referents in the discourse with particular referential and discourse-pragmatic properties. At the sentence level, indefinite *this* noun phrases show a robust tendency for referential readings (i.e. are specific, show wide scope, etc.). At the discourse level, referents introduced by indefinite *this* are more likely to be picked up in the subsequent discourse and to shift the topic. However, contrary to previous findings, subsequent pronominalization was shown not to be sensitive to the use of indefinite *this*. These results indicate that indefinite *this* functions as a discourse structuring mechanism, rather than representing a means to establish accessible or important discourse referents. Indefinite *this* is used by speakers as a cue that shifts hearers' predictions about the ensuing discourse. Taken together, the results furthermore suggest that anticipation-biases can arise simultaneously at the local, sentence level as well as at the discourse level. The next Chapter extends the present analysis by investigating the discourse-pragmatic effects of German noun phrases headed by indefinite *so 'n*.

## 4. German indefinite *so'n*

The semantic and pragmatic literature on German indefinite noun phrases has been primarily concerned with the referential properties of indefinite expressions headed by the simple indefinite article *ein* (*a(n)*). This Chapter focuses on indefinite noun phrases preceded by the indefinite determiner *so'n* ('so-a') that are used in conversational, informal German, as illustrated in (1).

- (1) In unserer Stadt gibt's *so'n Mann*<sub>i</sub>, er<sub>i</sub> ist nicht irgendeiner, nein, er<sub>i</sub> ist unser neuer Bürgermeister, er<sub>i</sub> sieht wie ein Vogelstrauß, [...] <sup>12</sup>.  
'In our city there is so-a man<sub>i</sub>, he<sub>i</sub> is not anyone, no, he<sub>i</sub> is our new mayor, he<sub>i</sub> looks like an ostrich, [...].'

Throughout this Chapter, I argue that indefinites headed by *so'n* are used to focus attention to a particular individual or property that will be elaborated upon in the subsequent discourse, rather than to an entity in the preceding discourse or perceivable situation. The use of *so'n* parallels the referential and discourse behaviour of English indefinite noun phrases headed by *this* in its indefinite use. I argue that similar to English indefinite *this* German indefinite *so'n* is used as a discourse-structuring device.

The aims of this Chapter are twofold. First, I will show that, contrary to indefinite noun phrases headed by the simple indefinite article *ein* ('a/an'), German indefinite descriptions preceded by *so'n* show a strong tendency for a referential interpretation (i.e. wide scope with respect to other operators, rigid reference, etc.). The theoretical considerations regarding the referential behaviour of indefinites preceded by *so'n* will be substantiated by an experiment (Experiment 1), which investigates their referential properties under two different operators: the intensional operator *will* and negation. Second, based on the findings of a second, web-based story continuation experiment (Experiment 2), which is similar to that discussed in Chapter 3 for English, I will show that the discourse structuring potential of the referents preceded by *so'n* can be spelled out in terms of referential persistence and topic shift potential. In light of the results of this survey, the relevant discourse contribution of *so'n* can be characterized in terms of the likelihood that the

---

<sup>12</sup> <http://Section.mzee.com/forum/archive/index.php/t-100000620.html> [visited: May 2010]

referent introduced in this way will be (frequently) mentioned again and will shift the topic in the subsequent discourse. In other words, upon encountering indefinite *so'n* the hearer is instructed to establish a permanent discourse representation as further information about the newly mentioned referent will follow- given that other parameters remain the same. The general conclusion of this Chapter patterns with the observations made with respect to English indefinite in Chapter 3, as the contribution of indefinite noun phrases headed by *so'n* cannot be thoroughly accounted for, unless the discourse structuring potential of this type of referring expression is integrated into the analysis. The stability of reference of these indefinite noun phrases comes from their discourse prominence in terms of discourse structuring.

In the following Sections, I will explore the characteristics and behaviour of *so'n* in more detail. This Chapter is organised as follows. In Section 4.1 I discuss different conventional uses of *so'n*, which have been presented in the traditional and more recent literature. I will show that examples like (1) involve a usage of *so'n* that has not been adequately captured in the existing literature, which focused on the intensifying and hedging usages of *so'n* in out-of-the-blue contexts. Section 4.2 is dedicated to the referential behaviour of noun phrases headed by *so'n* in contexts containing different types of operators at sentence level. The findings of a web-based judgement task (Experiment 1) will show that *so'n* triggers the wide scope readings of the nouns it precedes. I will then discuss in Section 4.3 the discourse properties of German indefinite *so'n*. Having established this background, I then report on the findings of a story continuation experiment (Experiment 2), which resembles the experiment presented in Chapter 3. The study investigates the discourse structuring potential of the indefinites preceded by *so'n* with respect to their referential continuity, their topic shift potential and the type of referring expression used to pick up these referents in the discourse. The final Section summarizes the key observations and results presented in this Chapter.

#### 4.1 Demonstrative uses of *so'n*

German *so'n* functions as an indefinite determiner alongside the simple indefinite article *ein* and the indefinite demonstrative *dieser* (von Heusinger and Deichsel, 2010). The fact that there are no studies on the diachronic development of *so'n* could be due to the fact that the occurrence of *so'n* is predominantly attested in informal registers of written and spoken German. Its spelling varies amongst linguists and in corpora of written language between the contracted form *son* and the cliticized form *so'n*. These observations reflect the fact that previous studies could not agree upon the exact status of this indefinite determiner, namely whether to analyse it compositionally, building on the semantics of the demonstrative *so* and that of the simple, cliticized indefinite article *'n*, or whether to analyse it as an independent determiner form (i.e. *so'n*). Henn-Mennesheimer (1986) and Lenerz and Lohnstein (2004), for example, adhere to the first view and analyse *so'n* as a mere combination of *so* with the different suffixed forms of the cliticized simple indefinite article *ein*. An argument for this analysis comes from the observation that neither the cliticized indefinite article, nor *so'n* can be used as pronouns, as illustrated in (2).

- (2) \*So'n/ 'n    lese    ich    immer    gerne.  
           So-a    read    I    always    readily  
           ' \*So-a I always readily read.'

On the contrary, Hole and Klumpp (2000) argue that *so'n* is not the mere contracted form of *so ein*, but rather a fully developed new article in German. One argument that speaks for this view derives from the different plural paradigms of *so'n* and of the simple indefinite article *ein*. The Duden grammar (1995) notes that the inflectional endings for gender and case in the singular of *so'n* are identical to those of the simple indefinite article *ein*, as Table 4.1 shows. Nevertheless, the (optional) plural paradigms for *so'n* and the simple indefinite article differ, as the indefinite article, but not *so'n*, lacks specialized plural forms, illustrated in Table 4.2<sup>13</sup>. Given this data, *so'n* could not have inherited its plural forms from the indefinite article, which does not realize them. Instead, Hole and Klumpp (2000) argue that *so'n* follows the general pattern of German pronominals, which inherit their plural paradigms from the identical nominative feminine singular forms.

<sup>13</sup> Please note that the genitive singular and the plural forms of *so'n* are rather marginal forms, as native speakers of German report.

Table 4.1: Singular paradigms of the indefinite article *ein* and *so'n*

Singular	Masculine		Feminine		Neuter	
	EIN	SO'N	EIN	SO'N	EIN	SO'N
Nominative	<i>ein</i>	<i>so'n</i>	<i>eine</i>	<i>so'ne</i>	<i>ein</i>	<i>so'n</i>
Genitive	<i>eines</i>	<i>(so'nes)</i>	<i>einer</i>	<i>(so'ner)</i>	<i>eines</i>	<i>(so'nes)</i>
Dative	<i>einem</i>	<i>so'nem</i>	<i>einer</i>	<i>so'ner</i>	<i>einem</i>	<i>so'ner</i>
Accusative	<i>einen</i>	<i>so'nen</i>	<i>eine</i>	<i>so'ne</i>	<i>ein</i>	<i>so'n</i>

Table 4.2: Optional plural paradigms of the indefinite article *ein* and *so'n*

Plural	EIN	SO'N
Nominative	∅	<i>so'ne</i>
Genitive	∅	<i>so'ner</i>
Dative	∅	<i>so'nen</i>
Accusative	∅	<i>so'ne</i>

The plural paradigm of *so'n* parallels that of adnominal possessive pronouns like *mein*, *dein*, *sein* ('mine, yours, his') and the negation *kein* ('no'), which are analysed (in a narrow sense) as articles (Eisenberg 1994). In this Chapter, I consider *so'n* as the contracted form of the demonstrative and the cliticized indefinite article *ein*. Despite the disagreement on its exact form, the literature unanimously admits the multifunctionality of *so'n*. In order to shed some light on the way in which the new function of *so'n* has emerged, I will begin by distinguishing between several uses of *so'n* and proposing a continuum which links the different uses. According to descriptive grammars, *so* has a broad variety of uses, as different as an (modal or degree) adverb, adverbial connective, subordinating conjunction, equative comparison and connective (cf. Burkhart 1982, Ehlich 1986, Zifonun, Hoffmann and Strecker 1997, König, Stark and Requardt 1990, Weinrich 2003). Its most prominent use, however, is argued to be that of a demonstrative expression that combines with nouns, adjectives and verbs (or their immediate functional projections). In this Chapter I will focus on instances in which demonstrative *so* combines with noun phrases.

As a demonstrative, *so'n* can point to an action or referent outside the text, it can refer back to certain points in the discourse, or it can occur without an exophoric reference or an antecedent. These different combinatory options of *so* correspond to different meanings

and functions of the expression it precedes. Further meaning differences arise depending on whether the expression headed by *so'n* is gradable or not. In what follows, I review some accounts of *so'n* as a demonstrative expression found in traditional grammars of German and the more recent literature. After reviewing these demonstrative uses, I will turn to the main focus of my investigation, namely those cases in which *so'n* is used in “out-of-the-blue” contexts, to introduce new entities that will play a prominent role in the subsequent discourse.

#### 4.1.1 Deictic and anaphoric uses of *so*

As already pointed out, common uses of *so'n* as a demonstrative expression involve deictic and anaphoric uses. For both uses, there is a difference in interpretation depending on whether the expression headed by *so* is gradable or not. In its deictic use, the exact characteristics of the lexical item preceded by *so* can be inferred based on the gesture that accompanies the expression, as illustrated in (3).

- (3) (with a pointing gesture) Ich hätte einen Tannenbaum, ungefähr *so groß*.  
 ‘I would like a fir tree, about this big.’

Ehlich (1986) argues that in order to identify the exact value that *so* points to, the hearer has to identify the object that correlates with this value first. Thus, the hearer has to assume in a first step that the speaker is pointing to an object and, in a second step, he has to deduce the intended properties on the basis of the referent and the pointing gesture. Accordingly, the hearer of sentence (3) must identify the object the pointing gesture refers to, i.e. the fir tree. Because *so* modifies the adjective *groß* (‘big’), the hearer knows that the relevant property of the object is height (and not width, for example). Additionally, the speaker’s pointing gesture is indicating the degree of height of the tree. When combined with expressions that are inherently gradable, as the adjective ‘big’, *so* points to a degree on a particular scale. However, when deictic *so* combines with a non-gradable expression (i.e. most nouns), *so* indicates a property, as illustrated in (4).

- (4) (pointing to a dog) So ein Hund habe ich auch.  
 ‘I have such a dog too.’

In example (4), the noun phrase headed by *so* refers to a dog. Since *so* modifies an expression that is not gradable, he points to a property of that particular dog (e.g., its height, breed or behaviour). As in the gradable case discussed above, the hearer infers the relevant properties of the dog based on the properties of the referent of the pointing gesture. Furthermore, the hearer has to assume a similarity relation between the referent of the pointing gesture and the referent intended by the speaker. Note that in comparison to the gradable expressions discussed above, in the case of non-gradable expressions, the head noun does not always fix the exact property pointed at. In other words, the context of the utterance has to be evaluated as well, to determine the exact property pointed at by the expression.

In its anaphoric use, *so* can combine with gradable as well as non-gradable expressions and point to a degree. In example (5), *so* modifies the gradable adjective *schwer* ('heavy') and we have to assume, as in the deictic use, that the degree *so* points at will be inferred from the antecedent, in this particular case *30 Kilo schwer* ('30 kilograms heavy'). In this sense, *Bobi's* weight is similar to *Gora's* (i.e. at least 30 kg).

- (5) Gora ist 30 Kilo schwer. Bobi ist auch so schwer.  
 'Gora weights 30 Kilo. Bobi is that heavy too.'

When combined with a non-gradable expression, anaphoric *so* refers to a property of the antecedent, while the relation between antecedent and the anaphor headed by *so* is based on similarity. In example (6), we can assume along the same lines with (4) that *so* refers back to the entire noun phrase *einen schönen Hund* ('a nice dog'). In other words, *so* indicates that *Pau's* dog is as nice as *Ruhl's*.

- (6) Ruhl hat einen schönen Hund. So ein Hund hat Paul auch.  
 'Ruhl has a nice dog. Paul has such a dog as well.'

Summing up the observations made up to this point, we have seen that demonstrative *so* can be used as a deictic and as an anaphoric expression. Whether *so* picks up a degree or a property of the referent depends on the type of expression it heads, i.e. if it is gradable or non-gradable. The next Section introduces a third use of *so*'n in which the demonstrative is neither accompanied by a pointing gesture nor does it refer back to an antecedent.

### 4.1.2 ‘Out-of-the-blue’ uses of *so'n*

Exactly like English demonstrative *this*, *so'n* can be used in indefinite or out-of-the-blue contexts as well. In these contexts, the noun phrase headed by the marker is neither accompanied by a pointing gesture, nor does it have an antecedent. Furthermore, in out-of-the-blue contexts, it is the head noun that bears the main stress, and not *so* or other constituents of the sentence. This signals that the information conveyed by the noun phrase is not presupposed, but discourse-new and hearer-new. As in the deictic and anaphoric uses described above, when preceding a gradable expression, out-of-the-blue *so* has an intensifying effect, and when it precedes a non-gradable expression, *so* has a hedging effect. In addition to these two uses, I will show that *so* displays a novel use, namely that of a discourse-structuring device. In other words, I will show that *so'n* introduces news entities, which will play a crucial role in the ensuing discourse.

Let us start with the intensifying effect<sup>14</sup>, which is triggered every time *so* heads adjectives or other gradable expressions (Thurmair 2001, Umbach and Ebert 2004). In both examples given in (7), *so* could be paraphrased by *sehr* (‘very’). Following this analysis, the intensifying effect of *so* has the same interpretation as the anaphoric and deictic use when combined with gradable expressions (i.e. it is a degree modifier).

- (7) (a) (the grandfather talking about his nephew)  
 Er ist *so* groß.  
 ‘He is so big.’

<sup>14</sup> Note that the term “intensifying” is not necessarily adequate, because when it precedes absolute adjectives (as for example “full” or “wet”, which determine a partially closed class), *so* favours their downgrading interpretation, as (77c) shows. As other absolute adjectives, *voll* (‘full’) cannot be further intensified. Thus, when modified by *very* or *so'n*, only a downgrading effect can be achieved. Umbach and Ebert (2004) suggest an analysis of *so* in terms of a degree modifier (i.e. THIS-DEGREE) that achieves its value from a silent demonstration. The silent demonstration is considered to point to an object, situation or property, as the information in the square brackets shows.

- (7c) (A: Put the apples in the fridge.)  
 B: Lieber nicht, der ist so voll [wie mein Magen].  
 ‘I would rather not, it is pretty full.’ [as my stomach]. (adapted from Umbach and Ebert 2004)

- (b) Er ist *so* ein Kind, (\*das in Indien geboren wurde).  
 ‘He is so-a child (that was born in India).’

Note that gradable expressions as the one illustrated in (7b), are evaluative nouns (e.g. ‘idiot’, ‘pedant’) and/ or sortal expressions (e.g. ‘child’) that refer to a temporary property or phase of an entity (Beyssade 2008). The underlying property of such expressions, which can hold to a greater or lesser extent in a certain context, makes these expressions gradable. Returning to example (7b), the predicatively used noun phrase *so ein Kind* (‘so-a child’) refers to an essential property characterizing the subject of the sentence and does not refer to an individual. This observation is underlined by the infelicitous continuation given in parentheses, in which the predicatively used noun phrase *so ein Kind* (‘so-a child’) cannot corefer with a relative pronoun. Concluding, when *so'n* is used in out-of-the-blue contexts, with no deictic gesture or anaphoric antecedent, it can head predicatively used nouns, whenever they are gradable expressions. Thus, the noun phrase headed by *so'n* does not refer to an individual, but to a property of the referent associated with the expression. In out-of-the-blue contexts, *so'n* accentuates or intensifies the most salient property introduced by the expression it precedes.

Not all predicatively used nouns behave alike, however. In example (8a), the noun *Mann* (‘man’), which is headed by the simple indefinite article refers to a particular individual and can be used predicatively. The *so'n* variant of example (8a) is given in (8b) and it does not refer to a specific individual. The presence of *so'n* is rendered infelicitous, as there is no particular property that can be modified by the expression in question. Furthermore, it seems that there is a mismatch in terms of givenness (i.e. new vs. given/old information) between the types of referring expression used for the subject referent and the referent in predicative position. The felicitous use of the subject noun phrases (i.e. *Paul*, *er*, *mein Freund*) presupposes a high activation status of the referent in question, whereas *so'n* signals a low activation status of the same referent.

- (8) (a) Paul/ er/ ?mein Freund ist ein Mann (...der in Indien geboren wurde).  
 ‘Paul/ he is a man (...that was born in India).’
- (b) ?Paul/ ?er/ ?mein Freund ist *so'n* Mann (...der in Indien geboren wurde).  
 ‘Paul/ he is so-a man (...that was born in India).’

Based on such examples, it could be argued that *so'n* cannot be combined with predicatively and identificationally used non-gradable expressions. However, taking a closer look at example (8b), it seems that this assumption is incorrect. In this case, the noun phrase *so'n Mann* ('so-a man') refers to one of the properties denoted by the subject noun phrases. The relation between the subject and the object referent in the examples in (8) is responsible for the acceptability of (8a) and the unacceptability of (8b). More precisely, in contrast to (8a), the noun phrase headed by *so'n* in (8b) must provide new information with regard to the subject referent in order to be acceptable. As one of the most salient properties of the noun phrases in (8b), i.e. *Paul* ('Paul'), *er* ('he'), *mein Freund* ('my friend') is the property of being male, the presence of *so'n* expresses redundant information in this case and is rendered infelicitous. Thus, *so'n* must precede a noun phrase that introduces new information about a particular entity. A case in point is example (9), where the subject noun phrase *ein/der Philosoph* ('a/the philosopher') does not necessarily refer to a male individual and the presence of *so'n* heading the nominal conveys new, non-superfluous information.

- (9) (a) Ein Philosoph/ der Philosoph ist ein Mann, der viel liest.  
'A/ the philosopher is a man that reads a lot.'
- (b) Ein Philosoph/der Philosoph ist so'n Mann, der viel liest.  
'A/ the philosopher is so-a man that reads a lot.'

Furthermore, consider a situation in which we want to express the fact that *Paul* behaves like a woman or like a jester, properties that are not necessarily salient properties of someone who bears the name *Paul*. Even though the nouns *woman* or *jester* are not gradable, *so'n* can precede them in this context, as they introduce new and relevant information about the referent in question.

- (10) Paul ist so'ne Frau (in Sachen Einkaufen)/ so'n Spaßmacher.  
'Paul is so-a woman (with regard to shopping)/ so-a jester.'

It can be argued that the use of *so'n* in (10) is metaphoric or evaluative in the sense of Strohmeyer (1907) or van Peteghem (1993), which means that the predicatively introduced noun phrase introduces a temporary property or aspect of the referent. In other words, the

fact that *Paul* is a woman or a jester is valid for a particular point in time. This aspect is accentuated if the sentence in (10) is further modified by a sentence that makes this property permanent (e.g. *seit 60 Jahren*, 'for 60 years'). Furthermore, the presence of *so*'n in this context underlines the evaluative attitude of the speaker towards the referent of the predicative noun phrase.

Besides the predicative and identificational contexts discussed above, noun phrases headed by *so*'n can be used as hedging expressions in out-of-the-blue contexts. As in the case of the intensifying effect, the hedging effect arises whenever the expression headed by *so*'n has no antecedent and is not accompanied by a pointing gesture, i.e. in out-of-the-blue contexts and when the entity preceded by *so* is non-scalar or non-gradable. Burkhardt (1982: 90) explains hedging occurrences of *so* as an indication of the speaker's insecurity with respect to the exact qualities or properties of the referent headed by *so*. According to his analysis, the noun phrase *so*'nen *Hund* ('so a dog') in (11) could be paraphrased as: [...] *that/ this kind of dog, you know what I mean [...]*. In a similar vein, Thurmair (2001) analyses the hedging function of *so* as a metalinguistic approximation.

- (11) Ich habe gestern *so*'nen *komischen weißen Hund* gesehen.  
 'I saw so-a strange white dog.'

Similarly, for a context like (12), in which a speaker wanted to buy paper clips but did not know the exact term used for these objects, he could use *so* to signal an approximation to the sense of the word to which it is attached.

- (12) (A customer in a store)  
 Haben Sie *so* eine *Klammer*?  
 'I would like so-a clip.' (adapted from Umbach and Ebert 2004)

Umbach and Ebert (2004) analyse the hedging use of *so* as being similar to its intensifying use, namely as an instance of a silent demonstration. While in its intensifying use *so* picks up a degree, in the hedging use, *so* picks up a property, as it combines with non-gradable expressions. If we return to example (12), we could state that the particular object the speaker is looking for is an instantiation of the object denoted by the noun phrase *so eine Klammer* ('so-a clip'). Thus, the property inferred from the object of the silent

demonstration shifts the denotation of *eine Klammer* ('a clip') to the denotation of *so eine Klammer* ('so-a clip'). The import of *so* is not to widen the denotation of the noun phrase it heads, but to shift the value of the demonstration to an element of the demonstration. We have seen so far that in out-of-the-blue contexts, *so'n* has an intensifying effect on gradable expressions and a hedging effect on non-gradable expressions. In the next Section I introduce an additional context in which *so'n* can be used and which will be accounted for in terms of discourse discourse structuring.

### 4.1.3 *So'n* as a marker of referential structure

No matter how intuitively correct an interpretation in terms of a silent demonstration might seem, a question that arises is on what grounds does the expression headed by *so'n* achieves its value and interpretation. Consider example (13), in which *so* heads the contracted form of the indefinite article *einen* ('a(n)', Masc.Sg.Acc) to precede the non-gradable expression *Lehrer* ('teacher'). Here, it is unclear how the silent demonstration actually picks out the intended object or referent (i.e. its value). Furthermore, it is unlikely for the speaker in (13) not to know the exact characteristics that determined her to fall in love with the teacher in question.

(13) Also, da gibt's *so'n*en Lehrer<sub>i</sub> in meine Schule, in den<sub>i</sub> ich verliebt bin.

'So, there is this teacher in my school whom I'm in love with.'

In example (13), the noun phrase headed by *so'n* introduces a referent, which is hearer-new and discourse-new, rather than referring back to a referent that represents shared knowledge between speaker and hearer. Even though the hearer cannot immediately identify the referent *so'n*en *Lehrer* ('so-a teacher'), he is nevertheless given a hint that the referent is likely to be picked up in the following discourse. This observation is highlighted in the mini-discourse in (14), which is taken from the sentence-continuation experiment reported in Section 4.3.2. In this example, the referent of *so'n*en *Jungen* ('so-a boy') is picked up in the subsequent discourse several times, signalling the preferential discourse status occupied by this referent. In such contexts, *so'n* does not provide further descriptive material to the sentence, as it could be left out without altering the meaning of the

sentence, but rather functions as a discourse-structuring device, contributing to the referential structure of the discourse.

(14) Draußen war es warm. Peter<sub>1</sub> hielt es zu Hause einfach nicht mehr aus und ging in die Stadt. Auf dem Weg dorthin sah er<sub>1</sub> so'n Jungen<sub>2</sub> einen Laden betreten. Er<sub>2</sub> hatte tätowierte Arme und schaute sich vorsichtig um. Als er<sub>2</sub> sich unbeobachtet glaubte, griff er<sub>2</sub> sich ein Paar Lederhandschuhe und stopfte sie unter seinen Pullover. Er<sub>2</sub> ging dann ruhig aus der Tür als ob nichts wäre.

'It was warm outside. Peter<sub>1</sub> didn't resist staying home and went downtown. On his way there he<sub>1</sub> saw so'n boy<sub>2</sub> entering a store. He<sub>2</sub> had tattooed arms and looked cautiously around. When he<sub>2</sub> felt unwatched, he<sub>2</sub> grasped a pair of leather gloves and stuffed them under his sweater. He<sub>2</sub> then went out of the store as if nothing had happened.'

I argue that *so'n* is an indefinite determiner that fulfils the function of a discourse-structuring device. In other words, *so'n* is used to focus or organize the discourse structure in such a way as to draw attention to particular properties of a referent. In this function, the hearer is instructed to establish a permanent discourse representation for the referent, as the distinctive characteristics about the referent in question will be elaborated upon in the subsequent discourse. I argued in Chapters 2 and 3, that the discourse structuring potential of a particular expression is reflected by two textual characteristics of its associated referent: (i) referential persistence and (ii) topic shift potential. These two properties will be empirically validated in the study presented in Section 4.3.

In its out-of-the-blue use, *so'n* resembles the English determiner *this* in its use. Interestingly, exactly like its English homologous expression, the use of *so'n* in indefinite contexts comes close to that of the simple indefinite article. One piece of evidence showing that *so'n* is indefinite comes from its incompatibility with typical definite contexts. As illustrated in examples (15a-c), when substituting *so'n* for the definite article in genitive construction, superlatives, or typical definite expressions like "the moon", the sentences are rendered infelicitous. These patterns show that despite being a demonstrative expression, noun phrases headed by *so'n* cannot be grouped with definite noun phrases.

- (15) (a) Ich habe die/ \**so'ne* Mutter meiner Freundin Teo getroffen, die in LA lebt.  
 'I met the/ \**so-a* mother of my friend Teo, who lives in LA.'
- (b) Ich habe den/ \**so'nen* größten Mann der Welt in Kalifornien getroffen.  
 'I met the/ \**so-a* tallest man in the world in California.'
- (c) Der/ \**so'n* Mond war nirgends zu sehen.  
 'The/ \**so'n* moon was nowhere to be seen.'

Further evidence for *so'n* being indefinite comes from existential-*there* sentences and *have*-sentences, which represent good testing grounds for indefiniteness. Indefinite *so'n* can appear in existential-*there* constructions in German, as illustrated in (16a), or as the object of inalienable *have*, as in (17a), patterning with the indefinite article, as exemplified in (16b) and (16b). When we attempt this with simple definite noun phrases as illustrated in (16c) and (17c), we find the resulting sentences unacceptable.

(16) *Existential there-sentences*

- (a) Als ich in die Schule ging, da gab's *so'nen* Jungen in meiner Klasse [...].
- (b) Als ich in die Schule ging, da gab's einen Jungen in meiner Klasse [...].
- (c) Als ich in die Schule ging, \*da gab's den Jungen in meiner Klasse [...].  
 'When I went to school, there was *so-a/ a/ \*the* boy in my class [...].'

(17) *Inalienable have-sentences*

- (a) Diese Stadt hat *so'n* Museum (das mich interessiert).
- (b) Diese Stadt hat ein Museum (das mich interessiert).
- (c) Diese Stadt hat \*das Museum (das mich interessiert).  
 'This city has *so-a/ a/ the* museum (that interests me).'

The tests presented above have been used in the literature to underline the indefinite status of *so'n* in out-of-the-blue contexts (i.e. in contexts in which the noun phrase headed by *so'n* neither does involve a pointing gesture, nor does it refer back to an antecedent expression). In light of the patterns presented above, I take *so'n* noun phrases used in out-of-the-blue contexts to be indefinite determiners that differ in distribution, function and effects from the standard (deictic and anaphoric) demonstrative use of *so*. Importantly,

however, *so'n* cannot be equated directly with standard indefinite noun phrases either. In what follows, I will show that *so'n* indefinites differ from simple indefinite noun phrases at least with respect to two aspects: first, *so'n*-indefinites introduce hearer-new and discourse-new referents with particular referential properties, and second, *so'n*-indefinites signal the discourse structuring potential of their associated referents.

#### 4.1.4 Summary

In Section 4.1, I discussed different adnominal uses of *so'n*. The most common uses of *so* as a demonstrative expression were reviewed, namely its deictic, anaphoric and ‘out-of-the-blue’ use. *So* was shown to function as a modifier that can combine with gradable as well as non-gradable expressions. Every time *so* combines with an expression that can be graded (the (a), (c), (e)) uses in Table 4.3), it induces a degree reading of its associated referents. When the expression headed by *so* is non-gradable (as in (b), (d), (f) in Table 15), *so* refers to a property that shifts the original denotation of the expression.

Table 4.3: Different uses of *so* in German with their characteristics

Use		Type of expression	Characteristics
(a)	Deictic	Gradable	Picks up a degree
(b)	Deictic	Non-gradable	Picks up a property
(c)	Anaphoric	Gradable	Picks up a degree
(d)	Anaphoric	Non-gradable	Picks up a property
(e)	Intensifier	Gradable	Shifts the standard (upwards or downwards)
(f)	Hedging	Non-gradable	Shifts/ extends the denotation
(g)	Indefinite	Non-gradable	Discourse-structuring device

By the end of the last Section, I introduced a novel use of *so* as an indefinite determiner, which can be found in contexts in which the noun it precedes is a non-gradable indefinite noun phrase. I suggested that previous accounts of *so'n* are not sufficient and fail to fully capture the conditions under which this indefinite is used. In the next sections I show in more detail that *so'n* fulfils two functions (which sets this expression apart from other German indefinite determiners): (i) it introduces a new referent with particular referential

properties in the discourse and (ii) it simultaneously signals that this indefinite noun phrase will give structure to the subsequent discourse.

## 4.2 Referential properties of indefinite *so'n*

We have seen so far that German indefinite *so'n* introduces brand-new referents in the discourse, similar to the simple indefinite article *ein* ('a/an'). It is, however, well established in the literature that indefinites headed by *ein* ('a/an') have many semantic-pragmatic properties that distinguishes them from other types of indefinites, as for example German *ein gewisser* ('a certain') and *ein bestimmter* ('a particular') (Ebert, Ebert and Hinterwimmer, in press). This Section compares the referential properties of indefinite *ein* with the referential properties of indefinite *so'n*. In particular, the question is whether we observe differences between the two types of indefinites in their scope-taking behaviour in various syntactic constructions? As introspective data on the interpretation of the two indefinites points to subtle and often conflicting results, I will substantiate these judgements with experimental evidence.

### 4.2.1 Different types of readings of indefinite *so'n*

This Section briefly reviews observations that concern the German simple indefinite *ein* ('a/an') in neutral contexts and in the context of scope islands and compares them with the interpretations of indefinites headed by *so'n* in the same contexts. As already discussed in Chapters 2 and 3, the notion of specificity is commonly used in the semantic literature as a referential notion, encompassing at least three separate phenomena: (i) referential specificity (in opaque contexts), (ii) scopal specificity and (iii) epistemic specificity<sup>15</sup>. Let us start by considering examples that show indefinite *so'n* to be scopally specific. First, in contexts involving existential quantifiers, simple indefinite noun phrases are generally interpreted as being ambiguous between a specific and a non-specific reading. The indefinite noun phrase *a book* in (18a) can be interpreted in two ways, as the referent headed by the simple indefinite article displays two readings. On the contrary, the indefinite noun phrase headed by *so'n* in (18b) is compatible only with an interpretation in

<sup>15</sup> For a more detailed discussion of specificity, see Chapter 3.2.1.

which the referent of the noun phrase does not covary with the referent of the universal quantifier (i.e. the specific reading).

(18) *Extensional universal quantifiers*

- (a) Jeder meiner Kollegen hat ein Buch von Eliade gelesen.

‘Each of my colleagues read a book by Eliade.’

*Reading*<sub>1</sub>: only one book overall [no covariation with colleagues]

*Reading*<sub>2</sub>: possibly many books overall [covariation with colleagues]

- (b) Jeder meiner Kollegen hat *so'n* Buch von Eliade gelesen.

‘Each of my colleagues read *so-a* book by Eliade.’

*Reading*: only one book overall [no covariation with colleagues]

Further evidence for the different behaviour of *so'n* indefinites comes from scope islands, which are syntactic structures that disallow wide scope readings for quantifiers. However, unlike quantificational phrases, *ein* indefinite noun phrases can scope out of islands such as conditionals and relative clauses (Fodor and Sag 1982, among others), while simultaneously allowing for intermediate scope above the island, but below a higher quantifier (Farkas 1981, Ruys 1992, Abusch 1994). Moreover, simple indefinites are not restricted to readings inside the scope island. Example (19a) exemplifies the different scope readings of the German simple indefinite *ein* (‘a/an’): on the wide scope reading, (*Reading*<sub>1</sub> in (19a)) the indefinite scopes above all other scope bearing elements in the sentence; on the intermediate scope reading (*Reading*<sub>2</sub> in (19a)) the indefinite scopes outside the relative clause but under the quantifier; and on the narrow scope reading (*Reading*<sub>3</sub> in (19a)) the indefinite scopes locally as it does not scope over the relative clause and the quantifier. In comparison to the various scope interpretations of indefinite *ein*, the narrow scope and intermediate scope readings of the *so'n* indefinite are disallowed, as illustrated in (19b).

(19) *Scope island escaping*

- (a) Jeder Junge hat jedes Buch gelesen, das ein Lehrer empfohlen hatte.

‘Every boy read every book that a teacher assigned.’

*Reading*<sub>1</sub>: a teacher > every boy > every book [wide]

*Reading*<sub>2</sub>: every boy > a teacher > every book [intermediate]

*Reading*<sub>3</sub>: every boy > every book > a teacher [narrow]

(b) Jeder Junge hat jedes Buch gelesen, das *so'n* Lehrer empfohlen hatte.

‘Every boy read every book that *so-a* teacher assigned.’

*Reading*<sub>1</sub>: a teacher > every boy > every book [wide]

The distribution of indefinites headed by *so'n* in sentences containing higher quantifiers underlines their preference for specific readings once more. Example (20b) shows that indefinites headed by *so'n* are unable to scope under a higher quantifier like *every*. The distributive reading is ruled out even in cases in which bound variables force this particular reading of the noun phrase preceded by *so'n*. Indefinite noun phrases headed by the simple indefinite article *ein*, on the contrary, allow for an intermediate scope reading in the same context, as witnessed in (20a).

(20) *Intermediate scope readings for ein(e) indefinites*

(a) [Jeder Junge]<sub>i</sub> hat ein Buch gesehen, das er<sub>i</sub> mochte.

‘Every boy read a book that he enjoyed.’

Continuation: Ruhl read “The Great Gatsby”, Paul “Harry Potter”, [...].

(b) [Jeder Junge]<sub>i</sub> hat *so'n* Buch gesehen, das er<sub>i</sub> mochte.

#Continuation: Ruhl read “The Great Gatsby”, Paul “Harry Potter”, [...].

(adapted from Ionin 2006: 202)

Up to this point, indefinites marked by *so'n* have been shown to display a preference for scopally specific readings, however, they are sensitive towards referential specificity as well. Referential specificity is defined in terms of the interpretation of the indefinite outside the scope of an operator, such as negation, an intensional verb or a modal (Dahl 1970, Karttunen 1976). This type of specificity expresses a contrast between a reading that allows for an existential entailment, and one that does not. The simple indefinite *ein* in sentences containing a negation, as in (21a) or an intensional operator, such as *want* in (22a) are ambiguous between a referentially specific and a non-specific reading. The different readings can be brought about by certain continuations. For example, continuation<sub>1</sub> in (21a) and (22a) bring about the non-specific interpretation of the indefinite *ein Buch* (‘a book’) and *einen Film* (‘a movie’) respectively, whereas continuation<sub>2</sub> in

(21b) and (22b) reverses these preference in favour of the referential specific reading. Indefinites marked by *so'n* in contexts containing a negator, as in (21b), or an intensional operator, as in (22b), are generally interpreted as being referentially specific, however, it seems that there are no clear judgements about whether the non-specific reading is strictly disallowed in these contexts or not.

(21) *Negation*

- (a) Es is nicht der Fall, dass Eva ein Buch von Eliade gekauft hat.  
 'It is not the case that Eva bought a book by Eliade.'  
 Continuation<sub>1</sub>: but there is no such book [referential non-specific]  
 Continuation<sub>2</sub>: there is such a book [referential specific]
- (b) Es is nicht der Fall, dass Eva *so'n* Buch von Eliade gekauft hat.  
 'It is not the case that Eva bought *so-a* book by Eliade.'  
 Continuation<sub>1</sub>: there is such a book [referential specific]  
 Continuation<sub>2</sub>: there is no such book [??referential non-specific]

(22) *Intensional operators*

- (a) Eva will einen Film über Eliade sehen.  
 'Eva wants to watch a movie about Eliade.'  
 Continuation<sub>1</sub>: but I can't find one [referential non-specific]  
 Continuation<sub>2</sub>: but I can't find it [referential specific]
- (b) Eva will *so'nen* Film über Eliade sehen.  
 'Eva wants to watch *so-a* movie about Eliade.'  
 Continuation<sub>1</sub>: but I can't find it [referential specific]  
 Continuation<sub>2</sub>: but I can't find one [??referential non-specific]

Let us consider one more example, in which the interpretation of indefinite *so'n* in sentences containing intensional operators becomes less straightforward as it might have seemed at first blush. In (23), both the indefinite *so'n* as well as indefinite *ein* ('a/an') are compatible with a continuation sentence that denies the identity of their referents, contrary to the intuition that a referent must be specific so that the speaker may refer to it by means of indefinite *so'n*.

- (23) Carrie will *so'nen/* einen Prinz auf einen weißen Ross heiraten. Aber sie hat noch keinen kennengelernt.  
 ‘Carrie wants to marry *so-a/* a prince on a white horse. But she has not met one yet.’

Even if the referent associated to the noun phrase *so'nen Prinz auf einen weißen Ross* (‘*so-a* prince on a white horse’) must take a non-existential and non-specific reading, the occurrence of *so'n* heading the indefinite noun phrase is still possible. It could be argued that the indefinite noun phrase actually refers to the ideal man *Carrie* pictures in her imagination (i.e. a blonde man who has a big belly and is a UPS driver). Thus, for *Carrie*, the referent of indefinite *so'n* is unique and clearly distinguishable from other men, even if she did not meet him in the actual world. Furthermore, the referent associated with indefinite *so'n* seems to play a pivotal role in the sentence, as more information about him has to follow. It seems that the speaker has to justify the use of a more elaborate form (i.e. the use of indefinite *so'n* instead of a simple indefinite) by further elaborating on the referent introduced in this way. We will return to this intuition in Section 4.3, where I report on the results from a story continuation study that investigates the discourse contribution of indefinites headed by *so'n* in terms of referential persistence and upcoming topic shift.

Returning to the third type of specificity, it is generally known that simple indefinites preceded by *ein* display an additional ambiguity in readings in contexts that do not contain operators. Scopal and referential specificity are closely related, but nevertheless distinct from the notion of epistemic specificity, which is also known as “speaker knowledge” or “identifiability” (Fodor and Sag 1982, Farkas 1994, von Heusinger 2002). An epistemic specific indefinite is considered to refer to an entity indistinguishable or known by the speaker or another salient agent in the discourse. The indefinite noun phrase in (24a), *a student in the Syntax class* is ambiguous between an epistemic specific reading, in which the speaker knows or is able to identify the individual associated with the indefinite noun phrase (i.e. the person that cheated on the exam) and a non-specific reading, in which the identity of the referent is not known to the speaker. On the contrary, indefinites headed by *so'n*, like in (24b), seem to pattern more like indefinite *this* or *a certain* in English, as they are more readily interpreted as referring to a specific referent. However, no clear judgements can be

made on the non-availability of non-specific readings of indefinites preceded by *so'n* in transparent contexts.

(24) *Epistemic specificity*

- (a) Teodora hat im Krankenhaus mit einem Arzt gesprochen.  
 ‘Teodora talked to a physician in the hospital.’  
 Reading<sub>1</sub>: it was Dioscoride [epistemic specific]  
 Reading<sub>2</sub>: I don’t know who it was [epistemic non-specific]
- (b) Teodora hat im Krankenhaus mit so’nem Arzt gesprochen.  
 ‘Teodora talked to so-a physician in the hospital.’  
 Reading<sub>1</sub>: it was Dioscoride [epistemic specific]  
 Reading<sub>2</sub>: I don’t know who it was [? epistemic non-specific]

The broad picture that emerges from the discussion of the referential properties of indefinite *so'n* is that: (i) it differs from indefinite *ein* in its scope-taking behaviour, and (ii) the availability of specific vs. non-specific readings seems to vary with the context. While indefinite *so'n* shows a robust preference for wide scope readings in sentences containing scope islands, its distribution and interpretation in sentences containing intensional operators, negation and in transparent or neutral contexts is less consistent. In the latter contexts, indefinite *so'n* seem to allow a non-specific interpretation as well. Theories on specificity often rely on fairly subtle judgments, when investigating the behaviour of indefinite noun phrases. As introspective data is not enough to account for the readings of indefinites in different contexts, I conducted an experimental study that investigated the distribution of referential specific vs. referential non-specific readings of indefinite noun phrases headed by *so'n* in sentences containing intensional operators and negation. The next Section summarizes the relevant findings of the experiment on the referential readings of *so'n* indefinites compared to their simple indefinite counterparts headed by *ein* (‘a/an’).

#### 4.2.2 Experimental evidence on the referential readings of indefinite *so'n*

In order to shed light on the issues pointed out above, this experiment explores whether indefinites headed by *so'n* show a systematic preference for wide-scope readings, or

whether they are compatible with narrow-scope readings as well. The concrete aims of the experiment are to test whether referentially specific readings in the scope of negation and the intensional operator *will* are available to *so'n* indefinites alongside functional readings<sup>16</sup>. The second aim of this experiment is to compare these data to the readings available for the simple indefinite *ein* ('a/an').

### Materials and Procedure

Fifty-one native speakers of German undergraduate students participated in the experiment. They were recruited from general education courses in linguistics, and had no prior exposure to formal semantics. It took about 20 minutes for a participant to complete the experiment. The total set of stimuli comprised 48 items: 24 test items and 24 fillers. Each test item consisted of two sentences. In the first sentence, I manipulated the type of embedding (negation vs. intensional operator *will*) and the type of referring expression used to introduce a referent (*so'n*-indefinite vs. *ein*-indefinite NP vs. simple definite NP), which resulted in six conditions, as illustrated in Table 4.4<sup>17</sup>.

Table 4.4: Sample experimental items in the six conditions with English translations

Type of referring expression	Negation	Intensional operator <i>will</i>
<i>so'n</i> indefinite	Es is nicht der Fall, dass Maria <b>so'n reiches afrikanisches Land</b> besucht hat. <i>'It's not the case that Mary visited so-a rich African country.'</i>	Maria will <b>so'n reiches afrikanisches Land</b> besuchen. <i>'Mary wants to visit so-a rich African country.'</i>
<i>ein</i> indefinite	Es is nicht der Fall, dass Maria <b>ein reiches afrikanisches Land</b> besucht hat. <i>'It's not the case that Mary visited a rich African country.'</i>	Maria will <b>ein reiches afrikanisches Land</b> besuchen. <i>'Mary wants to visit a rich African country.'</i>
Simple definite	Es is nicht der Fall, dass Maria <b>das reiche afrikanische Land</b> besucht hat. <i>'It's not the case that Mary visited the rich African country.'</i>	Maria will <b>das reiche afrikanische Land</b> besuchen. <i>'Mary wants to visit the rich African country.'</i>

<sup>16</sup> Note that epistemic modality can be expressed by a variety of verbs, adjectives and other modal auxiliaries, but because of lack of space the present study focuses only on the interaction between *so'n* and *will*. Accordingly, the findings of this study only indicate tendencies and preferences about the readings of *so'n* under negation and the modal *will*.

<sup>17</sup> Thank you to Edgar Onea and Alexeci Chibakov for their collaboration on the design of this study.

The methodology used in this experiment was a web-based Yes/No task. Participants were presented with one sentence at a time and were then asked to judge by choosing between one of the three categories: YES, NO and I DON'T KNOW, whether a second sentence matched the content of the first. The Yes/No task identifies the relationship between the two test sentences and the three answer possibilities presented to the participant. For example, after reading one of the stimulus items given in Table 4.4, participants were asked whether a particular rich African country does exist or not, i.e. *Folgt daraus, dass es ein bestimmtes reiches afrikanisches Land gibt?* ('Does it follow from this that there is a particular rich African country?'). A YES answer indicates that the existential presupposition holds, and that the critical noun phrase is interpreted as being referential, or referring to a particular referent. A NO answer shows that the critical noun phrase allows for a non-referential interpretation of the referent, whereas the third answer option, I DON'T KNOW, indicates that neither the referential nor the non-referential reading of the critical noun phrase is accessible to the participant. It is important to note at this point that in this experiment, I use the notion of *referential specific* for instances in which an (in)definite noun phrase is interpreted outside the scope of a scope-bearing element, and the notion of *referential non-specific* for cases in which the critical noun gets its interpretation within the scope of a scope bearing operator. I will not go into a discussion of the mechanisms that could be used to derive these two readings.

The experiment was fully counterbalanced across the two versions, so that each participant saw each test item in the negated version, and in the version with the intensional operator *will*. Each participant saw 12 critical items, as presented in Table 4.5. According to the numbers sketched out below, a participant saw two items of the negation+*so'n* type, intensional operator+*so'n* type, two items of the negation+simple indefinite type, intensional operator+simple indefinite article, and so on.

Table 4.5: The distribution of test items per participant

Conditions	Negation	Intensional operator <i>will</i>
<i>So'n</i> indefinite	2	2
<i>Ein</i> -indefinite	2	2
Definite noun phrase	2	2
Total	6	6

## Predictions

Following the observations made in the previous Section and based on the findings presented in the literature on the readings of noun phrases headed by the simple indefinite article *ein* ('a/an'), the predictions with respect to the referential readings of the critical noun phrases are as follows: When an indefinite noun phrase occurs in the scope of an intensional operator or negation, native speakers of German will interpret *so'n* indefinites and *ein* indefinites differently. More precisely, *so'n* indefinites will show a preference for referential specific readings, while *ein* ('a/an') indefinites will be ambiguous between a referential specific and non-specific reading. Definite noun phrases will show the strongest affinity for referential readings and will thus represent the baseline condition in this experiment.

## Results and discussion

Two participants were excluded from the analysis because they were not native speakers of German. This left a total of 49 responses for each of the two versions of the experiment (i.e. negation and intensional *will*). Participants' answers were analysed by two native speaker coders who worked independently. Disagreements were resolved by discussion.

I will first report on the findings of the referential readings of the critical referents in the context containing negation and then discuss the data on the intensional operator condition. Table 4.6 summarizes the results for the three types of referring expression tested in the context containing negation. The Table indicates the absolute and mean values of the responses for each condition.

Table 4.6: Mean and absolute values for the negation-condition

Answer NP-type	YES [specific reading ]	NO [non-specific reading]	I DON'T KNOW
<i>so'n</i> indefinite	117 (196) 60%	72 (196) 37%	7 (196) 3%
<i>ein</i> indefinite	44 (196) 22.5%	146 (196) 74.5%	6 (196) 3%
definite NP	175 (196) 89%	20 (196) 10%	1 (196) 1%

Several patterns can be observed based on the results presented above. First, in the *so'n*-condition, more than half of the participants chose a YES answer, which indicates a clear preference for a referential specific reading of the indefinite noun phrases marked in this way. However, contrary to the initial predictions, not all indefinites headed by *so'n* show the availability for referential specific readings. Second, the prediction with respect to the referential readings of *ein* indefinites was that these indefinites allow for both readings. The data in Table 4.6 shows that simple indefinite *ein* embedded in the same contexts, results more frequently in NO answers, indicating a robust preference for non-specific readings. This observation raises the question about the contextual factors that might play a role as well in determining the referential status of *ein* ('a/an') indefinites. Finally, definite noun phrases show the predicted preference for referential readings, as in 89% of the cases, definite noun phrases were interpreted as referring to a specific referent. In sum, the predictions stated at the beginning of this Section were partially confirmed by the experimental findings on the readings of different types of expressions embedded in a sentence containing negation.

The mean and absolute values for the three types of referring expressions (*so'n* indefinite, *ein* indefinite and simple definite noun phrase) for the context containing the intensional operator *will* are reported in Table 4.7. The findings show that, when given a sentence with a *so'n* indefinite, participants opted for a YES response 81% of the time, again indicating a preference for the specific reading of the indefinite marked in this way. Compared to the results from the negation-condition and in light of our initial predictions, it seems that while *so'n*-indefinites allow for non-specific readings in the intentional operator condition as well, they do so in less cases than in the negation-condition.

Table 4.7: Mean and absolute values for the intensional operator *will* condition

Answer NP-type	YES [specific reading]	NO [non-specific reading]	I DON'T KNOW
<i>so'n</i> indefinite	159 (196) 81%	35 (196) 18%	2 (196) 1%
<i>ein</i> indefinite	62 (196) 31.5%	130 (196) 66%	4 (196) 2%
definite NP	193 (196) 98.5%	3 (196) 1.5%	0

Indefinite noun phrases headed by *ein* ('a/an') resulted again mostly in NO answers (66%), indicating that the non-specific reading is more accessible than the specific one. However, the non-specific reading is less available compared to the negation condition. Definite noun phrases show an even stronger bias for YES answers compared to the negation condition. The predictions stated above are partially confirmed as *so'n* indefinites pattern more like simple definite noun phrases showing a preference for specific readings, while *ein* indefinites show the opposite preference.

Summarizing, a general observation that can be made when we compare participants' responses for each condition. The overall patterns observed in the context of the intensional operator *will* (Table 4.7) are stronger and thus more robust than with negation (i.e. Table 4.6). This effect could be due to the fact that negated information is less accessible than non-negated information (MacDonald and Just 1989). A similar observation was made within discourse representation theory (Kamp 1981), which assumed that discourse referents introduced in a negated context are represented in a special substructure, which makes these referents non-accessible in the subsequent discourse and unlikely to be subsequently pronominalized. However, this generalization was shown to be true for indefinite noun phrases, which *introduce* new discourse entities, but not for definite noun phrases, which *refer* to entities that are generally not discourse-new. According to this view, it is no surprise that the negation context has no major effect on the interpretation of the target items in the definite noun phrase condition.

A somewhat surprising observation is the overall stronger preference of *ein* indefinites for referential non-specific readings. These results parallel Geurts (2002) observation with respect to the fact that indefinite noun phrases are non-specific by default as they introduce new information. This is particularly true for sentences involving negation, which showed lower rates of non-specific readings of indefinites than sentences containing the intensional operator *will*. This remark extends to contexts in which the simple indefinite noun phrase is modified by adjectives (e.g. a rich African country). Even if modification generally facilitates the specific interpretation of noun phrases (Chierchia 2001), it seems that this aspect did not dramatically influence the general preference for non-specific readings of the simple indefinite noun phrases. If these findings receive confirmation from a more extensive experimental investigation, then we might expect to find an important contribution of the pragmatic context in the interpretation of indefinite noun phrases.

### 4.2.3 Summary

In this Section we observed that indefinites headed by *so'n* have particular referential properties that set them apart from simple indefinite noun phrases headed by *ein* ('a/an'). Table 4.8 summarizes the contexts and different readings of indefinites preceded by *ein* and by *so'n* in contexts that contain different operators.

Table 4.8: Referential readings of indefinit *so'n* and indefinite *ein*

	intensional	negation	universal quantifier	island escaping	functional	epistemic
<i>Ein</i> indefinites	specific/non-specific	specific/non-specific	covariation/no covariation	yes/ no	yes	specific/non-specific
<i>So'n</i> indefinites	specific	specific	no covariation	yes	no	specific/non-specific

Apart from their referential properties, which can be noticed in contexts that involve operators at sentence level, indefinite noun phrases headed by *so'n* have discourse properties, which become particularly evident in “neutral” contexts (i.e. such contexts that do not contain any operators). This observation is elaborated upon in the following Section.

## 4.3 Discourse properties of indefinite *so'n*

The previous Section established the preference of indefinite *so'n* for referential readings. In this Section, I will show that specificity is not enough to account for the distribution and discourse behaviour of these indefinites. In a first step I argue that in comparison to simple indefinite noun phrases, indefinites headed by *so'n* come with a felicity condition that their referents must be elaborated upon in the subsequent discourse. In light of the findings from a story-continuation experiment I show that referents headed by *so'n* are not only referentially more persistent, but are also more prone to shift the topic in the subsequent discourse.

### 4.3.1 Indefinite *so'n* and immediacy of reference

Recall example (13), repeated here as (25), which represents a neutral context, as it does not contain any sentence operator. Indefinite *so'n* takes wide scope by default in this case, however, a more striking observation is the fact that *so'n* introduces a new referent in the discourse (i.e. the teacher) about which further information is given in the subsequent sentence.

- (25) Also, da gibt's *so'nen Lehrer<sub>i</sub>* in meine Schule, in den<sub>i</sub> ich verliebt bin.  
 'So, there is so-a teacher<sub>1</sub> in my school whom<sub>1</sub> I'm in love with.'

In the story-fragment given in (26b) the presence of *so'n* is improper and unjustified, as no information about the referent preceded by *so'n* (i.e. the book) is given in the immediately following sentence. In contrast, whenever this condition is fulfilled and the referent headed by *so'n* is elaborated upon in the following discourse, as in (26c), the sentence is felicitous. The simple indefinite article, however, does not come with a felicity condition that requires its referent to be mentioned in the immediately following sentence, as (26a) and (26b) show.

- (26) (a) Gestern war ich in der Stadt und hab mir ein Buch gekauft. Dann bin ich heimgerannt und hab gegessen und bin ins Bett gegangen.  
 'Yesterday I went downtown and bought a book. Then I ran home and ate supper and went to sleep.'
- (b) ??Gestern war ich in der Stadt und hab mir *so'n* Buch gekauft. Dann bin ich heimgerannt und hab gegessen und bin ins Bett gegangen.  
 '??Yesterday I went downtown and bought so-a book. Then I ran home and ate supper and went to sleep.'
- (c) Gestern war ich in der Stadt und hab mir *so'n/* ein Buch gekauft. Dann bin ich gleich heimgerannt und hab's gelesen, und es war wirklich großartig.  
 'Yesterday I went downtown and bought so-a/ a book, and I ran home and read it, and it was truly a terrific book.' (adapted from Givón 1989: 182)

Further evidence for “immediacy of reference” (i.e. reference to an entity in the immediately following sentence) as an essential (but not necessary) condition for the use of indefinite *so'n* (as opposed to the indefinite article *ein*) comes from the observation that *so'n* often combines with a number of constructions that serve to further elaborate on a recently introduced discourse referent. In example (27), the restrictive relative clause comes with non-parenthetical information, as the information is relevant for the identification of the head noun referent. Left dislocations and adnominal modifiers are a common means to provide further information about the referent in question, as (28) and (29) show.

(27) Relative clauses:

Wir haben *so'n Lehrer, der* zu jedem Thema 1000 Folien hat<sup>18</sup>.  
 ‘We have so-a teacher, who has 1000 slides for each theme.’

(28) Left dislocations:

*So'n Typ, der bei mir* aus Versehen angerufen hat, er hat gefragt, ob wir Freunde werden sollen, ihm hätte meine Stimme gefallen<sup>19</sup>.  
 ‘So-a guy, who called me by accident, asked whether we could be friends, he said he liked my voice.’

(29) Adnominal modifiers:

Wir gehen wahrscheinlich ins 3D-Kino und schauen *so'n Film über* die wunderbare Welt der Ozeane [...] <sup>20</sup>.  
 ‘We are probably going to a 3D-theatre and watch so-a movie about the wonderful world of the oceans.’

We have seen so far that the felicitous use of *so'n* depends upon its referent being picked up in the following discourse. Consider examples (30) and (31) in which the referent associated to indefinite *so'n* is highly continuous in the subsequent discourse.

<sup>18</sup> <http://Section.traum-projekt.com/forum/11-einfach-so/67238-kosten-auf-folie-drucken-lassen.html> [viewed on February 2011].

<sup>19</sup> <http://famseite.spiderbee.com/art/stories/lily/3rd.htm> [viewed on February 2011].

<sup>20</sup> [http://verben.texttheater.net/Hauptmann\\_und\\_noch\\_fünf\\_Mann\\_auf\\_Landfahrt](http://verben.texttheater.net/Hauptmann_und_noch_fünf_Mann_auf_Landfahrt) [viewed on January 2011].

- (30) Da war so'n Typ, der war so schnell, das gibt's gar nicht. Ich konnte überhaupt nichts machen. Der war weg, bevor ich draußen war.<sup>21</sup>  
 'There was so-a guy, he was so quick, it's unbelievable. I couldn't do anything. He dissapeared by the time I was ouside.'
- (31) [...] aber das witzige war, ich hab da **so'nen mann** getroffen, der das gleiche problem hatte [...] von seiner familie ist die maschiene kaputt gegangen alo musste er es reparieren kommen. war auch zum ersten mal da [...]<sup>22</sup>.  
 '[...] but the funniest was that I met so-a man, who had the same problem [...] his family's car broke so he had to come and repair it. (he) was here for the first time'.

The examples above suggest that indefinite noun phrases headed by *so'n* tend to be taken up in the subsequent discourse. In this sense, the felicity condition on the use of indefinite *so'n* resemble the one proposed for English indefinite *this*, as discussed discussed in Chapter 3. Recall that the experimental evidence presented in the previous chapter showed that referents associated with indefinite *this* are: (i) more likely to be picked up in the following discourse and (ii) more prone to become topics, compared to indefinite noun phrased headed by the indefinite article. The question that arises at this point is whether indefinites headed by *so'n* show the same discourse structuring potential in terms of referential persistence and upcoming topic shift. The experiment reported in the next Section investigates the discourse contribution of indefinites headed by *so'n* and by the simple indefinite article along the same lines.

#### 4.3.2 The discourse structuring potential of indefinite *so'n*

Up to this point I showed that *so'n* fulfils at least two functions: on the one hand it introduces a new referent in the discourse universe and, on the other hand, it seems to signal that the referent is important or salient in the discourse, as more information about it is expected to follow. The first experiment reported in this Chapter brought evidence in favour of the observation that *so'n* indefinites show a preference for referential specific readings in certain contexts. It was already noted that the use of *so'n* is more often found in

<sup>21</sup> <http://ww3.ac-poitiers.fr/allemand/docs/Cinema/Schinkel.htm> [viewed on June 2011].

<sup>22</sup> <http://line-in-the-usa.jimdo.com/mein-blog/> [viewed in January 2011].

the informal register. Besides the study conducted by Wiese (2009) on Kietzdeutsch (i.e. a language variety found in the Berlin-area of Germany), there is no other study that systematically investigated the distribution of *so'n* in a larger corpus of German. Moreover, no previous study has investigated the discourse effects of *so'n* in giving structure to the subsequent text. In this Section, I report on a study on the discourse structuring potential of indefinites headed by *so'n* by using the textual methods introduced in Chapter 2, namely: (i) referential persistence and (ii) topic shift potential. Independently of these two metrics, I will also look at (iii) the type of referring form of the first anaphoric expression used to pick up these referents. In light of the findings from the study on the discourse effects of English indefinite *this*, the following three predictions about the discourse structuring potential of indefinites headed by *so'n* can be made. Despite the fact that indefinite *this* and indefinite *so'n* have different origins and functions, I expect them to pattern alike with respect to their discourse structuring potential.

*Prediction 1 (referential immediacy and referential persistence)*

- (a) Referents associated with noun phrases headed by *so'n* will be immediately picked up in the subsequent discourse (compared to their simple indefinite counterparts), as indicated by the discussion presented in Section 4.3.1.
- (b) If noun phrases headed by *so'n* are characterized by a discourse structuring potential, then their associated referents will show a higher frequency of subsequent mention (i.e. referential persistence) compared to their simple indefinite counterparts.

*Prediction 2 (topic shift potential)*

Referents of noun phrases preceded by indefinite *so'n* will more readily shift the topic in the subsequent discourse compared to their simple indefinite noun phrases (see the discussions in Section 2.4.2).

*Prediction 3 (referential type of the first anaphoric expression)*

If indefinite *so'n* is a discourse-structuring device that comes close to English indefinite *this*, then the prediction is that the German indefinite determiner will not impact the type of the first referring expression used to take up the referent (see discussion in Chapter 3.3) in another way compared to its simple indefinite

counterpart. Moreover, the presence of one of the indefinites will not contribute to a higher rate of subsequent pronominalization of their associated referents.

### 4.3.2.1 The experimental setup

#### *Participants*

Twenty-two native speakers of German participated in the experiment. They were recruited from different regions of Germany. It took about 30 minutes to complete the experiment. All participants were naïve with respect to the purposes of the experiment and did not take part in the first experiment on the referential properties of indefinite and definite noun phrases presented in Section 4.2.2.

#### *Design and Procedure*

The current study investigates the processing and comprehension processes involved while referring to understand the discourse contribution of referents introduced by indefinite *so'n* compared to their simple indefinite counterparts. The methodology used is the sentence-continuation task introduced in Chapter 2. Participants were given three-sentence stories and were asked to add five logical and natural-sounding continuation sentences to each of them. It is important to note that I used an open-ended sentence-continuation task that did not restrict participants' continuations in any way (i.e. by a pronoun prompt, a conjunction, etc.). As *so'n* is mostly used in conversational, casual German, the target and filler items were also designed to have a casual register.

#### *Materials*

The sentence-continuation experiment comprised 6 test items and 7 fillers. Each target story contained a three-sentence story. I manipulated the form of the direct object in the last sentence, which resulted in two conditions: one in which the direct object was realized as a *so'n*-marked indefinite (e.g. *so'nen Jungen*, 'so-a boy' in Table 4.9) and one in which the referent was realized as an indefinite noun phrase headed by the simple indefinite article (e.g. *einen Jungen*, 'a boy' in Table 4.9). This was a between-subjects manipulation, in that a participant either did a version of the study where all three targets contained simple indefinites, or a version where all three targets contained *so'n* indefinites. I will refer to these two versions as the '*so'n*-condition' and the '*ein*-condition'. The full list of target items used for each condition is given in Table 4.9.

Table 4.9: Experimental items (T1-T3) used in the sentence continuation experiment

TI1: <i>so'n</i> -condition	TI1: <i>ein</i> -condition
<p><i>Draußen war es warm. Peter<sub>1</sub> hielt es zu Hause einfach nicht mehr aus und ging in die Stadt. Auf dem Weg dorthin sah er<sub>1</sub> so'nen Jungen<sub>2</sub> einen Laden betreten.</i></p> <p>'It was warm outside. Peter<sub>1</sub> couldn't stay home and went downtown. On his way there he<sub>1</sub> saw <b>so'n boy</b><sub>2</sub> entering a store.'</p>	<p><i>Draußen war es warm. Peter<sub>1</sub> hielt es zu Hause einfach nicht mehr aus und ging in die Stadt. Auf dem Weg dorthin sah er<sub>1</sub> einen Jungen<sub>2</sub> einen Laden betreten.</i></p> <p>'It was warm outside. Peter<sub>1</sub> couldn't stay home and went downtown. On his way there he<sub>1</sub> saw <b>a boy</b><sub>2</sub> entering a store.'</p>
TI2: <i>so'n</i> -condition	TI2: <i>ein</i> -condition
<p><i>Letztes Jahr war ich am Meer und hab' einen Rettungsschwimmer<sub>1</sub> getroffen. Er<sub>1</sub> hatte den ganzen Tag zu tun und eines Tages hat er<sub>1</sub> so'n Mädchen<sub>2</sub> vorm Ertrinken gerettet.</i></p> <p>'I was at the seaside last year and I met a lifeguard<sub>1</sub>. He<sub>1</sub> was busy all day and one day he<sub>1</sub> saved <b>so'n girl</b><sub>2</sub> from drowning.'</p>	<p><i>Letztes Jahr war ich am Meer und hab' einen Rettungsschwimmer<sub>1</sub> getroffen. Er<sub>1</sub> hatte den ganzen Tag zu tun und eines Tages hat er<sub>1</sub> ein Mädchen<sub>2</sub> vorm Ertrinken gerettet.</i></p> <p>'I was at the seaside last year and I met a lifeguard<sub>1</sub>. He<sub>1</sub> was busy all day and one day he<sub>1</sub> saved <b>a girl</b><sub>2</sub> from drowning.'</p>
TI3: <i>so'n</i> -condition	TI3: <i>ein</i> -condition
<p><i>Die Gäste saßen im Wohnzimmer rum, ohne zu wissen, mit wem sie reden könnten. Julia<sub>1</sub> sah so'nen Mann<sub>2</sub> alleine auf dem Sofa sitzen. In der Zwischenzeit servierten die Gastgeber<sub>3</sub> Snacks.</i></p> <p>'The guests were sitting in the living room not knowing with whom to start chatting. Julia<sub>1</sub> saw <b>so'nen man</b><sub>2</sub> sitting alone on the couch. Meanwhile, the hosts<sub>3</sub> served snacks.'</p>	<p><i>Die Gäste saßen im Wohnzimmer rum, ohne zu wissen, mit wem sie reden könnten. Julia<sub>1</sub> sah einen Mann<sub>2</sub> alleine auf dem Sofa sitzen. In der Zwischenzeit servierten die Gastgeber<sub>3</sub> Snacks.</i></p> <p>'The guests were sitting in the living room not knowing with whom to start chatting. Julia<sub>1</sub> saw <b>a man</b><sub>2</sub> sitting alone on the couch. Meanwhile, the hosts<sub>3</sub> served snacks.'</p>

The first sentence of each story set the scene. In the second sentence, a first referent was introduced in subject position by means of a proper name. In the first test item (TI1) in Table 4.9, *Peter* is the global topic constituent, as this referent was mentioned at least once in grammatical subject position and was the referent the story was about. In the last sentence of each story, the critical referent was introduced as an indefinite noun phrase in direct object position (i.e. *so'nen Jungen*, 'so-a boy' vs. *einen Jungen* 'a boy' in TI1 in Table 4.9). The third test item contained an additional last sentence, in which no relevant referent was mentioned.

Both subject and direct object referents in the target sentences were human and of different gender (except for the first test item, in which both referents were male). It is important to

note that the critical referent was introduced in direct object position in the second or third sentence of the fragments. Many studies on reference resolution (Givón 1983, Ariel 1990, Arnold 1998, Gordon, Grosz and Gilliom 1993, among others) showed that referents mentioned in direct object position are less accessible or salient in the discourse compared to the subjects of the same sentence- given that other relevant characteristics (e.g. semantic roles, verb type, various sentence connectors, etc.) are kept constant. Thus, besides investigating the discourse structuring potential of *so'n* indefinites, the present study indirectly addresses the question of difference in discourse prominence between grammatical subject and direct object referents.

#### 4.3.2.2 Procedure and data analysis

The first five main clauses (including subordinate ones, if there were any) of each continuation story provided by the participants, were coded by two independent judges for three aspects of the indefinite direct objects: (i) their referential persistence, (ii) their topic shift potential, and (iii) the type of anaphoric referring expressions used to pick up the referent for the first time. The two judges agreed upon 95% of the cases; in the case of disagreement (which were mostly concerned with determining the topic), differences were resolved through discussion. The three factors mentioned above were employed as a means to establish the discourse structuring potential of indefinite *so'n* compared to its simple indefinite counterpart. Example (32) presents one response from the *so'n*-condition for target item 1 (TI1), and Table 4.10 illustrates the coding methods used.

(32) Example responses and coding methods from the story continuation experiment:

---

*Target item 1:*                      *Draußen war es warm. Peter<sub>1</sub> hielt es zu Hause einfach nicht mehr aus und ging in die Stadt. Auf dem Weg dorthin sah er<sub>1</sub> so'nen Jungen<sub>2</sub> einen Laden betreten.*

‘It was warm outside. Peter<sub>1</sub> didn’t resist staying home and went downtown. On his way there he<sub>1</sub> saw so’n boy<sub>2</sub> entering a store.’

---

*Continuation sentences:*

- S1:** *Er<sub>2</sub> hatte tätowierte Arme und (pro<sub>2</sub>) schaute sich vorsichtig um.*  
 ‘He<sub>2</sub> had tattooed arms and (pro<sub>2</sub>) looked cautiously around.’
- S2:** *Als er<sub>2</sub> sich unbeobachtet glaubte, griff er<sub>2</sub> sich ein Paar Lederhandschuhe und (pro<sub>2</sub>) stopfte sie unter seinen Pullover.*  
 ‘When he<sub>2</sub> felt unwatched, he<sub>2</sub> grasped a pair of leather gloves and (pro<sub>2</sub>) stuffed them under his sweater.’
- S3:** *Er<sub>2</sub> ging dann ruhig aus der Tür als ob nichts wäre.*  
 ‘He<sub>2</sub> then went out of the store as if nothing had happened.’
- S4:** *Peter<sub>1</sub> stellt sich ihm<sub>2</sub> in den Weg und (pro<sub>1</sub>) sagte nichts.*  
 ‘Peter<sub>2</sub> thwarted him<sub>2</sub> and (pro<sub>1</sub>) said nothing.’
- S5:** *Er<sub>2</sub> beschimpfte dann Peter<sub>1</sub> und (pro<sub>2</sub>) rannte davon.*  
 ‘He<sub>2</sub> then insulted Peter<sub>1</sub> and (pro<sub>2</sub>) ran away.’

Table 4.10: Coding methods for the continuation sentences in example (32)

Coding methods		Subject		Object ( $\pm$ so'n)		topic
		refer	per	refer	per	
Anaphoric forms and grammatical function		item / S	sum	item / S	sum	
S1	[pron <sub>2</sub> ] [pro <sub>2</sub> ] [Sub <sub>2</sub> ] [Sub <sub>2</sub> ]	0	0	2	2	Topic <sub>2</sub>
S2	(pron <sub>2</sub> ) [pron <sub>2</sub> ] [pro <sub>2</sub> ] (Sub <sub>1</sub> ) [Sub <sub>2</sub> ] [Sub <sub>1</sub> ]	0	0	3	5	(Topic <sub>2</sub> )
S3	[pron <sub>2</sub> ] [Sub <sub>2</sub> ]	0	0	1	6	(Topic <sub>2</sub> )
S4	[name <sub>1</sub> , pron <sub>2</sub> ] [pro <sub>1</sub> ] [Sub <sub>1</sub> , IO <sub>2</sub> ] [Sub <sub>1</sub> ]	2	2	1	7	(Topic <sub>1</sub> )
S5	[pron <sub>2</sub> , name <sub>1</sub> ] [pro <sub>2</sub> ] [Sub <sub>2</sub> , DO <sub>1</sub> ] [Sub <sub>2</sub> ]	1	3	2	9	(Topic <sub>2</sub> )

In the sentence-continuation study, I compared the discourse behaviour of the direct object referents to that of the subject referents of the sentence. The grammatical subject (which had the semantic role of Agent) was the referent that was mentioned first in the target item

and was indexed by 1. The direct object referent (i.e. the critical referent) was indexed by 2. The values of the three textual characteristics under investigation (referential persistence, topic shift potential and type of the first referring expression) for the example response in (32) are given in Table 4.10. The first column indicates the number of continuation sentence (S1 for continuation sentence one, etc.). The second column indicates the type of anaphoric expression (i.e. pronoun, name, definite NP, definite modified NP) used to refer back for the first time to the direct object and subject referents. Round brackets indicate embedded clauses and square brackets matrix ones.

The third and fourth columns indicate the coding methods for referential persistence for both the subject and direct object referents. I counted how often a referent is mentioned in the five continuation sentences. Note that referential persistence was calculated in two different ways: first, as number of occurrences in a sentence (item/ sentence) and, second, as the sum of occurrences up to a certain sentence (i.e. sum). This twofold distinction allows us to compare two aspects. First, the differences in the rates of subsequent mention between the subject and direct object for each continuation sentence. Returning to Table 4.10 above, we can notice that the subject referent is mentioned in S4 more often than the direct object referent (2 vs. 1 time), whereas the opposite holds for S1 (i.e. the subject referent was not mentioned at all, while the direct object referent was mentioned twice). Second, the occurrences within each sentence were added to a sum representing the referential continuity at a particular stage in the discourse. This measure method indicates at what stage in the continuation fragment we have more anaphoric expressions referring to the direct object referent than to the subject referent. In the example (32) coded in Table 4.10, this already happens in S2.

Finally, the last column indicates at which point in the discourse a direct object referent becomes the subject. In Table 4.10, this happens in S1. Whether this change is maintained in the subsequent sentences or not, is not relevant for the present study, which is indicated by the round brackets surrounding the direct object topic in S2 in Table 4.10. Note again that only direct objects becoming topics in matrix clauses were considered, as subordinate clauses behave differently and do not allow for topical referents (Maki, Kaiser and Ochi 1999).

### 4.3.2.3 Results and discussions

Two continuations were excluded from the analysis as participants added an ambiguous ( $n=1$ ) or nonsensical continuation sentence ( $n=1$ ). This left 10 continuations to be analysed for each condition.

#### Referential persistence

Several observations can be made on the basis of the findings of referential persistence of referents. Figure 4.1 displays the average number of anaphoric references to the direct object referents for each continuation sentence (S1-S5) for both conditions. The results presented in the Figure below confirm the predictions made above, as the referent of the direct object headed by *so'n* is more referentially persistent than the referent of the direct object preceded by the simple indefinite article *ein*. Up to S5, the referential persistence of the referents introduced by indefinite *so'n* is  $2/3$  higher than that of their indefinite counterparts.

The findings with respect to the referential persistence of referents offer interesting insights into the relation between the subject and direct object referent. Figure 4.1 shows the average number of anaphoric references to the subject and direct object referents up to the last continuation sentences (S5) for both conditions of the experiment.

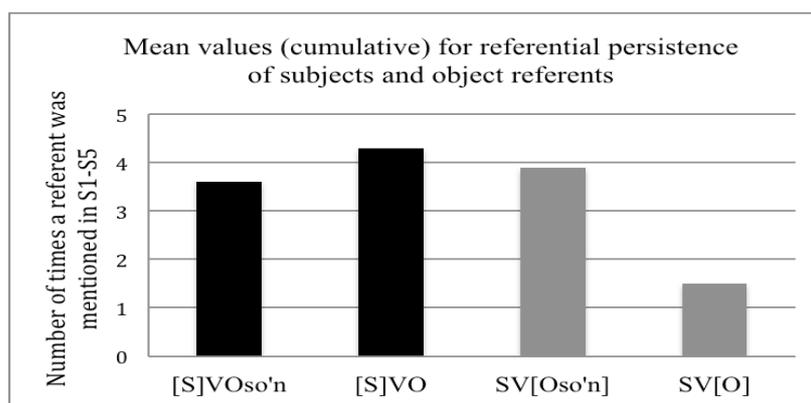


Figure 4.1: Mean values for referential persistence of subject and object in both conditions

For the *ein*-condition, in which the direct object referent was headed by the simple indefinite article, the difference in referential continuity between subject and direct object

referent is high, as referents associated with the subject referent are referentially more persistent. Participants preferred to continue the given fragments with the subject referent, whereas the direct object referent was rarely mentioned.

For the *so'n*-condition, however, the referential continuity value of the direct object referent is higher and even exceeds that of the subject referent. The increase in referential persistence of the direct object referents due to the presence of *so'n* has an impact upon the referential persistence of the subject referents. Indeed, taking a closer look at the referential persistence of subject and direct object referents for each target item separately reveals that the referential persistence of the subject referents is lower than that of the direct object referents, as Figure 4.1 shows. This observation underlines the privileged status the *so'n* referents occupy in the subsequent discourse and brings up the interesting question of whether *so'n* is responsible for the decrease in referential persistence of the subject referent or not? In other words, does the higher referential persistence of the referents of the direct objects affect that of the subject referents in the *so'n* condition? A more general question is, whether the referential persistence of one of the sentence's referents influences the referential persistence of other referents in the sentence, or not? Due to space limitations, I leave these questions for further research.

### **Topic shift potential**

The second textual characteristic investigated was the topic shift potential of the direct object referents headed by *so'n* compared to that of the direct object referents preceded by the simple indefinite article. The results presented in Figure 4.2 reveal two interesting patterns. A first overall generalization, which supports the findings made in Chapter 3 is that compared to the subject referents, both direct object types are less prone to become the topic in the immediately following matrix sentence. This observation is in line with other findings that report a clear tendency to continue with the grammatical subject of the preceding sentence, assuming that other factors (e.g. verb type) remain unchanged. Despite this general tendency, the *so'n*-referents display a higher probability to change the current topic from the first sentence-continuation on (in S1: 40% of *so'n*-referents vs. 15% for the *ein*-referents).

Second, we notice that the highest rate for the direct objects in the *so'n*-condition to become the aboutness topics in a following matrix clause is within the first three continuation sentences (S1-S3). In other words, the probability that a direct object headed by *so'n* becomes a topic in the following discourse increases only minimally after S3.

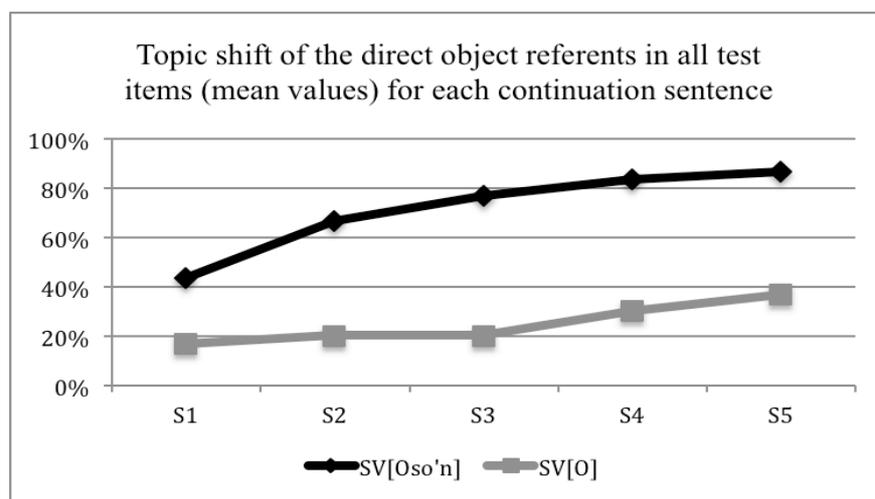


Figure 4.2: Topic shift of the direct objects

Third, Figure 4.2 shows that while almost all participants mentioned the referents of the *so'n* direct objects in grammatical subject position up to S5, the referents of the *ein* direct objects were mentioned as the topic of a matrix clause in less than 40% of the cases. In sum, prediction 2 is confirmed as well, as the *so'n*-marked referents became the new topics earlier and more often in the subsequent discourse than the referents introduced by the simple indefinite article.

As we did with the opposition presented in Chapter 3, between *this* indefinites and *a(n)* indefinites in English, we can observe that the indefinite headed by the simple indefinite *ein(e)* has a wider, less restricted distribution compared to its *so'n* marked counterpart. That means that simple indefinites are more prone to occur in both types of contexts: those in which their associated referents are less likely to be rementioned and less prone to shift the topic, but also in those in which their associated referents are subsequently mentioned and shifting the topic. On the contrary, *so'n* indefinites are more restricted to such contexts in which their associated referents show high values for both referential persistence and topic shift. We thus find again evidence for the generalization that unmarked forms are less restricted in their distributional patterns compared to marked forms (Givón 1981).

### Type of the first anaphoric expression

The last textual characteristic that was investigated was the type of referring expression used to pick up the direct object referents in the subsequent discourse. Tables 4.11 and 4.12 below display the absolute and mean values of this characteristic for both the *so'n* and *ein* referents.

Table 4.11: Type of referring expression of the first anaphoric item used to pick up the *so'n* referent

<i>So'n-referents</i> in TI1, TI2, TI3 (mean and absolute values)						
Ref. Type Sentence	Zero	Pronoun	Name	Def. unmodified NP	Indef. NP	Sum
S1	(0)	20% (6)	(0)	60% (18)	3.3 % (1)	83% (25)
S2	(0)	(0)	(0)	(0)	(0)	(0)
S3	(0)	(0)	(0)	6.7% (2)	(0)	6.7% (2)
S4	(0)	(0)	(0)	(0)	(0)	(0)
S5	(0)	(0)	3.3% (1)	3.3% (1)	(0)	6.7% (2)
Sum	(0)	20% (6)	3.3% (1)	70% (21)	3.3% (1)	96.7 % (29)

Table 4.12: Type of referring expression of the first anaphoric item used to pick up the *ein*-referent

<i>Ein-referents</i> in TI1, TI2, TI3 (mean and absolute values)						
Ref. Type Sentence	Zero	Pronoun	Name	Def. unmodified NP	Def. modified NP	Sum
S1	(0)	13.3% (4)	3.3% (1)	16.7% (5)	13.3 % (4)	46.7% (14)
S2	(0)	(0)	(0)	(0)	(0)	(0)
S3	(0)	3.3% (1)	(0)	3.3% (1)	6.7% (2)	13.3% (4)
S4	(0)	(0)	(0)	3.3% (1)	6.7% (2)	10% (3)
S5	(0)	(0)	(0)	(0)	3.3% (1)	3.3% (1)
Sum	(0)	16.7 (5)	3.3% (1)	23.3% (7)	30% (9)	73.3 % (22)

Ten different participants completed each of the two versions of the experiment, resulting in 30 continuation fragments to be analysed for the *so'n*-condition and 30 for the *ein*-condition. In the *so'n*-condition, one participant chose not to mention the direct object referent in a matrix clause, while for the *ein*-condition, there were 8 participants, which did not pick up the critical referent in a matrix clause. This observation is mirrored in the two

sums: 29 instances (96.7%) for the *so'n* direct objects and 22 instances (73.3%) for the *ein* direct objects. Six participants picked up the *so'n* direct object in S1 with a pronoun, which translates in the 20% in the first cell of the second column in Table 4.11. The sum 25 (representing 83%) of the first row in Table 4.11 shows that most participants mentioned the referent of the *so'n* direct object in S1. Several other observations can be made based on the results summarized in the two Tables above. First, the prediction that the referents of the *so'n* direct objects need more material to be taken up for the first time compared to the subject referents was confirmed. Participants chose only in few cases to mention the *so'n* or *ein* referents by more attenuated types of referring expressions as pronouns or names. A close inspection reveals that unmodified definite NPs prevail for the *so'n* referents (70% in the fourth column of Table 4.11), whereas in the case of *ein* referents, both modified and unmodified definite NPs were chosen (23.3% for definite unmodified NPs and 30% for definite modified NPs). Second, *so'n* direct object referents are mentioned for the first time more often in the first continuation sentence in comparison to the *ein* direct object referents (83% (25 instances) vs. 46.7% (14 instances) in S1).

As a whole, the findings from the third measure method show that the two types of indefinite noun phrases (headed by *so'n* vs. by the simple indefinite article *a(n)*) do not have a different impact upon the type of referring expression chosen to take up their referents again. In other words, both attenuated and less attenuated types of referring expression were chosen almost in the same proportion for the referents of indefinite *this* and *a(n)*. However, the presence of the indefinite noun phrases contributed to high rates of subsequent mention by means of explicit types of referring expressions (i.e. definite noun phrases).

#### 4.4 Summary

In this Chapter I showed that besides the proximal demonstrative *dieser* (Haspelmath 1997, Lyons 1999), German uses the determiner *so'n* to introduce hearer-new and discourse-new referents. I argued that indefinites headed by *so'n* show particular referential and discourse properties that set them apart from other indefinite expressions. First, I showed that noun phrases headed by *so'n* introduce new referents in the discourse, which display a strong preference for referential readings, as rigid reference, epistemic specificity, etc. These

theoretical considerations received further support from a web-based sentence interpretation task (Experiment 1) which showed that indefinite noun phrases headed by *so'n* tend to have wide scope interpretations, thus to presuppose the existence of their associated referents in contexts that contain intensional or modal operators. Furthermore, the findings of the study brought up interesting insights with respect to the interpretations of indefinites headed by the simple indefinite article under the same conditions, as this type of referring expression showed a strong preference for referentially non-specific readings, in spite of being further modified.

Second, I argued that the observation that indefinites headed by *so'n* allow for the realization of referential properties is not enough to fully characterize their function. As in the case of *this* indefinites in English, indefinites headed by *so'n* display a high discourse structuring potential, a property that becomes particularly evident in contexts that do not contain operators. More concretely, based on the findings of a sentence-continuation task (Experiment 2), which resembles the study reported in Chapter 3, I showed that referents headed by *so'n* are referentially more continuous in the subsequent discourse than their simple indefinite counterparts and that they are more prone to shift the topic of the discourse. Moreover, the presence of the indefinite noun phrases contribute to high rates of subsequent mention by means of explicit types of referring expressions (i.e. definite noun phrases).

Third, a more general observation is that referents associated to simple indefinites are not excluded from being rementioned in the subsequent discourse and from shifting the topic. In contrast, referents associated with *so'n* indefinites are less likely to remain unmentioned in the ensuing discourse and not to shift the topic. Indefinite noun phrases headed by the simple indefinite article are thus more flexible in occurring in both types of continuations, i.e. in those continuing with their associated referents and in those that do not. *So'n* indefinites on the contrary tend to be more strongly associated with only one type of continuation (i.e. one in which its associated referent is frequently rementioned and is shifting the topic). This observation provides further evidence for the generalization that unmarked forms (i.e. *ein* indefinites) are less restricted in their distribution compared to marked forms, i.e. *so'n* indefinites (Givón 1981). The aim of the next Chapter is to discuss an alternation found in Romanian with referents realized in direct object position in terms of their potential to give structure to the subsequent discourse.

## 5. *Pe*-marking of indefinite noun phrases in Romanian

Up to this point, I showed that two indefinite noun phrases, namely indefinite *this* in English and indefinite *so'n* in German impact the structure of the discourse after being introduced. Both, indefinite *this* and indefinite *so'n* were shown to display the potential to give structure to the subsequent discourse, as the referents associated with these types of referring expressions are more likely to be (frequently) picked up in the subsequent discourse and are more prone to shift the topic of the following discourse. In this Chapter, I will discuss another type of indefinite noun phrases, namely case-marked indefinite direct objects in Romanian, which fulfil a similar function by indicating predictable information in terms of frequency of re-mention and upcoming topic shift.

Various languages overtly mark their sentential direct objects non-uniformly and this phenomenon is known in the literature as differential object marking (DOM, Comrie 1979, Bossong 1985, Aissen 2003). This means that in a language that developed DOM, a particular class of direct objects receives a special case-marker, whereas the other class (generally) remains unmarked. Languages differ with respect to the exact (number of) features of the direct object argument towards which the DOM-system is sensitive to and with respect to the ways in which the same features interact in each language. In synchronic Romanian, for example, DOM is obligatorily present on personal pronouns and proper names, if they refer to human entities, while it is optional with nouns referring to human entities (Niculescu 1965, Manoliu-Manea 1993, Pană-Dindelegan 1997, Chiriacescu 2007, von Heusinger and Onea 2008, Stark and Sora 2008). The observations that DOM in Romanian is sensitive to the animacy and referential dimension of direct object arguments have led to the explanation that DOM functions as an individualizing device. Niculescu (1965) proposed that DOM occurs “in front of any person-denoting common noun, if it refers to a person that is already known, individualized for the speaker at the time of the communicative act”. Despite being intuitively correct, the view that marked direct object arguments refer to those referents that are individualised cannot explain cases in which DOM obligatorily occurs with bare quantifiers like *nimeni* (‘nobody’), *oricine* (‘whoever’), *cineva* (‘somebody’), which inherently send to non-individualised entities. Another problematic example for an analysis of DOM in terms of an individualising mechanism comes from its distribution with indefinite noun phrases,

which constitute the central part of this Chapter. Consider the sentences in (1), in which the verb *a auzi* ('to hear') may take both the *pe*-marked form, as in (1a), or the non-marked form, as in (1b).

- (1) (a) Petru l- a auzit pe un băiat.  
 Peter CL Aux. hear PE a boy  
 'Peter heard a boy.'
- (b) Petru a auzit un băiat.  
 Peter Aux. hear a boy  
 'Peter heard a boy.'

The alternation above cannot be explained in terms of individualization alone and has been attributed to the need for disambiguation between a specific and a non-specific reading of an indefinite noun phrase when it occurs in direct object position (Dobrovie-Sorin 1994). As it will become clear in Section 5.4, DOM with indefinite noun phrases extends its function from indicating specificity to giving structure to the discourse in terms of referential persistence and topic shift.

Rather than offering an explanation for all cases that allow for a variation in DOM, in this Chapter I focus on the seemingly optional distribution of the direct object marker with indefinite noun phrases. The aims of this Chapter are manifold. First, I will show that marked direct object arguments realized as indefinite noun phrases receive a specific interpretation in contexts containing operators. Second, I will argue that in transparent or neutral contexts, the variation in DOM cannot be accounted for in terms of specificity. By extending the analysis of indefinites to contexts of more than one sentence I will show that marked indefinite direct objects have the potential to give structure to the subsequent discourse, as their referents are referentially persistent and likely to shift the current topic of the discourse. I consider that marked indefinites display an affinity for referential readings due to discourse prominence; however, the availability of a referential reading is not an absolute requirement. I will argue that the discourse structuring potential constraint can prevent marked indefinites to scope under certain operators or quantifiers, but not others. The discussion presented in this Chapter brings favourable evidence for the fact that

DOM in Romanian is yet another device for giving structure to the subsequent discourse, alongside indefinite *this* in English and indefinite *so 'n* in German.

This Chapter is structured as follows: In Section 5.1 I introduce three factors, namely: animacy, referentiality and specificity, which have been considered the main triggers for DOM in general and in Romanian in particular. Then, I will present the distribution of *pe*-marking with different types of referring expressions. The picture that emerges in Section 5.2 is that, while pronouns and proper names represent instances that are obligatorily headed by *pe*, definite unmodified noun phrases and indefinite noun phrases show particular variances that cannot be accounted for that easily, unless the list of DOM-triggering factors is extended. Section 5.3 is dedicated to the distribution of *pe* with indefinite noun phrases. I will show that while *pe*-marked indefinites show a preference for referential readings (i.e. fixed reference, wide scope readings, etc.), specificity and other parameters previously mentioned in connection to DOM are not enough to account for all cases in which *pe* is allowed with indefinite noun phrases. In Section 5.4 I propose a new analysis for *pe*-marked indefinite noun phrases in Romanian, which is based on the findings of an experimental study that indicates that the referents of *pe*-marked indefinites are (i) referentially more continuous and (ii) good candidates to shift the topic of the discourse. The last Section concludes the present Chapter by showing that the distribution of DOM in Romanian cannot be thoroughly accounted for, unless the discourse structuring potential of the referents marked in this way is taken into account.

## 5.1 Criteria for differential object marking in Romanian

Romanian, like many other languages, exhibits Differential Object Marking (DOM, e.g. Niculescu 1965, Lazard 1984, Bossong 1985, Aissen 2003), which means that some direct objects are overtly (case) marked, whereas others remain unmarked<sup>23</sup>. In Romanian, DOM is obligatory for some noun classes, ungrammatical for others and optional for a third type (Niculescu 1965, Farkas 1978, Dobrovin-Sorin 1994, Pană-Dindelegan 1997, Mardale 2002, Chiriacescu 2007, von Heusinger and Onea 2008, Stark and Sora 2008). The DOM-marker in Romanian is realized by the particle *pe*, which is homophonous with the locative spatial preposition *pe*, meaning ‘on’.

---

<sup>23</sup> Other languages may display exceptions from this observation, as different types of direct object arguments can be marked with different case forms (Aissen 2003, Malchukov and de Hoop 2007).

Direct object arguments in Romanian are often accompanied by clitic doubling<sup>24</sup>. The obligatory occurrence of the clitic pronoun, limited in Spanish to contexts in which the direct object is realized as a strong pronoun, extends in synchronic Romanian to strong noun phrases like proper names, pronouns, definite descriptions, strong quantifiers, and indefinite noun phrases, all of which must be simultaneously *pe*-marked (Gierling 1997, Gramatica Limbii Române 2005, Chiriacescu 2007). The simultaneous occurrence of clitic doubling and DOM has led many researchers to the conclusion that these instances represent one phenomenon. In an earlier study (Chiriacescu 2007), I showed that despite many instances of overlap, the diachronic development and the current distribution of *pe*-marking and clitic doubling is different. I will thus restrict my analysis in this Chapter to *pe*-marking.

Previous approaches on DOM cross-linguistically and Romanian in particular have focused on the function and distribution of the (case) markers cross-linguistically. The two most prominent theories on DOM are the Ambiguity Dissertation and the Transitivity Hypodissertation. The proponents of the Ambiguity Dissertation or the Differentiating Hypodissertation (Comrie 1975, Silverstein 1976, Moravcsik 1978, Bossong 1985, Aissen 2003) consider that DOM originated in the need to highlight the direct object of a transitive sentence by means of a special marker, in cases in which the object could be confounded with the subject (or indirect object) of the sentence. This development occurs due to communicative pressures having to do with ambiguity avoidance (Bever 1970, Levy and Jaeger 2007, among others) and receives further support from more general economy principles found in languages, which treat idiosyncratic structures as being formally marked, whereas canonical structures remain unmarked (Sperber and Wilson 1995). According to this view, DOM encodes *the relation* between two arguments of a transitive construction and not the properties of a particular argument. In SVO languages like Romanian, it is largely acknowledged that prototypical (direct) objects are usually lower in

---

<sup>24</sup> Note that the Romanian clitic system is richer than that of other Romance languages. Object pronominal clitics have distinct morphological inflections for dative and accusative cases, whereas subject referents can never be doubled by a co-indexed weak pronoun (Steriade 1980, Klein 2007). The studies investigating object clitic doubling in Romanian concentrate on their distribution and status within a noun phrase. With respect to their function, direct object clitic pronouns have been most prominently accounted for in terms of specificity (Sportiche 1992) and focus projection (Gierling 1997). Other approaches describe object clitic pronouns as carrying a D-feature, which makes them heads with properties that resemble operators (Kallulli 2000), or as agreement markers (Géraldine Legendre 2000).

animacy and/ or definiteness than prototypical subjects. Thus, only when the direct object is higher or equal to the subject in terms of animacy and definiteness, a special marker is needed to distinguish between the two arguments. Romanian DOM generally applies to direct objects, which are specified for the semantic feature [+human]. Proponents of the Ambiguity Dissertation predict that [+definite] objects will be the prime targets for *pe*-marking. This expectation is borne out for proper names and personal pronouns, but not for all direct objects realized as definite and indefinite noun phrases, which may remain unmarked even when the subject referent is [-definite].

The second main approach to DOM, The Transitivity or Indexing/ Coding Approach (Hopper and Thompson 1980), analyses this phenomenon as indicating particular semantic and/ or pragmatic properties of a given argument, irrespectively of its relation to other arguments present in the sentence. Under this view, DOM-marking does not signal the non-prototypicality of the direct object (i.e. its resemblance with the subject in terms of animacy, definiteness or specificity), but is rather considered an *indicator of high transitivity* (Hopper and Thompson 1980). More precisely, this approach suggests that a direct object is overtly marked if it represents a prototypical argument in a transitive sentence that represents a ‘salient event’. For Naess (2004) and von Heusinger and Kaiser (2010), the role played by the direct object referent, especially in terms of affectedness, is decisive for its case-marking. Affectedness is roughly defined as involvement in the verbal event and change of state in the participant as a result of this event. An affected referent is an individualized referent, which in turn correlates with its definiteness. An affected direct object referent is considered to be more prototypical and tends to receive more grammatical marking than a less-affected, non-prototypical direct object referent. Translated to the Romanian data, this means that a verb with a *pe*-marked direct object contributes to a higher degree of transitivity of the sentence containing this object (e.g. 2a). Conversely, to express low transitivity, a non-*pe*-marked form will be used, or an alternative construction, as for example a reflexive verb and a prepositional object as illustrated in (2b).

- (2) (a) Ion (îl) întâlnește (pe) un prieten.  
 Ion CL meets PE a friend.  
 ‘John meets a friend.’

- (b) Ion se întâlnește cu un prieten.  
 Ion SE meets with a friend.  
 ‘John meets (himself) with a friend.’

We should keep in mind at this point that *pe*-marking seems to be a means to signal the discourse prominence of the direct object argument in terms of giving structure to the subsequent discourse. For both, the Ambiguity and the Transitivity theories, the most important synchronic semantic and pragmatic factors triggering DOM are animacy, definiteness/ referentiality and specificity. These factors are largely conceived in the typological literature as scales or hierarchies (Comrie 1975, Bossong 1985, Haspelmath 1997, Aissen 2003, among others). On these scales, DOM occupies continuous sectors of different lengths, starting at the more prominent part of one or more of these scales, as presented in (3). The scales represent useful tools for characterizing DOM, as they can predict the case-marking of a referent, whenever it occupies a high position on one or more of these scales.

### (3) Prominence scales for DOM

#### Animacy scale

human	>	animate	>	non-animate
-------	---	---------	---	-------------

#### Referentiality scale

pers. pron.	>	PN	>	def. DP	>	indef DP	>	non-arg DP
-------------	---	----	---	---------	---	----------	---	------------

#### Specificity scale

specific	>	non-specific
----------	---	--------------

#### Topicality scale

topic	>	non-topic
-------	---	-----------

De Hoop and Narasimhan (2005), de Hoop and Malchukov (2007) and Klein and de Swart (2010) showed that the factors or argument features (e.g. animacy, definiteness, specificity, topicality) generally used to account for the distribution of DOM are not weighted similarly and do not interact in the same way cross-linguistically. They propose a

differentiation between strong and soft constraints to account for the distribution of DOM. Strong constraints obligatorily trigger or block the occurrence of overt DOM marking, as they are intrinsic to an argument and are not prone to change, whereas soft constraints allow for an alternation in DOM. The differentiation between strong and soft constraints can explain most instances of Romanian DOM as well.

An example for a strong or hard constraint in Romanian is based on the DP-type (Klein and de Swart 2010), as the object marker *pe* appears with personal pronouns, irrespective of other argument factors, such as animacy. In example (4), the personal pronoun *ea* ('she') must be *pe*-marked, even though it refers back to a non-animated referent.

- (4) Televiziunea m -a ales pe mine, nu eu pe ea.  
 Television CL Aux. chosen PE me not I PE her  
 'Television has chosen me, not I it.'

(Klein and de Swart 2010)

*Pe*-marking with proper names represents another strong constraint in Romanian, as the semantics of names obligatorily triggers DOM, as illustrated in (5). It seems that *pe*-marking does not contribute an additional feature to proper names but rather functions as an agreement marker.

- (5) L -am văzut pe Ion/ Donald Duck/ \*caiet.  
 CL Aux. seen PE John/ Donald Duck/ \*notebook  
 'I have seen John/ Donald Duck/ the notebook.'

Toponyms form a class of proper names that generally do not host *pe*. Not even in cases in which proper names referring to names of countries or cities are used metonymically denoting the inhabitants of a city is the occurrence of *pe* strongly preferred<sup>25</sup>.

<sup>25</sup> There are further additional conditions triggering the *pe*-marking of proper names, including metonymical shifts, metaphorical transfers, but I do not discuss them at this point as they are not of primary concern for the purpose of this dissertation. See Gramatica Limbii Române (2005) or Chiriacescu (2007) for an ample discussion of these instances.

The variation between semantically human noun phrases, on the one hand, and inanimate noun phrases, on the other hand, represents another strong constraint that holds in Romanian, as only the former can take the object (case) marker. The contrast between animate and inanimate objects becomes evident when comparing (6a-b) to (6c).

- (6) (a) Am văzut \*(-o) \*(pe) femeia frumoasă.  
 Aux. seen CL PE woman beautiful.  
 ‘I have seen the beautiful woman.’
- (b) Am văzut (\*-o) (\*pe) pisica frumoasă.  
 Aux. seen CL PE cat beautiful  
 ‘I have seen the beautiful cat.’
- (c) Am văzut (\*-o) (\*pe) biserica frumoasă.  
 Aux. seen CL PE church beautiful  
 ‘I have seen the beautiful church.’

The referent of *the woman* in (6a) is marked with *pe*, as it is specified for the semantic feature [+human]. On the contrary, in (6b) and (6c) the presence of the marker *pe* is strictly blocked, as the definite modified noun phrases do not denote human referents. The presence or absence of the clitic pronoun does not save any of the two sentences in (6b) and (6c). Because *pe*-marking targets those direct objects that are specified for the semantic feature [+human], I will leave out inanimate direct object referents in the following.

Besides the two strong constraints mentioned above, namely animacy and *pe*-marking with proper names and pronouns, which obligatorily trigger the *pe*-marked form or its unmarked counterpart, Romanian DOM is sensitive to soft or fluid constraints as well. Such constraints allow for a less rigid distribution of the *pe*-marker and the presence or absence of overt case marking does not render the expression (un)grammatical, but changes the interpretation of the direct object. In Romanian, a case in point for such a soft or fluid constraint represents the specific vs. non-specific reading of indefinite noun phrases in different contexts in which a *pe*-marked indefinite noun phrase often bears a specific reading, while its non-marked counterpart is associated with the non-specific reading of its

referent (Dobrovie-Sorin 1994). The distribution of *pe*-marking with indefinite noun phrases will be discussed in Section 5.3.

DOM was accounted for in terms of different argument features, most commonly animacy, referentiality and specificity that trigger the *pe*-marked or the non-marked form on the direct object. While such factors are useful and necessary to account for the distribution of *pe*-marking in Romanian, I will argue that they are not sufficient to explain all instances of DOM. I will focus on a seemingly optional variation found with indefinite noun phrases in neutral contexts (i.e. contexts that do not contain any sentence operators), as in (7).

- (7) (a) Petru l -a văzut pe un băiat.  
 Petru CL Aux. see PE a boy  
 ‘Peter saw a boy.’
- (b) Petru a văzut un băiat.  
 Petru Aux. saw a boy  
 ‘Peter saw a boy.’

The presence of *pe* in (7a) and its absence in (7b) can be accounted for neither in terms of specificity, nor in terms of individuation. In this Chapter I argue that discourse structuring plays a pivotal role in the distribution of *pe*-marking in Romanian, especially in such contexts that allow for variation in *pe*-marking. More concretely, I will show that *pe*-marked direct objects, in contrast to their unmarked counterparts, are discourse prominent in terms of discourse structuring potential, as they are: (i) referentially highly persistent, and (ii) prone to become the topic in the subsequent discourse. Before investigating the distribution of *pe*-marking with indefinite noun phrases in greater detail, I will outline the conditions of *pe*-marking with definite noun phrases<sup>26</sup>. The discussion does not include an analysis of bare singular nouns, but see von Heusinger and Chiriacescu (2011) for further details.

<sup>26</sup> For an ample discussion of the distribution of DOM in Romanian in general, see Farkas (1987), Gramatica Limbii Române (2005), Chiriacescu (2007) and von Heusinger and Onea (2008).

## 5.2 The distribution of *pe*-marking with definite expressions

In this Section I discuss the contribution of the differential object marker *pe* with definite (un)modified noun phrases, which allow for particular cases of variation that will be accounted for in terms of different soft and hard constraints.

### 5.2.1 DOM with definite modified noun phrases

Human direct objects realized as definite noun phrases are usually *pe*-marked, if they are further modified. The subject referent in (8A), introduced by means of the indefinite noun phrase *un băiat* ('a boy') is taken up in the continuation sentences (8a) and (8b) by means of the same definite modified noun phrase *băiatul bolnav* ('the sick boy'). If no other semantic and/or syntactic restrictions are present in the sentence, definite modified noun phrases are generally preceded by *pe*, as in (8a). Constructions in which the definite modified direct object remains unmarked, like in (8b), are marginal and rather dispreferred in synchronic Romanian, according to most studies (Farkas 1978, Chiriacescu 2007, von Heusinger and Onea 2008, among others).

- (8) A: Un băiat merge la doctor. (A boy goes to the doctor.)  
 (a) Doctorul îl examinează pe băiatul bolnav.  
 Doctor.DEF CL examines PE boy.DEF sick  
 'The doctor examines the sick boy.'  
 (b) ?Doctorul examinează băiatul bolnav.  
 Doctor.DEF examines boy.DEF sick  
 'The doctor examines the sick boy.'

In von Heusinger and Chiriacescu (2009) we tested whether the preference for *pe*-marking with definite modified noun phrases is systematic or whether we find cases of variation. In a *Google* survey and in a corpus containing Romanian newspaper articles, we analysed 650 examples of DOM with *pe*-marked and non-*pe*-marked definite modified noun phrases in relation to three verbs, namely *a omorî* ('to kill'), *a critica* ('to criticize') and *a impresiona* ('to impress'). Phraseologies and repetitions were excluded from the investigation to avoid any false results. An example of our search method with the verb *to impress* is given in (9).

- (9) (a) Am impresionat -o pe femeia [...].  
 Aux. impressed CL PE woman.DEF  
 ‘I have impressed the [...] woman.’
- (b) Am impresionat femeia [...].  
 Aux. impressed woman.DEF  
 ‘I have impressed the [...] woman.’

The examples above are similar except that in (9a) the definite direct object is preceded by *pe* and doubled by a clitic, while in (9b) the same direct object is neither marked with *pe* nor doubled by a clitic pronoun. As it would have been complicated and difficult to test, we did not specify the modifier of the definite noun *femeia* (‘the woman’).

Table 5.1 summarizes the findings of the corpus analysis with respect to the frequency of occurrence of the *pe*-marker with definite modified noun phrases for the three verbs. For unmarked direct objects, Table 5.1 makes further distinctions between different syntactic (i.e. the presence of the possessive dative) or semantic restrictions (i.e. collective readings of the noun phrase) that block the occurrence of *pe*. The last column of the same Table contains definite modified noun phrases that allow for both the *pe*-marked and the non-*pe*-marked forms.

Table 5.1: Frequency of human *pe*-marked and non-*pe*-marked definite modified NPs

Verbs	Total	<i>Pe</i> -marked objects	Non- <i>pe</i> -marked objects		+/- marked objects
			Syntactic restriction	Semantic restriction	
To kill	63 (100%)	49 (77%)	10 (15%)	1 (1.6%)	3 (4.8%)
To criticise	138 (100%)	113 (82%)	7 (5%)	15 (11%)	3 (2%)
To impress	108 (100%)	92 (85%)	4 (3.8%)	10 (9%)	2 (1.9%)
Total	309 (100%)	254 (82%)	21 (6.8%)	26 (8%)	8 (2.6%)

Several observations about definite modified noun phrases in direct object position can be made on the basis of the findings of the corpus analysis. First, the investigation confirmed prior observations, as human definite modified direct objects governed by the three verbs above were preceded by *pe* in most cases. Second, in about 13% of the cases, the presence

of *pe* on the definite modified noun phrase is strictly blocked by two hard constraints: a syntactic and a semantic restriction. The syntactic restriction blocking the appearance of *pe* on the definite noun phrase involves the possessive dative, which cannot co-occur with DOM in general (see discussion in the next Section). The semantic restriction that renders the *pe*-marked construction infelicitous is found with definite noun phrases that bear a collective reading, or with noun phrases that represent metonymical shifts (e.g. *El a impresionat presa straină*, ‘He impressed the foreign press’), which trigger an inanimate interpretation of the noun phrase and therefore block *pe*. Third, the last column of Table 9 underlines the existence of marginal cases of variation (8 out of 254, representing roughly 3% of the total occurrences) in which *pe* optionally heads the modified definite noun phrase. In the example given in (10), the modified definite noun phrase is non-*pe*-marked, even though similar objects are generally preceded by *pe*.

- (10) [...] a omorât copilul de numai șase ani al [...].<sup>27</sup>  
 Aux. killed boy.DEF of only six years of  
 ‘[...] he killed the six-year-old boy of [...].’

The semantic or syntactic restrictions mentioned above do not hold in (10), as the sentence contains neither a possessive dative, nor does the modified definite noun phrase bear a collective or inanimate reading. In the remainder of this Chapter I will focus on DOM with definite unmodified and indefinite noun phrases that allow for more variations in *pe*-marking, which cannot be accounted for in terms of a bundle of semantic features of the argument alone.

## 5.2.2 DOM with definite unmodified noun phrases

The distribution of *pe*-marking with definite unmodified noun phrases is more complex, as different types of constraints play a role.<sup>28</sup> For the sake of simplicity and in order to offer a unitary analysis of *pe* with definite noun phrases, most analyses on *pe*-marking with definite noun phrases abstract away from cases in which the marker *pe* is optional (Farkas

<sup>27</sup> <http://Section.6am.ro/iasi/tanara-de-19-ani-din-iasi-retinuta-pentru-ca-si-ar-fi-ucis-copilul> [viewed on 09.11.2008]

<sup>28</sup> Existing accounts on DOM with definite unmodified noun phrases consider genericity (Dobrovie-Sorin 2007) and individualization (Stark 2008) as relevant features licensing the one or the other form.

1978, Klein and de Swart 2010). Thus, most analyses consider the presence of the marker *pe* on the definite noun phrase obligatory, disregarding cases in which the marker is blocked or is in seemingly free variation. In the following, I sketch out the different distributional contexts of DOM-marking with definite unmodified noun phrases, underlying the existence of instances in which *pe*-marking is optional. I argue that definite unmodified noun phrases represent a case of a fluid alternation (i.e. allowing for optionality) inside an otherwise split domain (i.e. that generally triggers *pe*-marking). Furthermore, following von Heusinger and Chiriacescu (2011), I argue that the alternation is affected by the different discourse behaviour of the *pe*-marked form compared to the non-marked one.

### 5.2.2.1 Hard constraints

At least two hard constraints are responsible for the unacceptability of *pe*-marking with definite unmodified noun phrases, namely possessor raising constructions and collective readings of noun phrases. First, at sentence level, *pe* is ruled out in possessor raising constructions as in (11), even though the main triggering factors for DOM (animacy and definiteness) are met. Note that adjectival modification (i.e. the adjective *buna*, ‘good’ in (11)) does not save the sentence from ungrammaticality.

- (11) Maria își înțelege (\*pe) buna prietenă.  
 Maria herself understands PE good.DEF friend  
 ‘Maria understands her good friend.’

The noun phrase *buna prietenă* (‘good girlfriend’) in (11) is strongly individuated and receives the definite article. Constructions involving dative possessives represent the unmarked modality to convey possession in Romanian. As in many other languages in the Balkan Sprachbund (see Tomić 1996 for Bulgarian and Macedonian, Alexiadou 2001 for Greek), pronominal possessive relations in the domain of the noun phrase can be expressed not only by dative possessives as in (11), but also by full pronominal possessive modifiers like in (12).

- (12) Maria o înțelege pe prietena ei<sub>MARIA/?</sub> [dar nu pe a mea].  
 Maria CL understands PE friend.DEF her [but not mine]  
 ‘Maria understands her friend [but not mine].’

In comparison to (11), the definite noun phrase *prietenta ei* (‘her girlfriend’) in (12) bears two readings. In the first and marked reading, the pronoun *ei* (‘her’) is coreferent with the subject of the matrix sentence, *Maria*. In this case, the sentence has the same propositional meaning as sentence (11) above. For this reading, sentence (12) adds a discursive contrast, emphasizing the fact that the direct object referent is Maria’s friend but not mine/ yours, etc. The second and unmarked reading of the definite noun phrase in direct object position is one in which the pronoun *ei* (‘her’) is not bound to the grammatical subject referent, but rather to another female referent outside the discourse.

The second context in which *pe*-marking cannot precede the direct object is when the definite unmodified noun phrase accommodates a collective reading (i.e. Cruse’s 1986 “groups”) as in (13). Collective nouns, as for example ‘the committee’, ‘the press’, or ‘government’ are generally analysed as a unit, rather than from the level of the individuals that form the group and therefore cannot be headed by *pe*.

- (13) Paul a criticat vehement comisia.  
 Paul Aux. criticized vehemently committee  
 ‘Paul vehemently criticized the committee.’

### 5.2.2.2 Soft constraints

Besides the two hard constraints discussed above, soft constraints play a role in *pe*-marking with definite unmodified noun phrases as well. Recall that in contrast to the former, the so-called soft constraints do not apply in a categorical way. Thus, instead of ruling out the *pe*-marked form or its unmarked counterpart, soft constraints may trigger the *pe*-marked form or the unmarked one, depending on the intended reading of the noun phrase. Demonstratively used noun phrases, immediate situation uses and generic readings of noun phrases represent three contexts in which the presence of *pe* triggers a particular reading of the noun phrase it precedes.

First, as any other definite expression, unmodified definite noun phrases can be used demonstratively, as shown in (14). Such uses typically involve a pitch accent on the determiner, which suggests that these noun phrases retain a special status in the discourse.

(14) Context: Mary and Peter are watching a show about [the mayor of Braşov<sub>i</sub>]. Peter says to Mary pointing [to the man on TV<sub>i</sub>]:

- (a) Un prieten de-al meu îl cunoaste [pe primar<sub>i</sub>].  
 A friend of mine CL knows PE mayor
- (b) ??Un prieten de-al meu cunoaste [primarul<sub>i</sub>].  
 A friend of mine knows mayor.DEF  
 ‘A friend of mine knows the mayor.’

In (14), the mayor is indirectly present in the utterance context since he appears on TV and referring to him demonstratively is possible only with *pe*, as witnessed in (14a). It seems that for a *pe*-marked construction to be felicitously used in a context like (14), the referent should have been previously introduced in the discourse or be accompanied by a pointing gesture. Referring to an entity by means of a non-*pe*-marked construction, as in (14b) implies that the speaker refers to some other mayor and not the one already mentioned in the preceding discourse. As this interpretation is incompatible with the context given in (14), in which the mayor is discourse old, sentence (14b) is infelicitous.

A second context in which *pe*-marking gives rise to a difference in readings is in immediate situational uses in the sense of Hawkins (1978), as shown in example (15). In comparison to context (14), in which *pe* can be felicitously used only when the referent is familiar to both conversation participants, in the immediate use, the referent in question does not have to be familiar to both discourse participants. Rather, *pe* can be used in a context as (15) whenever the hearer can immediately identify (e.g. see) the referent upon hearing the sentence.

(15) Context: A policeman is standing guard outside a prison, which is surrounded by a fifteen-foot wall. Suddenly, he hears the voice of a colleague policeman from the other side:

- (a) Prinde-l pe fugar.

- |     |                      |             |         |                             |
|-----|----------------------|-------------|---------|-----------------------------|
|     | Catch-CL             | PE          | runaway |                             |
| (b) | Prinde fugar         | -ul.        |         |                             |
|     | Catch                | runaway.DEF |         |                             |
|     | ‘Catch the runaway.’ |             |         | (adapted from Hawkins 1978) |

*Pe*-marking represents a soft constraint in (15), as both sentences (15a) and (15b) are felicitous, but are associated with different readings. The *pe*-marked form can be used only in a context in which the hearer can see the escaped prisoner. Thus, in sentence (15a), the policeman is not only informed of the existence of the escaped prisoner, but s/he is also instructed to locate the referent in the immediate situation of utterance. On the contrary, if the first policeman utters sentence (15b), the hearer does not have to see the runaway to understand what happened and how he should react. He would have to look where the prisoner actually is. In both (15a) and (15a), the *pe*-marked referents are understood as referring to a unique individual that has the relevant property in the particular context.

Third, as already pointed out by Dobrovie-Sorin (2007), *pe* is semantically relevant for definiteness in contexts like (16) below. If the noun *femeia* (‘the woman’) bears a generic reading, the marker *pe* cannot precede it, as happens to be the case in (16b). On the contrary, the sentence with *pe* in (16a) triggers the non-generic reading of the same noun.

- |      |     |                                   |          |    |           |                       |
|------|-----|-----------------------------------|----------|----|-----------|-----------------------|
| (16) | (a) | Mihai o                           | adoră    | pe | femeie.   |                       |
|      |     | Mihai CL                          | worships | PE | woman     |                       |
|      |     | ‘Mihai worships the/ that woman.’ |          |    |           | [non-generic]         |
|      | (b) | Mihai                             | adoră    |    | femeia.   |                       |
|      |     | Mihai                             | worships |    | woman.DEF |                       |
|      |     | ‘Mihai worships women.’           |          |    |           | [generic]             |
|      |     |                                   |          |    |           | (Dobrovie-Sorin 2007) |

Genericity and uniqueness representing soft constraints, neither explain the optionality of DOM in examples (14-16), nor do they account for cases in which generic interpretations or uniqueness readings would not be available anyway.

### 5.2.2.3 Neutral contexts

Up to this point it was shown that *pe*-marking is subject to hard constraints (i.e. such contexts which either obligatorily block or require the *pe*-marked form) and soft constraints (i.e. contexts in which both the *pe*-marked and the non-*pe*-marked form are acceptable and give rise to different readings of the noun phrase). Besides these two contexts, *pe*-marking with definite unmodified noun phrases can be used in neutral contexts as well. Such contexts were neglected in the literature on DOM in Romanian, as both direct object forms can be used seemingly interchangeably. One case in point is illustrated in example (17). Romanian has a grammatical rule that is responsible for the ungrammaticality of the enclitic definite article on the unmodified noun phrase in the presence of other prepositions<sup>29</sup>. While most prepositions always block the attachment of the enclitic definite article on unmodified nouns (note the form *la doctor* ('to doctor') in (17A)), in the case of *pe* as a case marker, the above mentioned constraint also holds and gives rise to an alternation. Speakers of Romanian can either drop the marker *pe* or keep the definite enclitic article *-ul* (DEF.masc), as in (17a), or they can drop the definite article and keep *pe*, as illustrated in (17b).

(17) A: Un băiat merge la doctor. (A boy goes to the doctor.)

(a) Doctorul examinează băiatul.

Doctor.DEF examines boy.DEF

'The doctor examines the boy.'

(b) Doctorul îl examinează pe băiat.

Doctor.DEF CL examines PE boy

'The doctor examines the boy.'

Both sentences (17a) and (17b) represent different possibilities of expressing very similar referential categories. As it could be noticed so far, Romanian shows the same variation for definite modified direct objects and unmodified definite direct objects. Nevertheless, the

<sup>29</sup>In addition to the cases in which the direct object realized as an unmodified definite noun phrase cannot be preceded by *pe*, there also exist certain contexts in which the presence of the DOM marker is obligatory. Archaic uses of certain terms in direct object position found in written texts at the beginning of the 20<sup>th</sup> century constitute an exception in the sense that such expressions can be simultaneously suffixed by the definite article and host *pe* (Chiriacescu 2007).

contrast is different: for modified definite objects the alternation concerns the marker *pe*, but not the definite article. For unmodified definite objects, the alternation affects both: the marker *pe* and the definite article. Moreover, the semantic-pragmatic conditions are probably quite different: For modified definite direct objects, the form without the *pe*-marker is marginal, while for unmodified definite direct objects, both forms are acceptable. Depending on the context and language register, Romanian speakers tend to prefer one alternative to the other. The crucial observation is, however, that both sentence (17a) and (17b) are grammatical and have the same propositional content, and the presence of *pe* seems to be optional. In von Heusinger and Chiriacescu (2009) we investigated the discourse effects of *pe*-marked objects compared to their unmarked counterparts, by counting their subsequent coreferential expressions. We showed that the optionality in *pe*-marking can be accounted for in terms of the preferential status of *pe*-marked referents in the discourse, as such referents are referentially more persistent in the subsequent discourse. Thus, it seems that *pe*-marking with a definite noun phrase is employed by the speaker whenever he intends to elaborate upon the referent introduced in this way.

In this Section I discussed the distribution of *pe*-marking with different types of referring expressions. It was shown that while pronouns, proper names and definite modified noun phrases are generally headed by the direct object marker *pe*, definite unmodified noun phrases are optionally *pe*-marked, depending on additional syntactic and semantic restrictions. In the next Section I discuss the distribution of *pe*-marking with indefinite noun phrases showing, focusing on neutral contexts.

### **5.3 *Pe*-marking with indefinite noun phrases**

The literature on DOM in Romanian considers specificity one of the main triggering factors for indefinite noun phrases (Farkas 1978, Dobrovin-Sorin 1994, von Heusinger and Onea 2008, among others). This Section summarizes some major approaches to DOM with indefinite noun phrases, drawing a line between indefinite noun phrases used in opaque contexts (i.e. contexts that contain particular sentence operators) and indefinite noun phrases used in plain and transparent contexts (i.e. contexts that do not contain any operators). I will show that *pe*-marked indefinite noun phrases get referential readings in opaque contexts, as they generally take wide scope, have fixed reference, etc. In contrast, in plain and transparent contexts, indefinites allow for a variation between the *pe*-marked

form and the non-*pe*-marked form. The crucial observation will be that specificity in terms of scope taking over other operators is not sufficient to account for the alternation in such contexts. In the remainder of this Chapter I will argue that the seemingly optional distribution of *pe* with indefinite noun phrases can be explained if we consider the broader discourse in which they occur. I bring theoretical and empirical evidence that indefinite noun phrases headed by *pe* are a means to give structure to the discourse, as their associated referents are more likely to be picked up and to shift the topic in the subsequent discourse.

### 5.3.1 Specificity

The notion of specificity plays a major role in the analysis of *pe*-marked indefinite noun phrases. The concept of specificity was introduced in the late 60's (see Karttunen 1968, Fodor 1979, Jackendoff 1972) by transferring the *de re-de dicto* distinction applied to definite noun phrases under verbs of propositional attitudes, as illustrated in (18a) and (18b), to indefinite noun phrases as in (18c) and (18d).

- (18) (a) Mary wants to marry the winner- but he does not like her at all. [de re]  
 (b) Mary wants to marry the winner- so she waits for the match to end. [de dicto]  
 (c) Mary wants to date a Mexican- even though he was rude to her. [specific]  
 (d) Mary wants to date a Mexican- so she hopes to meet one soon. [non-specific]

In pre-theoretical terms, a specific reading of an indefinite noun phrase can be paraphrased as “the certainty of the speaker about the identity of the referent”. This understanding finds its reflection in the semantic properties of specific indefinite noun phrases, which are assumed to take wide scope, to be referential terms, and to trigger the presupposition that there is an individual satisfying the description. As researchers focussed on one or more of these properties of indefinite noun phrases, different views on specificity emerged. This being so, it is no surprise that the concept of specificity has been used as a general term for several phenomena such as: (i) scopal specificity, (ii) epistemic specificity, (iii) partitive specificity<sup>30</sup> and (iv) specificity as noteworthiness. Whether these phenomena refer to

<sup>30</sup> Milsark (1974) introduces *partitive specificity* as a third distinction in the semantic typology of specificity. He argues in favour of a distinction between a weak (existential) reading and a strong (presuppositional or partitive) reading. The considerations regarding partitive specificity cannot be applied to the Romanian data in the same manner because Romanian overtly marks partitivity by means of the indefinite article *un* (‘one’) - inflected for gender and number- and by the partitive preposition *din/ dintre* (‘of’). Such determiners can

related phenomena or to the same concept is a matter of debate and I will not go into this discussion any further as it is not central to the aim of this dissertation (but see von Heusinger 2011 for an ample analysis and review of specificity).

The first main approach to specificity is the so-called *scopal specificity account*, which explains the contrast between a specific and a non-specific use in terms of scope interactions between the indefinite description and some other operator in the sentence. Take for example sentence (19a), which can be interpreted in two ways, as the two continuations, (19b) and (19c), indicate. The different behaviour with respect to operators motivate the two possible readings of the indefinite *a house* in (19a) above. The indefinite noun phrase scopes over the negator operator in (19b) and receives widest scope, as it is a constant that does not participate in scope ambiguities, whereas the same indefinite scopes under the operator in (19c).

- (19) (a) The boy didn't see a house.  
 (b) There is a house the boy didn't see.  
 (c) The boy saw no house.

In opaque contexts, indefinite noun phrases headed by *pe* are generally interpreted as specific, whereas unmarked noun phrases are ambiguous between a specific and a non-specific reading (see Farkas 1978, Dobrovie-Sorin 1994, Farkas and von Heusinger 2002, Chiriacescu 2007, Klein and de Swart 2010). In example (20), the direct objects are locally bound by an existential operator, which is realized in subject position. The *pe*-marked indefinite in (20b) is specific and can refer to a particular woman only, taking wide scope. On the contrary, the non-*pe*-marked direct object in (20a) is ambiguous between a wide and a narrow scope reading. In intensional contexts, *pe*-marking is sometimes possible, however the reading between this construction and its non-*pe*-marked counterpart are again significantly different.

---

project only partitive noun phrases (Comorovski 1995). If a partitive noun phrase is realized as a human direct object, *pe*-marking is obligatory, which is not surprising since partitives include a definite or familiar domain of quantification for the indefinite description, they are expected to behave like definite noun phrases with respect to *pe*-marking (which, as we have seen above, generally host *pe* whenever other semantic and/ or syntactic restrictions do not apply). For similar observations regarding DOM marking and partitivity in Spanish, see Brugè and Brugger (1996).

(20) *Extensional operators (universal quantifiers)*

(a) Toți bărbații iubesc o femeie.

All men love a woman

'All men love a woman.'

[wide/ narrow scope]

(b) Toți bărbații o iubesc pe o femeie.

All men CL love PE a woman

'All men love a/ this woman.'

[wide scope only]

(21) *Intensional operators*

(a) Ion caută o secretară.

John looks for a secretary

'John looks for a secretary.'

[wide/ narrow scope]

(b) Ion o caută pe o secretară.

John CL looks for PE a secretary

'John looks for a secretary.'

[wide scope only]

The indefinite noun phrase *o secretară* ('a secretary') in (21a) could refer to a specific as well as a non-specific secretary, while the same noun phrase preceded by *pe* in (21b) bears only a specific interpretation. It is important to stress that although indefinite direct objects are generally interpreted as having wide scope readings, not all wide scope indefinites host *pe*.

The wide and narrow scope readings mentioned above arise in the presence of operators, such as universal quantifiers, verbs of propositional attitudes, negation, conditionals, or other operators. In contexts that lack any operators (i.e. in plain and transparent contexts), indefinite noun phrases are assumed to be lexically ambiguous between a quantificational and a referential interpretation. This type of specificity is generally known as *epistemic specificity* (Karttunen 1969, Fodor and Sag 1982, Farkas 1994, among others). The fact that the indefinite *a boy* in (22A) has two different readings is underlined by the two possible continuations given in (22b) and (22c), whereas the first continuation is the specific one, and the latter the non-specific.

- (22) A: A boy is playing in the yard.  
 (a) His name is James.  
 (b) I don't know him.

Ludlow and Neale (1991) assume that the difference between a referential and a quantificational use arises from the amount of information available to identify the referent. There is a widespread opinion that in order for an indefinite noun phrase to be used with a specific interpretation the speaker must have *a particular individual in mind* (e.g. Kasher and Gabbay 1976, Fodor and Sag 1982, Kratzer 1998, Yeom 1998, van Rooy 1999). Accordingly, the speaker uses a specific indefinite if he wants to indicate that he knows a particular individual, without presupposing that the hearer knows the identity of that referent as well. Conversely, a non-specific indefinite is quantificational and does not refer.

The epistemic use proposed by Farkas (1994) is an extension of Fodor and Sag's (1982) approach, as the notion of epistemic specificity can be applied to cases that are not limited to referential uses of indefinites. *Pe*-marked indefinite noun phrases in plain and transparent contexts have been thought of in terms of epistemic specificity as well. The indefinite noun phrase in (23a) is analysed as an instance of epistemic specificity, as it is said to reflect the knowledge of the speaker about the identity of the referent. The sentence without *pe*, on the contrary, is again ambiguous between a specific and a non-specific reading, as illustrated by the translation of sentence (23b).

- (23) (a) Petru l -a văzut pe un băiat. (Dar eu nu știu pe cine)  
 Peter CL Aux. seen PE a boy. But I don't know who.  
 'Peter saw a boy (but I don't know whom).' [specific]
- (b) Petru a văzut un băiat. (Dar eu nu știu pe cine)  
 Peter Aux. seen a boy. But I don't know who.  
 'Peter saw a boy. (But I don't know whom).' [specific/ non-specific]

The studies dealing with the behaviour of *pe*-marked indefinites in plain and transparent contexts limited their analyses to one-sentence contexts. This proves to be insufficient, when analysing natural occurring sentences in larger sentence fragments, as in (23) above.

If *pe* were to show a contrast with respect to epistemic specificity as defined above, we would expect (i) the unmarked case (23b) to allow both the specific and the non-specific reading of the direct object referent, and (ii) the *pe*-marked form in (23a) to only allow the specific reading. More concretely, the indefinite direct object preceded by *pe* in (23a) is analysed as referring to a particular boy, say Mike, which the speaker ‘knows’, while the indefinite in (23b) could either refer to a boy the identity of which is unknown to both speaker and hearer, or to a particular boy. However, an epistemic specific reading is generally incoherent with the assertion that the speaker does not know the referent. As both sentences (23a) and (23b) are compatible with the continuation sentence given in parenthesis, which denies the speaker’s knowledge of the identity of the referent, it seems that having something in mind is not a prerequisite for *pe*-marking. Thus, the presence of *pe* in (23a) cannot be motivated in terms of speaker’s knowledge of the identity of the referent.

Following the ideas developed in von Stechow (2002) and Kornfilt and von Stechow (2009) with regard to “anchored specificity”, von Stechow and Onea (2008) analyse *pe*-marking in Romanian in these terms. “Anchored specificity” was developed to cover a wide range of uses of specificity, providing a means to distinguish and relate to each other several subtypes of specificity. Anchored indefinite noun phrases may be bound to other discourse referents besides the speaker. This means that there must be a salient function or relation between the referent of the indefinite noun phrase and a visibly present discourse referent (i.e. the “referential anchor”). Crucially, any anchored indefinite noun phrase is specific and the nature of the referential anchor determines the specificity subtype involved. The anchor can be on the speaker, on another referent in the current utterance, or on a non-atomic discourse referent.

When the anchor is on the speaker of the sentence, the referent of the indefinite noun phrase is known to or identifiable by the speaker. Assuming speaker anchoring, the sentence (24) would yield the reading *I know who the neighbour is*. In other words, there is a function such that the referent of the indefinite noun phrase is the value of the function, which takes the speaker as an argument. This type of specificity has the widest scope because the anchor is fixed outside the scope of any sentential operators or scope islands. Speaker anchored specificity resembles the epistemic specificity notion discussed above as both involve a certain degree of speaker knowledge. However, the notion of specificity

anchored to the speaker can be extended to encompass more contexts, not being limited to a function that has the reading ‘what the speaker has in mind.’

(24) Peter: John saw a certain neighbour of his.

Speaker-anchored reading: Peter knows who the neighbour is.

Locally-anchored reading: John knows who the neighbour is.

A second subtype of anchored specificity is one in which the referential anchor of the indefinite noun phrase occurs within the current sentence. Thus, in example (24), the indefinite *a certain neighbour of his* could be anchored to the subject of the matrix sentence, *John*. Local anchoring indicates that the indefinite cannot get wide scope over some operator that has scope over its anchor. However, it still expresses specificity as it is anchored to a discourse referent. Accordingly, locally anchored indefinites can have narrow or intermediate scope. Another type of indefinites are non-anchored indefinites which have narrow scope and are mere existential in opaque contexts. In plain and transparent contexts they are neutral towards specificity.

With this fine-grained structure of specificity in place, von Heusinger and Onea (2008) propose the specificity scale presented in Table 5.2 to account for *pe*-marking in Romanian. Their predictions towards the specificity scale below are that *pe*-marked indefinite direct objects found in opaque contexts express specificity as they are locally bound to a referent in the sentence.

Table 5.2: Scale of specificity (von Heusinger and Onea 2008)

Speaker-anchored	Locally anchored	Non-anchored	Incorporated
------------------	------------------	--------------	--------------

The prediction with respect to indefinite noun phrases in plain and transparent contexts, which constitute the focus of this dissertation, is that *pe*-marked indefinites express specificity because their referent is speaker-anchored. On the contrary, non-*pe*-marked referents are non-specific because they are not anchored (neither to the speaker, nor to any other visibly present referent in the text). If this analysis of indefinite noun phrases in transparent contexts is correct, then we would expect the referent of the specific indefinite *pe o asistentă* (‘PE a nurse’) in (25) to be functionally dependent on another expression in the context. However, such a reading does not hold, as the sentence can be continued with

a comment that inhibits the anchoring of the referent to the speaker or to the subject of the matrix sentence.

(25) Rareș a văzut-o *pe o asistentă* ieșind din cabinet, noi nu știm cine este ea. Important este că ne putem furișa înăuntru acum.

‘Rareș saw PE a nurse leaving the practice, we don’t know who she is. What matters is that we can sneak inside now.’

It seems that an analysis which involves a certain degree of speaker knowledge or which presupposes an anchoring relation between the speaker and the referent of the indefinite description cannot properly disambiguate between the two types of indefinites in direct object position in plain and transparent contexts as in (25) above.

Most importantly, however, an analysis of *pe*-marked indefinites in terms of anchored specificity does not include the special status retained by such referents in the subsequent discourse. Consider the two examples in (26) and (27), which are taken from two different newspaper articles, which describe the same shooting event. Note that the indefinite noun phrases found in direct object position constitute minimal pairs, in the sense that their descriptive material and the verbs selecting these objects are almost the same.

(26) *pe*-marked indefinite NP in a larger discourse segment with English translations<sup>31</sup>

[S1] Neculai Florea, de 40 de ani, viceprimarul satului Horodniceni, și-a pus poliția pe cap după ce l-a împușcat cu un pistol cu gloanțe de cauciuc **pe un tânăr** din localitate.

[S2] Incidentul s-a petrecut în noaptea de 10 spre 11 februarie, la discoteca ce aparține soției viceprimarului Florea și a fost reclamat la poliție în cursul după amiezii, la ora 15:40.

[S3] La ora respectivă, **Vasile M.**, de 24 de ani, din comuna Horodniceni, **pro** s-a adresat postului de poliție reclamând că **pro** a fost împușcat în picior de viceprimarul Neculai Florea.

[S4] Din primele verificări efectuate s-a stabilit că în cursul nopții, la discoteca

[S1] The 40-year-old Nicolae Florea, the vice mayor of the Horodniceni village, angered the police after he shot **a young man** from the same village with a gun with plastic bullets.

[S2] The incident took place on the night of February 10<sup>th</sup> in the discotheque, whose owner is Florea’s wife, while the police were notified at 15:40 o’clock in the afternoon.

[S3] At that time, the 24-year-old **Vasile M.**, from the Horodniceni village complained to the police that **he** was shot in the leg by the vice-mayor Neculai Florea.

[S4] In keeping with initial findings, it was established that during the night an altercation

<sup>31</sup> <http://Section.obiectivdesuceava.ro/index.php?ids=26841andpage=articol> [viewed on September 1, 2009]

viceprimarului, pe fondul consumului de alcool, a avut loc o altercație, iar Neculai Florea a folosit pistolul cu gloanțe de cauciuc împotriva lui **Vasile M.**, pe care **l-a împușcat în picior, rănindu-l.**

[S5] Viceprimarul Neculai Florea susține că a fost nevoit să facă uz de armă, întrucât a fost agresat **de tânărul în cauză.**

[S6] A spus că în cursul nopții de 10 spre 11 februarie, în discoteca administrată de soția lui a izbucnit un scandal între două grupuri rivale de tineri.

[S7] "Soția mea m-a chemat și am intervenit ca să liniștesc apele.

[S8] Am încercat să stau de vorbă, să-i calmez, dar **băiatul acela** m-a lovit în piept și era cât pe ce să [...].

took place at the vice mayor's discotheque due to alcohol consumption. Neculai Florea used his gun with plastic bullets against **Vasile M.**, **whom** he shot in the leg, hurting **him.**

[S5] The vice-mayor Neculai Florea maintains that he had to make use of his gun, as the mentioned young man shoved him.

[S6] He said that on the night of February 10<sup>th</sup>, a scuffle broke out between two rival groups of young men in the discotheque administered by his wife.

[S7] "My wife called me and I came to calm down the situation.

[S8] I tried to talk to them, to calm them down, however, **that boy** hit me in the chest and **he** almost [...].

(27) non-*pe*-marked indefinite NP in a larger discourse segment with English translations<sup>32</sup>

[S1] Viceprimarul Neculai Florea, din comuna Horodniceni, este cercetat de poliție după ce în noaptea de sâmbătă spre duminică a împușcat în picior **un tânăr** de 24 de ani la discotecă.

[S2] Viceprimarul, care este membru PNG, a scos pistolul pentru a interveni într-o încăierare între tineri, care avea loc în discoteca familiei sale.

[S3] El este asociat unic, iar soția sa administrator.

[S4] Poliția a stabilit că **tânărul împușcat, Vasile Mihai**, pe fondul consumului de alcool, **pro** a fost implicat într-un scandal, iar viceprimarul a intervenit pentru a-**l** stopa.

(no further coreferential expressions)

[S1] The vice mayor Neculai Florea from the village Horodniceni, is verified by the police after he shot **a 24-year-old young man** in the leg in the night from Saturday to Sunday in a discotheque.

[S2] The vice mayor, who is a PNG member, took his gun out in order to intervene in a quarrel which started in his family's discotheque between some young men.

[S3] He is the owner and his wife the administrator.

[S4] The police found out that the **young man, Vasile Mihai**, was shot due to alcohol consumption, and that **(he)** was involved in a scuffle, and that the vice mayor intervened in order to stop **him.**

(no further coreferential expressions)

The referent of the direct object preceded by *pe* in (26) is introduced in the sequence [S1] as *pe un tânăr din localitate* ('PE a young man from the village'). Within the next sentences, [S2] through [S8], the referent is mentioned nine times. The referent *un tânăr de 24 de ani* ('a 24-year-old young man') is introduced in the second newspaper article,

<sup>32</sup> <http://Section.9am.ro/stiri-revista-presei/2007-02-13/un-viceprimar-a-impuscat-un-tanar-in-discoteca.html>

[viewed on September 1, 2009]

(27), in sentence [S1] as a non-*pe*-marked indefinite noun phrases and is taken up in the subsequent text only three times, in [S4]. It seems that the *pe*-marked referent is more referentially persistent than its unmarked counterpart. Another interesting observation is that the critical *pe*-marked and non-*pe*-marked referents are not mentioned in the immediately following sentences (i.e. [S2]), but only after the second continuation sentence on (i.e. in [S3] in (26) and in [S4] in (27)). Note furthermore that the *pe*-marked referent in (26) becomes the sentential topic in the ensuing discourse, whereas the non-*pe*-marked referent does not achieve this status in (27). In sum, we observe that *pe*-marked referents are more frequently mentioned and are more prone to shift the topic in the subsequent discourse.

On an intuitive level, the observation is that *pe*-marked indefinites are more important than their unmarked counterparts. Building on the work of Prince (1981), Maclaran (1982), Fodor and Sag (1982) and others, Ionin (2006) introduces an additional discourse-pragmatic type of specificity, which she calls *specificity as noteworthiness*. In her view, a specific indefinite noun phrase is used whenever the speaker intends to signal that the referent realized in this way is more important or salient in the discourse on the basis of a particular property of that referent. The noteworthiness of a referent can be expressed by means of a predicate, a particular adjectival modification, a previous statement elaborating upon the noteworthy property, or from subsequent coreferential expressions. Ionin (2006) uses the concept of noteworthiness in order to account for the distribution of indefinite *this* in English. Indefinite *this* expresses discourse specificity (Perlman 1969, MacLaran 1982, Ionin 2006) by introducing new referents in the discourse that are of greater importance for the text segment and therefore specified, elaborated or taken up in the subsequent discourse (see the discussion in Chapter 3).

Ionin's (2006) analysis of specificity as noteworthiness applied to Romanian DOM in plain and transparent contexts seems to correctly predict that indefinite noun phrases host *pe*-marking if they are noteworthy, which is reflected in the referential persistence or continuity of the referents marked in this way. However, an analysis in terms of specificity as noteworthiness as proposed by Ionin (2006) would have to account for the fact that both indefinites in (26) and (27) above are taken up in the following discourse. In other words, the predictions concerning the noteworthiness of indefinite noun phrases that host *pe* should be modelled in such a way as to encompass the fact that the non-*pe*-marked

indefinite referent is taken up in the discourse as well. Furthermore, an analysis of *pe* in terms of noteworthiness should be able to account for the fact that the *pe*-marked referents are not immediately picked up after being introduced in the discourse, but with a delay.

In sum, every subclass of specificity discussed in this Section can host *pe*, and in all these cases *pe* disambiguates more or less between specific and non-specific readings. The most stable and systematic observation was made with respect to indefinite noun phrases in opaque contexts, which are generally used with a specific reading if they are *pe*-marked. However, the correlation between *pe*-marking on indefinite noun phrases and specificity is imperfect, as we find cases of specific objects that do not host *pe* and epistemically non-specific indefinite objects that allow for *pe*-marking. These observations indicate that the seemingly optional distribution of *pe*-marking with indefinite noun phrases can be accounted for, if further factors are taken into account. In the remainder of this Chapter I propose that the optionality can be explained if the larger *discourse context* in which indefinite noun phrases are used is considered. In some respects, my analysis builds up and extends Ionin's (2006) concept of noteworthiness. Like her analysis, the analysis put forth in this dissertation assumes that *pe*-marked referents retain a special status in the discourse. However, in the present investigation I focus on the role played by a *pe*-marked indefinite in direct object position in larger discourse units in terms of several concrete effects, such as referential persistence, topic-shift potential and the type of anaphoric referring expression used to pick up these referents in the subsequent discourse. In the next Section I review several studies that explored case-marking on direct objects in terms of topicality at sentence level.

### 5.3.2 Topicality and *pe*-marking with indefinite noun phrases

Having seen that the correlation between *pe*-marking on indefinite noun phrases and specificity is imperfect, let us now explore whether referentiality may play a role in the distribution of *pe*-marking with indefinite noun phrases in Romanian. Many authors have pointed out that DOM in general, and the specific uses of indefinite noun phrases in particular, are related to the topical status of these indefinites compared to their non-marked counterparts. It is important to stress out before we go any further that the notion of topicality is not used in a unitary way among researchers. Leonetti (2003) and Escandell-

Vidal (2009) use the term topicality in the Reinhartian “aboutness” sense, signalling given information. They argue that the contribution of topical referents is to “introduce prominent participants in the discourse, [...] indicating that the individualization of the referents is relevant for utterance interpretation” (Leonetti 2003:8). Naess (2004) discusses topicality in terms of salience and defines it as a characteristic of entities that are more important in terms of human perception in general. De Swart (2007:138) defines topicality in terms of “the centrality of an entity in the discourse or with the readiness with which an entity presents itself to the speaker as a topic of conversation”. More recently, Nikolaeva (2001) and Dalrymple and Nikolaeva (2011, to appear) propose an elaborated notion of topicality, which deals with “the construal of the referent as pragmatically salient so that the assertion is made about this referent”. The aim of this Section is not to offer an exhaustive overview of the notion of topic, but rather to present several relevant studies that accounted for DOM cross-linguistically in terms of topicality (for a more ample discussion of the notion of topic in general, see Chapter 2). I will argue that former analyses cannot account for the distribution of *pe*-marking with indefinites in Romanian.

Following Laca (1987) who argued that topicality is as an indispensable notion for the analysis of DOM in Spanish, Leonetti (2003) and Escandell-Vidal (2009) propose a more elaborated analysis of the differential object marker *a* in Spanish and Balearic Catalan along these lines. Spanish and Balearic Catalan DOM marking resembles (but is still significantly different from) Romanian, as case-marking depends on the argument features animacy and definiteness (Laca 1987, Leonetti 2003, von Heusinger 2008). Leonetti’s (2003) notion of topicality or prominence is part of the information structure, but simultaneously of event structure (when talking about autonomous vs. incorporated constructions). Leonetti (2003) considers that the topical status of *a*-marked direct objects forces a specific or generic reading of the indefinite description (as it is generally the case with indefinite noun phrases in topical position which require referentially autonomous entities), and not the other way round. Leonetti (2003) draws support for his analysis from certain syntactic constructions, which necessarily trigger *a*-marking, from the diachronic evolution of DOM in Spanish and from semantic considerations. More precisely, he shows that clitic-left dislocated constructions require *a* and that subjects and *a*-marked objects show parallel characteristics, as both can be realized as bare plurals only if they are further modified by a restrictive modifier. Moreover, like subjects, the correlation of the presence of *a* and the occurrence of a secondary predicate is presented as an additional argument to

treat the preposition a topic marker. Further evidence involves binding or scope asymmetries stemming from the presence or absence of the marker. Example (28) is from Romanian and shows how case-marking on the object can turn into a ‘prominent binder’ for a pronoun inside the indirect object. Sentence (28a), in which the direct object is *pe*-marked, favours a reading in which the possessive *ei* (‘hers’) is coreferent with the direct object referent. On the contrary, the unmarked indefinite in (28b) is rather interpreted as not being coreferent with the possessive pronoun, thus referring to another female referent outside the context. In (28a), the case-marker underlines the topic status of its referent, making the indefinite noun phrase a prominent antecedent for an anaphoric pronoun.

- (28) (a) Am trimis -o pe o fată<sub>1</sub> la şeful ei<sub>1</sub>.  
 Have send CL PE a girl at boss her
- (b) Am trimis o fată<sub>1</sub> la şeful ei<sub>2</sub>.  
 Have send a girl at boss her  
 ‘I send a girl to her boss.’

In sum, Leonetti (2003) and Escandell-Vidal (2009) focus more on the semantic factors, or on the referential properties of different noun phrases that contribute to the topical status of direct object arguments. For the present analysis of *pe*-marking in Romanian, I employ a distinction between the semantic properties of a direct object argument, on the one hand, and the discourse status the referent associated with it retains in the subsequent discourse, on the other hand. I will show that the association between the semantic features of a referent and its discourse status is less than perfect, as we find semantically prominent referents that are not discourse prominent.

An alternative account on DOM in terms of topicality was advanced by Nikolaeva (2001) and Dalrymple and Nikolaeva (2011, to appear). They propose a cross-linguistic analysis of optional case-marking in some nominative-accusative languages, elaborating upon former accounts on DOM as topicality. As Leonetti (2003) and Escandell-Vidal (2009), they analyse topicality in the “aboutness” sense. A topical referent is one that is the current center of interest in the conversation. Furthermore, the notion of topic discussed by Dalrymple and Nikolaeva (2011, to appear) somewhat resembles the backward-looking center used by Prince and Beaver (2004), as the topical referent connects the present utterance with the preceding one. Dalrymple and Nikolaeva (2011, to appear) make a

distinction between primary topics, which are highly prominent and typically realized as subjects (see Comrie 1975, Keenan 1976, Givón 1983, Lambrecht 1994, among others) and between secondary topics, which are less prominent (see Givón 1983 for a similar categorization of topics). According to this division, a sentence may have more than one topic. The (inherent) semantic factors considered the triggering factors for DOM are not sufficient to determine the topicality of a referent, which rather depends on the pragmatic salience of the referent, which in turn relies on the speaker. In their functional analysis, topical objects are marked not because they have to be distinguished from subjects, but because languages tend to mark topics.

The analysis of Dalrymple and Nikolaeva (2011, t.a.) investigates the syntax of objects associated with different information structural roles (topic vs. non-topic/ focus). In her analysis of Ostyak, Nikolaeva (2001) convincingly shows that object agreement is sensitive to focus structure, an observation that supports her proposal. However, *pe*-marking in Romanian is not sensitive to focus structure, but is dependent on semantic factors as well. Thus, Nikolaeva's (2001) analysis cannot be applied to all instances of *pe*-marking and cannot be adopted for the analysis of case-marked indefinites in Romanian.

Dalrymple and Nikolaeva (2011, t.a.) consider brand-new referents unidentifiable for the addressee and thus non-topical entities. This is a reason why indefinite noun phrases do not make good topics. Exceptions to this rule are pragmatically anchored indefinites that are linked to another identifiable entity in the consciousness of the interlocutors. Such referents are roughly the same as the specific indefinites defined by Enç (1986). The pragmatic link to a previously established referent (i.e. a recoverable relation or a subset relation) ensures that the referent of the specific indefinite expression is identifiable to the speaker. This situation is different for non-specific indefinites, as neither the speaker nor the addressee can identify the referent. Their definition of topic as a referent that the utterance is presupposed to be about, states that non-referential expressions cannot be topics. In other words, topical status forces a specific interpretation of an indefinite noun phrases (see Erteschik-Shir 1997, Geurts 2002, Leonetti 2003, and reference cited therein for extended discussions). In Section 5.3.1 I discussed several types of specificity in connection to Romanian *pe*-marked indefinites. I showed that neither locally anchored specificity, nor speaker-anchored specificity can explain the optionality encountered with indefinite noun phrases in transparent contexts. Thus, Dalrymple and Nikolaeva's (2001, t.a.) proposal

cannot be applied to the distribution of *pe*-marking with indefinite noun phrases in Romanian. Furthermore, an analysis in terms of topicality should enumerate the defining properties of topical elements.

Despite employing a somewhat different terminology, the studies accounting for DOM in terms of topicality, have several common denominators. First, in many languages DOM was shown to have its origin as a marking device for topical arguments. In a first stage, only topical referents host the case-marker, whereas non-topical referents remain unmarked. At a next stage, DOM spreads, applying to non-topical elements that have (semantic) features typical for topics. As a consequence, the role of information structure is diminished lost, DOM being eventually dependent on semantic features alone. Second, much research on DOM addressed the relationship between the referential properties of a noun phrase and its topical status (Leonetti 2003, Naess 2004, De Swart 2007). Based on the observation that topical referents are generally semantically prominent in terms of animacy, definiteness and specificity, these studies have proposed a distinction between the semantic prominence and topicality or discourse prominence.

### 5.3.3 Summary

Summarizing, I showed that direct objects realized as pronouns and proper names are always *pe*-marked in synchronic Romanian. The picture of the distribution of *pe* becomes more complex moving along the referentiality scale to the right, as case-marking for more explicit types of referring expressions (especially for definite unmodified and indefinite noun phrases) cannot be accounted for in terms of animacy and referentiality alone. Besides situations in which the morphologically marked form and the unmarked one are in complementary distribution, excluding one another, it was shown that definite (unmodified) noun phrases can occur in free variation as well, in the sense that both forms are allowed in the same contexts, without triggering different readings. Lastly, I showed that specificity is not enough to account for the cases of variation with indefinite noun phrases.

Table 5.3 summarizes the referential contexts in which direct objects receive *pe*-marking in Romanian. Besides the referring type of the direct objects, the Table also makes a clear distinction in the domain of indefinite nominal phrases with respect to different types of

specificity. The distinction is, on the one hand, between specific and non-specific indefinite noun phrases according to which non-specific indefinites do not host *pe*. On the other hand, Table 5.3 includes a further distinction in the realm of specific indefinites, namely between contexts containing sentence operators and contexts that do not contain any (i.e. transparent contexts). As it was shown, *pe*-marked indefinite noun phrases show a clear preference for referential readings in contexts that contain sentence operators, whereas the non-*pe*-marked forms are generally neutral towards this preference. In plain and transparent contexts, the distribution of the direct object marker was shown to be less predictable, as animacy, definiteness and specificity cannot thoroughly account for its distribution. Neither topicality (in the sense of Leonetti 2003 or Nikolaeva 2001), nor other global parameters (like the lexical properties of the governing verb or secondary predications, etc.) are general enough or useful to explain the variation found with indefinite noun phrases.

Table 5.3: Referentiality and Specificity Scales for [+human] *pe*-marked direct objects

definite noun phrase	> indefinite noun phrase		
	scopal specificity	epistemic specificity	non-specific
+/-	+	+/-	-

In what follows, I account for the problematic examples involving indefinite noun phrases in transparent contexts by adding a discourse-pragmatic parameter on the list of the factors licensing DOM in Romanian. Exactly like English indefinite *this* and German *so'n*, *pe*-marking will be shown to function as a discourse structuring device.

#### 5.4 The discourse structuring potential of *pe*-marking in Romanian

The preceding Sections showed that the distribution of *pe*-marked indefinite direct objects in transparent contexts in terms of epistemic specificity, anchored specificity or topicality cannot explain the privileged status *pe*-marked referents have in the subsequent discourse. Applied to one-sentence contexts, a motivation of *pe* in terms of epistemic specificity or speaker anchoring seems intuitively correct, but insufficient to account for the discourse structuring potential of a *pe*-marked referent in a larger discourse unit, as example (29) shows.

(29) Rareş a văzut-o *pe o asistentă* ieşind din cabinet, noi nu ştim cine este ea. Important este că ne putem furişa înăuntru acum.

‘Rareş saw PE a nurse leaving the practice, we don’t know who she is. What matters is that we can sneak inside now.’

In (29), the indefinite noun phrase *o asistentă* (‘a nurse’) preceded by *pe* could be interpreted as referring to a specific nurse the speaker knows; however, the continuation sentence definitely denies this knowledge. In this dissertation I will propose an alternative analysis of *pe* in terms of a device that gives structure to the discourse and contributes to the overall coherence of the discourse. The contrast between the referential persistence of the *pe*-marked indefinite direct objects and the unmarked ones observed above is surprising from the point of view of the existing literature, which aims for an explanation of DOM in Romanian in terms of a set of triggering factors which apply at sentence level. The crucial question, of course, is what exactly it is about these two direct object forms that make the difference in terms of presence or absence of the *pe*-marker in transparent contexts. More concretely, does the referential persistence of the *pe*-marked referent reflect a higher degree of accessibility or prominence of the referent marked in this way? Before turning to this question in more detail, we also have to worry about whether the contrast in referential persistence observed above is a general one, or not, i.e. one that can be replicated across examples and be confirmed systematically by a representative corpus. To address these concerns, a web-based story continuation experiment was carried out. The findings of the experiment are reported in the next Section.

#### 5.4.1 The sentence-continuation experiment on *pe*-marking

The sentence-continuation study was designed to investigate whether there is a systematic difference in terms of discourse structuring effects between the *pe*-marked referents and the non-*pe*-marked ones. The textual parameters developed in Chapter 2 were used to measure the discourse structuring potential of indefinite noun phrases in direct object position. To this end, the (i) referential persistence, the (ii) topic shift potential of the *pe*-marked referents and (iii) the type of referring expression used to pick up the referent for the first time were determined. Recall that different investigations (Kehler et al. 2008, Kaiser 2010, among others) showed that referential persistence, on the one hand, and the type of anaphoric expression used for a given referent, on the other hand, do not always coincide

and thus possibly indicate different properties of a referent. Starting from the premise that *pe*-marked referents show a similar discourse structuring potential as English indefinite *this* and German *so 'n*, following predictions can be made:

*Prediction 1 (referential immediacy and referential persistence)*

- (c) If they are used as discourse structuring devices, referents introduced by noun phrases headed by *pe* will be more readily picked up in the subsequent discourse (compared to their simple indefinite counterparts).
- (d) In Chapter two, we discussed referential persistence (i.e. the frequency of subsequent mention) as a reflection of a discourse-structuring device. If referents introduced by noun phrases headed by *pe* display the potential to give structure to the discourse, then their associated referents will show a high frequency of subsequent mention (i.e. referential persistence).

*Prediction 2 (topic-shift potential)*

Referents of noun phrases preceded by indefinite *pe* will show the potential to shift the topic in the subsequent discourse (see discussion in Chapter 3.2.2).

*Prediction 3 (referential type of the first anaphoric expression)*

- (c) If the referent associated with indefinite *pe* is a marker of accessibility in the sense of Gernsbacher and Shroyer (1989) and other accessibility-based accounts, the likelihood of subsequent mention of this referent by means of a pronoun should be high.
- (d) If, on the contrary, *pe*-indefinites are discourse-structuring devices, then they will not impact the type of the first referring expression used to take up the referent (see discussion in Chapter 2.4.3). In other words, the mere presence of *pe* will be neutral towards the first type of anaphoric expression used to refer back to its referent. This prediction derives from recent findings presented by Kehler et al. (2008) and Kaiser (2010), which showed that the expectancy of subsequent mention of a particular referent does not necessarily have to coincide with the type of referring expression used to pick that referent up. For example, the likelihood that a referent will be mentioned in the following discourse does not necessarily mean that this referent is more likely to be realized by a reduced type of referring expression (e.g. a pronoun).

### 5.4.1.1 The experimental setup

#### *Method*

The experiment employed a multi-sentence story-continuation task. The elicited sentence-continuations are interpreted as a measure of participants' expectations about the development of the discourse, given the events introduced in the initial mini-context, the differential marking of the critical items, and the participants' prior experience with the way discourses naturally unfold.

#### *Participants*

Twenty-three native Romanian speakers participated in the experiment. They received no incentive for taking part in the experiment. There was no time limit to complete the experiment, but a time frame of twenty minutes was proposed in the instructions to encourage participants to write spontaneous continuations, moving relatively quickly through the items.

#### *Design and Procedure*

The methodology used in this experiment was the off-line sentence-continuation task already introduced in Chapter 2. Participants' task was to read the given sentence fragments and add five logical and natural-sounding sentence-continuations for each of them. The first two sentences of each test item set the context of the story, and contained individual references to two characters. The first character was the clearly established topic of the story fragments, whereas the second character was the critical item realized as an indefinite noun phrase in direct object position. Furthermore, note that I used an open-ended sentence-continuation task with no pronoun prompt.

#### *Materials*

Six target stories were constructed and no fillers were added. Each target story consisted of three sentences and I only manipulated the realisation form of the critical direct objects, which resulted in two conditions: one in which *pe* heads the indefinite direct object and one in which the indefinite direct object remains unmarked or "bare", as illustrated in Table 5.4.

Table 5.4: Sample experimental items in the two conditions

TI1: <i>pe</i> -marked condition	TI1: non- <i>pe</i> -marked condition
<p><i>Aseară a fost extraordinar de cald. Pentru că nu mai rezista în casă, Graur s-a hotărât să se ducă în oraș. Pe drum l-a văzut pe un copil intrând într-un magazin.</i></p> <p>‘It was extraordinarily warm outside yesterday evening. Because it was unbearable for him to stay home anymore, Graur decided to go downtown. On his way there he saw <i>pe</i> a child entering a store.’</p>	<p><i>Aseară a fost extraordinar de cald. Pentru că nu mai rezista în casă, Graur s-a hotărât să se ducă în oraș. Pe drum a văzut un copil intrând într-un magazin.</i></p> <p>‘It was extraordinarily warm outside yesterday evening. Because it was unbearable for him to stay home anymore, Graur decided to go downtown. On his way there he saw a child entering a store.’</p>
TI2: <i>pe</i> -marked condition	TI2: non- <i>pe</i> -marked condition
<p><i>Anul trecut când am fost la mare am cunoscut un salvamar<sub>1</sub>. (pro<sub>1</sub>) Era tot timpul activ. La sfârșitul sejurului meu, (pro<sub>1</sub>) a salvat-o pe o fată<sub>2</sub> de la înec.</i></p> <p>‘Last year when I was at the seaside I met a lifeguard<sub>1</sub>. (He<sub>1</sub>) was very active all day long. At the end of my stay there, (he<sub>1</sub>) saved <i>PE</i> a girl<sub>2</sub> from drowning.’</p>	<p><i>Anul trecut când am fost la mare am cunoscut un salvamar<sub>1</sub>. (pro<sub>1</sub>) Era tot timpul activ. La sfârșitul sejurului meu, (pro<sub>1</sub>) a salvat o fată<sub>2</sub> de la înec.</i></p> <p>‘Last year when I was at the seaside I met a lifeguard<sub>1</sub>. (He<sub>1</sub>) was very active all day long. At the end of my stay there, (he<sub>1</sub>) saved a girl<sub>2</sub> from drowning.’</p>
TI3: <i>pe</i> -marked condition	TI3: non- <i>pe</i> -marked condition
<p><i>Oaspeții stăteau neliniștiți în sufragerie, neștiind cu cine să înceapă o conversație. Ana l-a văzut pe un bărbat stând singur lângă canapea. Între timp, chelnerii aduceau mâncarea.</i></p> <p>‘The guests were anxiously sitting in the living room, not knowing whom to start a conversation with. Ana saw <i>pe</i> a man sitting alone on the couch. Meanwhile, the waiters were bringing food.’</p>	<p><i>Oaspeții stăteau neliniștiți în sufragerie, neștiind cu cine să înceapă o conversație. Ana a văzut un bărbat stând singur lângă canapea. Între timp, chelnerii aduceau mâncarea.</i></p> <p>‘The guests were anxiously sitting in the living room, not knowing whom to start a conversation with. Ana saw a man sitting alone on the couch. Meanwhile, the waiters were bringing food.’</p>

The first two sentences of each story fragment set the context and contained an individual reference to a character, other than the new referent of the direct object introduced in the third sentence. This first-introduced character was the global topic of the mini-discourse, as it was mentioned in subject position (at least in the first sentence) and was the referent the

story was about. Although the inclusion of two context sentences made it difficult to control every aspect of these discourses, it provided the advantage of creating a more natural discourse. In the third stimulus sentence, the critical referent was introduced as an indefinite noun phrase in direct object position (i.e. the target item). The third target item (TI3) contained an additional last sentence in which a third referent was introduced (e.g. *chelnerii* ‘the waiters’).

#### 5.4.1.2 Procedure and data analysis

The first five main clauses (including subordinate ones, if there were any) of each continuation story provided by the participants, were coded by two independent judges for three aspects of the indefinite direct objects: (i) their referential persistence, (ii) their topic shift potential, and (iii) the type of anaphoric referring expressions used to pick up the referent for the first time. The two judges agreed upon 92% of the cases; in the case of disagreement (which were mostly concerned with determining the topic), differences were resolved through discussion. The three factors mentioned above were employed as a means to establish the discourse structuring potential of the *pe*-marked and non-*pe*-marked direct object referents. Example (30) presents one response for the *pe*-condition for test item 2 (TI2), and Table 5.5 illustrates the coding methods used. I used subscript 1 for the subject referent, the lifeguard, and subscript 2 for the direct object referent, the girl. The type of referring expression (e.g. *pro*, clitic, personal pronoun, definite noun phrase and definite modified noun phrase) and the grammatical function of the anaphora of both subject and object referents are listed in Table 5.5. Round brackets mark subordinate clauses and square brackets main clauses. Referential persistence is measured by referents mentioned per sentence (referent / S) and the sum of all items up to S5 (i.e. a cumulative measure). Comparing the sums allows us to see at what stage in the discourse we have more anaphoric expressions referring to the direct object referent than to the subject. In our example (30), the referent of the direct object exceeds the subject referent in persistence in S3. And finally, I verified in what sentence the direct object referent becomes the aboutness topic in a main clause, which I called ‘topic shift’. In example (30) above, this happens in sentence-continuation 3 (i.e. S3).

(30) Example responses (with English translation) for an item in the *pe*-condition

### Sample experimental item

*Anul trecut când am fost la mare am cunoscut un salvamar<sub>1</sub>. (Pro)<sub>1</sub> era tot timpul activ. La sfârșitul sejurului meu, (pro)<sub>1</sub> a salvat-o pe o fată<sub>2</sub> de la înec.*

*'Last year when I was at the seaside I met a lifeguard<sub>1</sub>. He<sub>1</sub> was very active all day long. At the end of my stay there, he<sub>1</sub> saved PE-a girl<sub>2</sub> from drowning.'*

### Sample continuation sentences for the *pe*-condition

S1	<i>M-a mirat ca (pro)<sub>1</sub> a reusit sa o<sub>2</sub> salveze, pentru ca (pro)<sub>1</sub> era un burtoș.</i> 'I was surprised that he <sub>1</sub> could save her <sub>2</sub> , because he <sub>1</sub> was a big-bellied man.'
S2	<i>Dar cum (pro)<sub>1</sub> a vazut-o pe fata<sub>2</sub> ca (pro)<sub>2</sub> striga dupa ajutor, (pro)<sub>1</sub> nu a stat pe ganduri si (pro)<sub>1</sub> s-a dus dupa ea<sub>2</sub>.</i> 'But as soon as he <sub>1</sub> saw PE-girl <sub>2</sub> screaming for help, he <sub>1</sub> did not hesitate and went after her <sub>2</sub> .'
S3	<i>Fata<sub>2</sub> era inconștienta cand (pro)<sub>2</sub> a fost adusa la tarm.</i> 'The girl <sub>2</sub> was unconscious as she <sub>2</sub> was brought to shore.'
S4	<i>(pro)<sub>2</sub> nu avea mai mult de 10 ani.</i> 'She <sub>2</sub> was not older than 10.'
S5	<i>Parintii ei<sub>2</sub> i-au multumit salvamarului<sub>1</sub>.</i> 'Her <sub>2</sub> parents thanked the lifeguard <sub>1</sub> .'

Table 5.5: Coding methods for the continuation sentences in (30)

	Coding methods	Subject		Object		topic
		refer per item / S	sum	refer per item / S	sum	
S1	Anaphoric forms and grammatical function (pro <sub>1</sub> , Cl <sub>2</sub> ) (pro <sub>1</sub> ) (Sub <sub>1</sub> , DO <sub>2</sub> ) (Sub <sub>1</sub> )	2	2	1	1	Topic <sub>1</sub>
S2	(pro <sub>1</sub> , NP <sub>2</sub> ) (pro <sub>2</sub> ) [pro <sub>1</sub> ] [pro <sub>1</sub> , pron <sub>2</sub> ] (Sub <sub>1</sub> , DO <sub>2</sub> ) (Sub <sub>2</sub> ) [Sub <sub>1</sub> ] [Sub <sub>1</sub> , DO <sub>2</sub> ]	3	5	3	4	Topic <sub>1</sub>
S3	[def unmod NP <sub>2</sub> ] (pro <sub>2</sub> ) [Sub <sub>2</sub> ] (Sub <sub>j2</sub> )	0	5	2	6	Topic <sub>2</sub>
S4	[pro <sub>2</sub> ] [Sub <sub>2</sub> ]	0	5	1	7	(Topic <sub>2</sub> )
S5	[def mod NP <sub>3</sub> , pron <sub>2</sub> , def unmod NP <sub>1</sub> ] [Sub <sub>3</sub> , attr <sub>2</sub> , IO <sub>1</sub> ]	1	6	1	8	(Topic <sub>3</sub> )

The first aspect investigated was the *referential persistence* of the story's referents. I counted how many times each referent was mentioned in the main and subordinate clauses of the continuation sentences. I also looked at how referential persistence relates to grammatical role and thus I calculated the referential persistence of the subject and the direct object of the story fragments given by me. The persistence of the subjects was

calculated as well, as I wanted to test how this value correlates with and influences the persistence value of the direct object. The occurrences within each sentence were added to a sum representing the referential persistence at a particular stage in the discourse (see Table 5.5 for an illustration). Additionally, the average was calculated by dividing the overall sum of all stories by the number of participants and the sentences (for *pe*-marked direct objects 11 participants and for unmarked direct objects 9 participants).

*Topic shift* was the second aspect coded. The first instance in which a direct object referent became the aboutness topic in a main clause was treated as an instance of topic shift (see Table 5.5, S3, last column for an example). Whether this change was maintained in the subsequent sentences or not, was of no relevance at this point. Important was which continuation sentence constituted the ‘turning point’ for the direct object to become the topic. Note that the discourse topic in the three stimuli for this experiment was established by introducing one character as the subject, and then referring to this character with a zero or pronoun in the following two sentences. Thus, the subject referent was the entity the stimulus item was about and was the prominent referent.

Topicalized direct objects (which also function as sentence topics) were disregarded since I had only two such instances. Potential topic shifts occurring in subordinate clauses were not taken into consideration because of the general assumption that subject referents mentioned in subordinate clauses are less prominent than subject referents introduced in main clauses (Cooreman and Sanford 1996, Kaiser 2000, Miltsakaki 2003). In a third step, the *type of anaphoric expression* used to refer back for the first time to the direct object referents was analysed. For each continuation sentence I coded whether the direct object (if present) was picked up as a zero (pro), clitic pronoun, proper name, or as a (modified or unmodified) definite noun phrase.

### 5.4.1.3 Results and discussion

23 participants provided continuations for the story fragments. Three continuations were excluded from the analysis because the participants added an ambiguous continuation sentence ( $n=2$ ) or wrote nothing at all ( $n=1$ ). This left a total of 20 continuations that could be analysed: 11 for the *pe*-condition and 9 for the unmarked condition.

### Number of anaphoric references – referential persistence

The first method for measuring the discourse prominence of direct objects was the number of anaphoric references. Figure 5.1 displays the mean values for the subject and object referents in both conditions up to the last continuation sentence (S5). The prediction concerning the referential persistence of the direct objects was confirmed, as the *pe*-marked referents were more persistent than the referents of the unmarked direct objects.

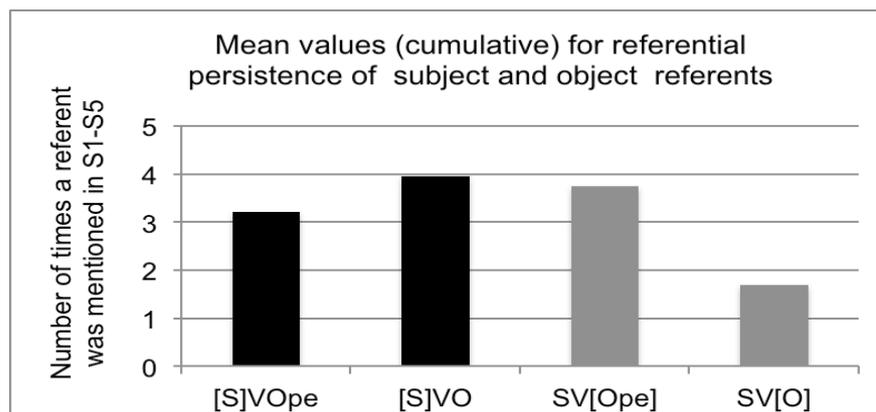


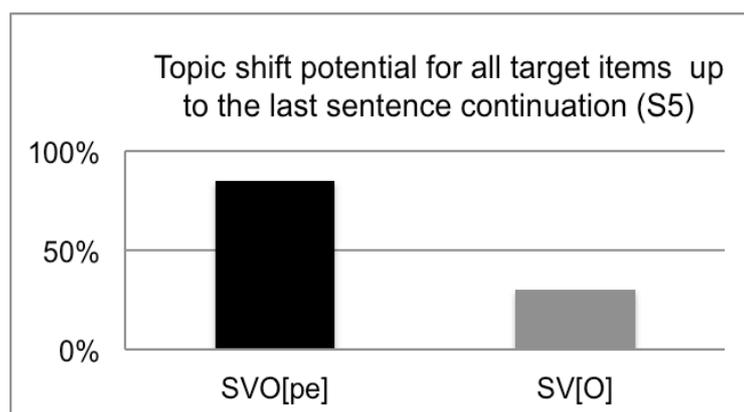
Figure 5.1: Referential persistence of subjects and object referents up to S5

In the non-*pe*-marked condition an overall strong tendency to continue elaborating upon the subject referent can be noticed. Figure 5.1 moreover indicates that for the *pe*-condition, the persistence of the referents of the case marked direct objects slightly exceeded the continuity of the subject referents. Different studies (Kertz 2006, Kehler et al. 2008) showed that depending on the coherence relation given in a sentence, the subject or the object referent of that sentence is associated with an increased tendency to appear in the following discourse. The somewhat unexpected observation that *pe*-marked referents are more referentially persistent than the subject referents of the same sentence, might be due to the coherence relations established between the sentences in which the target referents occurred. At this point, I will not pursue a more ample analysis of this option, but leave it for further research. In sum, participants preferred a continuation story that evolved around the referent of the subject, thus taking it up more often, unless the direct object referent was *pe*-marked.

## Topic shift

The second textual characteristic investigated was the topic shift potential of the indefinite noun phrases in direct object position. Recall that each mention of a direct object as an aboutness topic was counted as an instance of topic shift. The counts for the topic shift potential are cumulative.

The data in Figure 5.2 reveal several patterns. First, the referent of the *pe*-marked direct object displays a stronger preference to become the topic in the continuation sentences (S1-S5) than the referent of the non-*pe*-marked direct object referent. Second, while almost all participants mentioned the referent of the *pe*-marked direct object sooner or later as a topic in the continuations they provided, the unmarked direct object became a topic in less than 50% of cases.



5.2: Topic shift potential of *pe*-marked and non-*pe*-marked referents

Furthermore, Figure 5.3 shows that the referent of the unmarked direct object was never picked up in subject position in the first two continuation sentences (S1 and S2) provided by the participants. On the contrary, the referent of the *pe*-marked direct object was picked up in the first two continuation sentences, even though the rate was not high.

The findings concerning the topic shift potential of direct objects confirmed the initial predictions, as the referents of the *pe*-marked direct objects displayed a higher expectancy to be mentioned again as topics in a main clause in comparison to the unmarked ones.

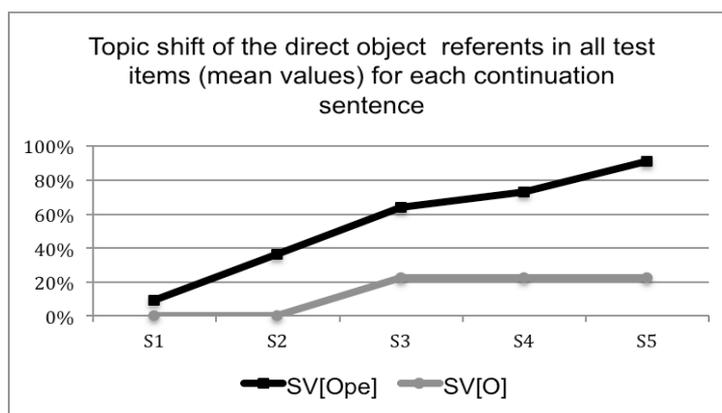


Figure 5.3: Results: Topic shift

In sum we can notice that indefinites headed by the simple indefinite article (i.e. the non-*pe*-marked indefinite) have a wider, less restricted distribution compared to their unmarked counterparts. That means that the simple indefinite forms are more prone to occur in both types of contexts: those in which their associated referents are less likely to be rementioned and to shift the topic, but also in those in which their associated referents are subsequently mentioned and shifting the topic. On the contrary, *pe*-marked indefinites are more specialized for such contexts in which their associated referents show high values for both referential persistence and topic shift. We thus find again favourable evidence for the generalization that the unmarked forms are not as restricted in their distribution compared to the marked forms (Givón 1981).

**Type of anaphoric referring expression**

The last textual characteristic coded was the type of referring expression used to refer back to the target referents for the first time in the continuing discourse. Tables 5.6 and 5.7 display the absolute and mean values of this characteristic for both conditions.

Table 5.6: The type of referring expression of the first anaphoric referent

<i>Pe</i> -marked direct objects in TI1, TI2, TI3- mean values/ items						
Ref. Type Sentence	Zero	CL	Pronoun	Def. unmodified NP	Def. modified NP	Sum
S1	3% (1)	9% (3)	(0)	27% (9)	9% (3)	48% (16)
S2	(0)	3% (1)	(0)	15% (5)	(0)	18% (6)
S3	(0)	(0)	(0)	6% (2)	15% (5)	21% (7)
S4	(0)	3% (1)	(0)	3% (1)	3% (1)	9% (3)
S5	(0)	(0)	(0)	(0)	(0)	(0)
Sum	3% (1)	15% (5)	(0)	52% (17)	27% (9)	97% (32)

Table 5.7: The type of referring expression of the first anaphoric referent

unmarked direct objects in TI1, TI2, TI 3- mean values/ items						
Ref. Type Sentence	Zero	CL	Pronoun	Def. unmodified NP	Def. modified NP	Sum
S1	(0)	4% (1)	(0)	7% (2)	19% (5)	26% (7)
S2	(0)	(0)	(0)	(0)	(0)	(0)
S3	(0)	(0)	(0)	11% (3)	19% (5)	33% (9)
S4	(0)	(0)	(0)	15% (4)	7% (2)	22% (6)
S5	(0)	(0)	(0)	4% (1)	11% (3)	15% (4)
Sum	(0)	4% (1)	(0)	37% (10)	56% (15)	96% (26)

The three test sentences of the *pe*-condition were continued by 11 participants, yielding 33 stories with five or more continuation sentences each, while for the non-*pe*-condition 27 continuation stories could be evaluated. Only one direct object referent in each condition was not mentioned at all in the following sentences, as the two sums show: 32 instances of 33 continuations (97%) for *pe*-marked direct objects and 26 instances of 27 continuations (96%) for unmarked direct objects. Furthermore, only one participant picked up the *pe*-marked direct object in the first continuation sentence with a zero anaphor, which yields the 3% in the first cell of Table 5.6. The sum 16 (representing 48%) of the first line in Table 5.6 shows that almost half of the participants took up the *pe*-marked direct object in the first sentence (S1). The sum 17 (representing 52% in Table 5.6) of the column for definite unmodified noun phrases shows that more than half of the participants have referred back to the *pe*-marked direct object for the first time with an unmodified definite noun phrase. Several other observations can be made on the basis of the two Tables above. First, Tables 5.6 and 5.7 indicate that the *pe*-marked referents are mentioned for the first time earlier in the continuation sentences compared to the referents of the unmarked direct objects (48% (16) vs. 26% (7) in S1). In contrast to that, the majority of the non *pe*-marked referents are mentioned for the first time in S3 (33%) and the probability that the referent will be mentioned in S4 and S5 for the first time is higher than for the non-*pe*-marked one. Second, more reduced types of referring expression are used to refer for the first time to the *pe*-marked objects. There are more clitic anaphoric expressions for *pe* than for non-*pe*, and zero anaphors are never used to take up the referents of the non *pe*-marked referents. Yet, the absolute numbers are low in both cases and the difference is not very clear, due to the small numbers. Third, an overwhelmingly high number of full definite noun phrases were found for both direct object types. This observation corresponds to the low accessibility of direct objects compared to the subjects. A closer inspection reveals that unmodified noun

phrases prevail for the *pe*-marked referents, whereas in the non-*pe* marked instance modified noun phrases prevail. A certain tendency towards the initial prediction can be observed, namely that *pe*-marked direct objects are associated with a higher level of activation and thus the anaphoric link to them needs less descriptive material. However, more less specified types of referring expressions for the *pe*-marked referents were expected.

Overall, the results of the analysis of the anaphoric forms used to refer back to the direct object referents have two main implications. First, the referents of the marked as well as the unmarked direct objects are taken up for the first time in the continuation sentences more often by means of a definite (unmodified) noun phrase than by means of any other type of referring expression. Different experiments (e.g. Arnold 1999, Francik 1985, Karmiloff-Smith 1985, Givón 1983) have demonstrated that speakers use fewer personal pronouns when there are more salient individuals that match the gender of the pronoun to avoid ambiguities in reference. Note that the stimuli sentences were constructed in such a way that the two introduced referents (i.e. the subject and the direct object referent) were of different genders. However, even if they could have opted for more reduced forms, participants chose more specified forms to refer again to the direct object referents. This finding supports Arnold and Griffin's (2007) claim that not only gender, but also the presence of multiple referents in the discourse reduces the use of less specified types of referring expressions to refer back to one of them. In our case, the first-mentioned subject noun phrase, which is the clearly established topic constituent, inhibits the potential to systematically take up the direct object referent by a more reduced form in the first continuation sentence. Second, the results showed only minor effects of the direct object type upon the referring expression type used to refer back to it. In other words, reduced forms (clitics) as well as more descriptive forms (definite unmodified and modified noun phrases) were used to refer to both: the referent of the *pe*-marked direct object as well as to the referent of the unmarked direct object. Such results point towards an explanation in which the likelihood of a referent to be mentioned again and its probability to be mentioned with a particular type of referring expression in the subsequent discourse should be kept apart. Indeed, Kehler et al. (2008) accounted for Stevenson et al.'s (1994) seemingly contradictory experimental results by providing an explanation which differentiates between two types of expectations, namely an expectation towards the subsequent mention of a referent and an expectation concerned with the form of referring

expression that a speaker would use to take up that referent again. Kaiser (2010) observed a similar distinction that pertains to the likelihood of pronominalization on the one hand and the likelihood of subsequent mention on the other hand.

## 5.5 Summary

In this Chapter I discussed the discourse structuring potential of *pe*-marked indefinite direct objects in Romanian. As *pe*-marking corresponds to DOM in Romanian, I firstly motivated the need to depart from single-sentence analyses to account for its distribution with indefinite noun phrases. A body of research on DOM notes that *pe* with indefinite noun phrases constitutes an instance of a fluid domain, arguing that in this case; specificity is the triggering factor licensing the use of the marker *pe*. Second, I showed that indefinite noun phrases in direct object position allow for the realization of referential properties such as specificity and wide scope readings. However, I showed that specificity is not sufficient to account for *pe*-marking with indefinite noun phrases in Romanian. More concretely, I argued that *pe*-marked direct objects are also sensitive to the discourse context they appear in, changing the discourse properties of their referents. Based on the findings of an off-line story continuation study, I showed that *pe*-marked referents are devices for structuring the discourse that are interpreted by hearers as signals of high referential continuity. At the same time, *pe*-marked referents are more prone to shift the topic of the discourse they appear in. These observations parallel the experimental findings discussed in Chapters 3 and 4, as the indefinites headed by indefinite *this* in English and indefinite *so'n* in German were shown to be referentially continuous and were more likely to shift the discourse topic. Thus, the necessity to distinguish between the expectancy of referential persistence and topic shift potential, on the one hand, and the type of the first anaphoric expression, on the other hand, is once again underlined by the investigation presented in this Chapter. It was argued throughout this dissertation that the discourse structuring potential of the three different indefinite types is mirrored by the first two factors, but not by the latter.

The indefinite markers discussed in Chapters 3, 4 and 5 show striking similarities with respect to their strong tendency for referential meanings and discourse prominence. The next Chapter will draw upon data from other languages to show that these two characteristics of a referent are realized by different morphological means cross-linguistically.

## 6. Referential properties and discourse structuring

In the previous Chapters, I presented the characteristics, referential properties and discourse structuring function of noun phrases headed by indefinite *this* in English, *so'n* in German and *pe*-marking in Romanian, as illustrated in (1a-c)

(1) (a) He saved this girl from drowning.

(b) Er hat so'n Mädchen vom Ertrinken gerettet.  
He Aux. so-a girl from drowning saved  
'He saved so-a girl from drowning.'

(c) El a salvat -o pe o fată de la înec.  
He Aux. saved CL PE a girl from drowning  
'He saved PE a girl from drowning.'

Despite their different functions at sentence level, I showed that these types of referring expressions share several characteristics. First, these noun phrases are indefinite, as they can occur as pivots in several constructions that allow only for indefinite expressions (e.g. existential *there*-sentences, the combination with warmth-nominals, etc.), and because they introduce discourse-new and hearer-new referents. Second, in comparison to the least marked indefinite noun phrases headed by the simple indefinite article, the indefinite markers analysed in this dissertation have a limited distribution and allow for more subtle interpretations. More precisely, I showed that at sentence level they display a strong and robust tendency for referential properties, i.e. rigid reference, wide scope relative to other operators, fixed reference, etc. I argued for each of the three indefinite expressions that they display the semantics of indefinites and have an additional constraint that leads to their referential stability (in the sense of Brasoveanu and Farkas 2010). Generally, whether a particular determiner is referentially stable or variable, depends on different constraints. According to certain accounts, definite noun phrases achieve referential stability via uniqueness (Russel 1905), while partitives achieve it through familiarity (Roberts 2003), and so on. Throughout this dissertation, I argued that noun phrases headed by English indefinite *this*, German *so'n* and Romanian *pe* achieve referential stability by means of

their discourse structuring potential. Based on several studies on accessibility, topicality and salience that were discussed in Chapter 2, the notion of discourse structuring potential was introduced as an umbrella term for two textual characteristics of a referent that pertain to the following discourse, namely (i) referential persistence and (ii) the potential to change the current topic of the discourse in a next matrix clause. In light of the findings of three web-based story continuation experiments, I showed that referents headed by the three markers are highly referential persistent and show a strong tendency of becoming topics in the upcoming discourse. The evidence from the third measure method investigated, namely the first anaphoric expression used to pick up for the first time the referents associated with the indefinite noun phrases, revealed no difference between the marked and unmarked indefinites. Overall, I argued that despite indicating a high discourse structuring potential, the marked indefinite noun phrases do not signal the accessibility of their associated referents. Moreover, it seems that any attempt to characterize these markers, without investigating their behaviour in larger discourse units, cannot thoroughly account for their distribution and function.

The aims of this Chapter are manifold. First, I summarize and compare the theoretical and empirical claims put forth in the last three Chapters. Based on these results, I will elaborate on the notion of discourse structuring potential and its relevance for the better understanding of the behaviour of indefinite noun phrases headed by indefinite *this*, *so 'n* and *pe*. Another aspect that will be elaborated upon in this Chapter is the different morphological means employed cross-linguistically to indicate discourse structuring. It was argued so far that Romanian uses the differential object marker *pe* to fulfil this function, while German and English use the demonstrative for properties, *so 'n* and the demonstrative for objects *this*, respectively for the same function. Based on the diachronic development of differential object marking in Romanian (Chiriacescu 2007), I will argue that at least since the beginning of the 19<sup>th</sup> century, *pe*-marking in Romanian was used to mark discourse prominent referents, a function that is still observable in synchronic Romanian in such contexts that allow for optionality (i.e. with direct objects realized as definite and indefinite noun phrases). Moreover, different studies on DOM cross-linguistically (Niculescu 1959, Lace 1987, Dalrymple and Nikoleava 2011) have suggested that DOM markers entered the language to indicate differences in information structure. I will suggest that the primary function of DOM in Romanian was and is to distinguish between referents that will become more or less prominent in the subsequent discourse. Languages that do

not have DOM, as English and German develop other means (e.g. demonstrative determiners) to realize this function.

At the end of this Chapter, I attempt to answer the question why English, German and Romanian need indefinite noun phrases that give structure to the subsequent discourse? I will show that the languages under investigation make a distinction in their determiner system between definite and indefinite, but do not distinguish between discourse prominent and discourse non-prominent referents. I argue that the need of languages to distinguish between referents that are more likely to be referentially persistent in the ensuing discourse and such referents that are neutral or underspecified towards this prediction is an essential distinction that languages employ for discourse structuring purposes.

This Chapter is organized as follows. In Section 6.1 I will briefly compare the referential properties of the three indefinites describes in this study. Section 6.2 continues with the comparison of the experimental findings presented in Chapter 3, 4 and 5. More precisely, I show that indefinite *this*, *so'n* and *pe* and are used as discourse structuring devices. Furthermore, when comparing the experimental data, it can be observed that the three indefinite noun phrases differ with respect to their strength in giving structure to the subsequent discourse. Section 6.3 discusses the relation between referential properties and discourse prominence. The final Section summarizes the key observations made in this Chapter.

## 6.1 Referential properties

The investigation of the behaviour of indefinite noun phrases in opaque contexts showed that indefinites headed by English indefinite *this*, German *so'n* and *pe* in Romanian resemble the behaviour of simple indefinite noun phrases headed by *a(n)*, as they pass several traditional indefiniteness tests. Nevertheless, we cannot account for the distribution of the special indefinites central to the present study on a par with the distribution of *a(n)*, as the two determiners were shown to differ with respect to their referential properties at sentence level. In contexts that involve sentence operators and in contexts that lack them, noun phrases headed by the simple indefinite article take scope freely, while noun phrases preceded by the indefinite markers discussed here were shown to display a strong and

robust tendency for referential readings (i.e. specific, wide, non covarying endings). Table 6.1 summarizes the readings of the three indefinite noun phrases in different contexts.

Table 6.1: Referential properties of indefinite noun phrases headed by *this*, *so 'n* and *pe*

	referential	island escaping	independent	interme- diate	functional	pair- listing	epistemic
<i>this</i>	specific	wide>narrow	no covariation	no	no	no	specific
<i>so 'n</i>	specific	wide>narrow	no covariation	no	no	no	specific/ non-specific
<i>pe</i>	specific	wide>narrow	no covariation	no	no	no	specific/ non-specific

Several observations can be made in light of the data reported in Table 36. First, in contexts involving existential quantifiers, simple indefinite noun phrases are most commonly analysed as being ambiguous between a specific and a non-specific reading. On the contrary, the indefinite noun phrases headed by *this*, *so 'n* and *pe* are compatible only with an interpretation in which their referents do not covary with the referents of the universal quantifier (i.e. the specific reading). Further evidence for the scopally specific interpretation of these indefinite noun phrases comes from their behaviour under scope islands. Unlike simple indefinite noun phrases, which escape scope islands such as conditionals and which allow intermediate scope above the island, but below a higher quantifier (Farkas 1981, Ruys 1992, Abusch 1994), the indefinite noun phrases central to this dissertation can be interpreted only outside the scope island and thus taking the wide scope reading. Second, in addition to being scopally specific, the indefinite noun phrases central to the present analysis were shown to be sensitive towards referential specificity as well, which means that they favour the existential reading of their associated referents in certain contexts. Finally, the three markers discussed here differ with respect with their behaviour in neutral contexts (i.e. in contexts that do not contain operators at sentence level). More precisely, while noun phrases headed by English indefinite *this* allow only for an epistemic specific reading in such contexts, it seems that German *so 'n* and Romanian *pe*-marking allow for both the epistemic specific and non-specific reading.

An interesting and important observation worth to be pointed out is that the referents of the indefinite noun phrases in the three languages represented in Table 6.1 differ with respect

to their referential strength. That means, that some of the indefinite noun phrases discussed in this dissertation show a tendency for referential readings more readily than others. The different contexts in which the behaviour the referents of the indefinite noun phrases was tested show that English indefinite *this* is specific (referentially, scopally and epistemically), whereas German *so'n* and Romanian *pe* seem to allow for non-specific readings of their referents as well.

Overall, however, because the three types of indefinite markers discussed in this dissertation are sensitive with respect to each of the three notions of specificity represented in Table 4.1 (i.e. referential specificity, scopal specificity and epistemic specificity), many researchers concluded that their import is to show specificity. However, the fact that specificity (and its different instantiations) is not sufficient to account for the distribution and function of these indefinite noun phrases becomes evident at discourse level, in contexts of more than one sentence. In larger discourse units, the referential readings of these indefinite noun phrases can be neutralized and their referents seem to achieve referential stability by means of another mechanism, namely their discourse prominence in terms of discourse structuring potential. In the next Section, I briefly discuss the findings from the three sentence-continuation experiments, which sustain these observations.

## 6.2 The discourse structuring potential

To test whether indefinite noun phrases headed by English indefinite *this*, German indefinite *so'n* and Romanian *pe* differ with respect to their discourse contribution from their simple indefinite counterparts, I investigated their behaviour in contexts larger than two adjacent sentences. As the interpretation of indefinite noun phrases generally does not rely on an antecedent expression, I employed a forward-looking perspective to determine the discourse effects of indefinite noun phrases. In doing so, I explored three characteristics of referents associated to indefinite noun phrases that pertain to the following discourse, namely: (i) referential persistence (i.e. the likelihood of a referent to be picked up in the following discourse several times), (ii) topic shift potential (i.e. the tendency of a referent to become the topic in the subsequent discourse), and (iii) the type of the first anaphoric expression used to pick up the referents of the indefinite noun phrases in the ensuing discourse. Recall that the discourse structuring potential of the referents of the indefinite

noun phrases was tested relative to the discourse behaviour of indefinite noun phrases headed by the simple indefinite article. The major findings of the three experiments on the discourse behaviour of referents headed by indefinite *this*, *so 'n* and *pe* are summarized in the following.

The first textual characteristic indicating discourse prominence in terms of discourse structuring potential was *referential persistence* or continuity. This property underlines the likelihood of the referents associated to the marked indefinite noun phrases to be frequently picked up in the subsequent discourse compared to their simple indefinite counterparts. Despite being realized in a non-prominent position (i.e. as indefinite noun phrases in direct object position), the referents of the marked indefinite noun phrases, which were mentioned in grammatical direct object position, became better competitors for the referents realized in subject position than the referents of the direct objects realized as indefinite noun phrases headed by the simple indefinite article. In other words, the marked indefinite noun phrases were shown to boost the discourse structuring potential of their associated referents. This observation does not hold for the simple indefinite noun phrases, which did not show high rates of subsequent mention.

Comparing the results of the three sentence-continuation experiments brings up interesting questions about the relation between the sentence's referents in terms of referential persistence. A closer inspection of the next mention bias of the sentence's referents reveals several interesting patterns. First, recall that most (psycho-)linguistic investigations compute the probability of subsequent mention of a referent locally, that is, between two adjacent sentences. For example, in most instantiations of Centring Theory, the referent realized in subject position in the current sentence is the preferred referent to be picked up in the immediately following sentence. This tendency was confirmed by the present investigation as well. More concretely, the subject referent on the test items showed higher rates of immediate subsequent mention compared to the direct object referents. The presence of the indefinite markers did not reverse this pattern. Second, besides investigating local (i.e. between two adjacent sentences) next mention biases, the present studies explored the frequency of subsequent mention as well. Specifically, the number of times a referent was picked up in the next five following sentences was determined as well. Surprisingly, the results from this investigation showed that in larger discourse segments, marked indefinite noun phrases not only exceeded their unmarked counterparts in

frequency of subsequent mention (i.e. referential persistence), but even exceeded their subjects in this characteristic. This observation is particularly true for the *pe*-marked direct objects in Romanian and the referents of German indefinite *so 'n*. However, the presence of the indefinite marker *this* in English does not impact the referential persistence of its associated referents that strongly. That means that independently of the special marking, English direct objects seem to be more referential persistent, regardless of the presence of indefinite *this* compared to German and Romanian direct objects. These results indicate that the indefinite markers boost the referential persistence of their referents, but not in an equal manner. Referents associated with *so 'n* and *pe* indefinite noun phrases become better competitors for the subject referents than the direct object referents preceded by English indefinite *this*. Furthermore, the generalisation that the referents preceded by the three markers are always more persistent in the subsequent discourse is not entirely true. There are certain contexts in which the referents' persistence is dampened, despite the presence of the marker. The generalization deriving from this observation is that in a generalized event, in which a particular referent does not play a pivotal role, the presence of a marker like German *so 'n*, for example, will have a weaker impact upon its referents' persistence, than in a context in which the referent plays an important role in the subevents of the event schema initially introduced. Further investigations are needed to determine the factors that contribute to the reversal of the referential persistence pattern.

The second characteristic used to determine the discourse prominence of the indefinite noun phrases in terms of discourse structuring potential was *topic shift*. I focused on the way in which different types of indefinite noun phrases affect topic transitions between two adjacent sentences and in larger discourse segments. Recall that I used the term 'topic' as an information structural device, which stands for the entity the current sentence makes a predication (Strawson 1964, Reinhart 1982). According to this view, the topic position is distinct from the grammatical subject position, despite the general tendency of topics to be mentioned in grammatical subject position (Hockett 1958). Thus, the topic shift potential was measured in terms of the likelihood that a referent becomes the aboutness topic in a matrix clause in the subsequent discourse. Before we compare the results for topic shift potential, recall that previous work showed that next mention biases are generally modulated by accessibility and that indefinite noun phrases correspond to lower rates of immediate subsequent mention compared to their definite counterparts (see the discussions in Chapter 2). In light of these observations, the prediction for the three languages was that:

(i) definiteness and topic position affect local (i.e. sentence-by-sentence) topic transitions, and that (ii) indefinite noun phrases have (at best) an impact upon non-local topic transitions (i.e. non-adjacent matrix sentences). In other words, topic referents, which were realized in prominent positions (i.e. as subjects, introduced by proper names), were expected to show local effects on topic continuation, while the non-prominent indefinite direct objects were expected to show an impact on a non-local basis.

The overall findings of the experiments indicate that both the initial topic position and the type of indefinite noun phrase used affect language production in terms of topic planning. First, the results support previous findings on the preference of topic continuations between two adjacent sentences, as the initial topic was picked up as an aboutness topic in the first sentence continuation in more than 50% of the cases across conditions. Second, overall effects in terms of subsequent topic shift occurred up to the third sentence continuation, which means that less topic shifts occurred after this point. Third, it seems that the type of indefinite noun phrase used is a means to indicate non-local (between non-adjacent sentences) topic shifts. Specifically, the results suggest that marked indefinite noun phrases were more prone to be mentioned in topic position, while simple indefinite noun phrases were picked up in topic position only marginally. However, marked indefinite noun phrases showed non-local effects on topic transitions, as their associated referents were picked up in this position with a delay of one sentence. These findings challenge traditional accounts on topic continuity as a means to indicate accessibility or salience. Such accounts did either not consider the discourse effects of less-prominent referents, or computed accessibility by looking at a referent's status in the immediately preceding sentence (e.g. Ariel 2001, most instantiations of Centering Theory).

Fourth, the findings reveal that the referents associated with the three indefinite noun phrases do not display the same strength in terms of topic shift. Direct objects are more prone to become the next topic in English, regardless of the presence or absence of indefinite *this*, whereas for German and Romanian, the presence of *so'n* or *pe* enhances the topic shift potential of the indefinite noun phrases in direct object position. Whether these observations show that English direct objects generally differ from direct objects in Romanian and German, is left for further research. However, if this observation is true, then it would imply that the correlation between topic(hood) and grammatical subject position is less strong in English, compared to Romanian and German. Both Romanian and

German are more topic-prominent than English. One reason for this correlation might be due to English word order being more rigid than German or Romanian word order. Thus, when a referent is mentioned in subject position (or perhaps in sentence initial position) in English may not be as informative as it is in a language like Romanian or German that has flexible word order. Several studies on this topic hinted at this difference (Hemforth 2000), but a suitable evaluation of the tenability of this suggestion awaits future research.

The findings from the third textual characteristic investigated show that the referents of the three indefinite noun phrases under investigation are not likely to be pronominalized in the subsequent discourse. These results confirm other recent findings that showed that less prominent referents are less likely to be immediately picked up in the subsequent discourse by less explicit types of referring expressions. Rather, the referents associated with the indefinite noun phrases headed indefinite *this*, *so 'n* and *pe* were taken up in the following discourse by means of definite (modified) noun phrases, similar to their simple indefinite counterparts. These findings show that the two types of indefinite noun phrases (the marked indefinite noun phrases vs. the unmarked ones) do not have a different impact upon the type of referring expression chosen to take up their referents again.

In sum, the results from the three sentence-continuation studies presented in this dissertation show that from a production perspective, the three indefinite markers are indicators of the discourse structuring potential of their associated referents. Discourse structuring potential is understood in terms of the referential intention of the speaker to signal to his addressee that further information about the referents associated to these indefinite noun phrases will follow. Crucially, the findings of the studies presented in the preceding Chapters showed that indefinite noun phrases headed by the special markers vary with respect to the strength of their discourse structuring potential, as summarized in Table 6.2 below.

Table 6.2: The discourse structuring potential of indefinite noun phrases

	<i>This</i>	<i>So'n</i>	<i>Pe</i>	<i>Simple indefinite artic</i>
<i>Referential persistence</i>	medium	high/medium	high	underspecified
<i>Topic shift</i>	medium	high	high	underspecified
<i>Type of anaphoric expression</i>	definite NP	definite NP	definite NP	definite NP

Table 6.2 shows that Romanian *pe*-marking displays the highest values for both referential persistence and topic shift potential, whereas English indefinite *this* shows the lowest values for these two characteristics. Overall, indefinite noun phrases headed by the simple indefinite article are underspecified for referential persistence and topic shift potential. That means, the referents associated with this type of indefinite expression are less likely, but not excluded, from being referentially persistent and becoming the topic in the following discourse. This observation provides further evidence for the generalization that unmarked forms (i.e. simple indefinites) are less restricted in their distribution compared to marked forms (Givón 1981).

Summarizing, the findings presented here show that we have to distinguish between two main types of expectations, which are triggered by different factors at different discourse levels. The first type of expectation, which is generally discussed in the literature on pronoun resolution in terms of accessibility, can be derived from structural factors such as a particular syntactic position and a certain argument structure. Based on such factors, hearers make probabilistic expectations on a sentence-by-sentence basis about those referents that are more probable to be mentioned next and more prone to be pronominalized. The second type of expectation, which is important for the present analysis, comes about by discourse structuring devices such as the three types of indefinite noun phrases discussed in this dissertation. Such linguistic devices target larger discourse units in that they indicate which referent will be mentioned over larger discourse segments. It is important to note that the discourse structuring potential is not to be confounded with accessibility or activation, as different types of referring expressions that are associated with a high discourse structuring potential are not necessarily more activated or accessible compared to referring expressions associated that are not associated with a discourse structuring potential. Moreover, I argue that the expectancy of continuation (i.e. the

forward pointing potential) is a property defining indefinite noun phrases, but that at the same time it is a fine-grained feature that differentiates between several sub-types of such expressions. In other words, a referent mentioned for the first time by means of an indefinite noun phrase is more or less prone to be mentioned again in the following discourse and to shift the topic. We furthermore observed that the three languages under investigation do not differ with respect to the expectancy types they employ, as both sentence-by-sentence effects and discourse structuring effects are realized. What is different is the formal marking employed by each language for each function and the relative weighting that languages assign to each of them. The main suggestion resulting from the presented evidence is that expectancy operates in parallel at different levels of representation.

In light of these findings, one general question that arises is whether the discourse structuring potential of the referents headed by the three determiners discussed in this dissertation can be calculated from a speaker-oriented or an addressee-oriented perspective. That is, do speakers use the three indefinite markers for the benefit of the hearer, or rather egocentrically, for their own purposes? The literature on the topic has been providing mixed evidence on the interplay between these two views in reference processing. On the one hand, there has been abundant evidence that the production mechanism has access to estimations about the addressee's expectations, as speakers make detailed assumptions about common ground between themselves and their addressees, adapt their word pronunciation based on their relationship with the addressee, use linguistically less complex types of referring expressions when a particular entity is known by the addressee (e.g. Chafe 1976; Gundel et al. 1993, Prince 1981, Grosz et al. 1995, Clark 1996, Eckert 2004, Arnold 2007). On the other hand, different studies have shown that sometimes it is difficult to disentangle speaker-internal and addressee-oriented production strategies, as any shared information is available to the speaker alone and could thus be used "selfishly" for his own information organisation purposes only (Francik 1985, among others). For example, it was shown that speakers do not always avoid ambiguity with *that*-complementizers (Ferreira 2003), or that the rate of pronominal reference can be low for a highly salient referent, if two semantically human referents (of different gender) are in the context (Arnold and Griffin 2007). Even though the present studies did not explicitly address the question of whether the choice of using indefinite *this*, *so'n* and *pe* over their simple indefinite counterparts is a speaker-internal or an addressee-oriented phenomenon,

the findings from the three experimental studies point towards an explanation in terms of audience-tailored discourse structuring in which speakers take into consideration what information is important for their addressees to access. This view is consistent with Clark and Marshall's (1981) and Clark and Brennan's (1991) suggestion that for felicitous reference and discourse planning, speakers and comprehenders must establish certain kinds of mutual knowledge by updating their *common ground* (Lewis 1969, Karttunen 1976) on a moment-by-moment basis. The presence of *this*, *so'n* and *pe* on indefinite noun phrase serves as a clue for the addressee that a referent marked in this way has a high potential to give structure to the subsequent discourse. That is, upon encountering the special indefinites, the addressee develops expectations about the high referential persistence and the topic shift potential of the referents associated with these types of referring expressions. That these expectations are borne out was shown on the basis of the experimental findings discussed above. Recall that the multi-sentence story-continuation tasks used to test the discourse structuring potential of referents marked by indefinite *this*, indefinite *so'n* and *pe* has the advantage of combining comprehension and production processes. That means that participants' responses were made on the basis of the mental representations and the expectations they generated while reading the stories.

Another aspect that speaks for an addressee-oriented interpretation of the data at hand comes from the seemingly optional distribution of *this*, *so'n* and *pe*-marking with simple indefinite noun phrases. At any time in the discourse, a speaker could, in principle, opt for the simple indefinite forms instead of using the marked ones. The Griceian maxim of quantity (Grice 1975) states that speakers tend to use the least complex type of referring expression that is informative enough to fulfil their communicative purposes. For both speakers and hearers, processing a linguistically complex and marked type of referring expression results in a higher processing cost than processing a linguistically less complex, unmarked type of referring expression (Gordon, Grosz and Gilliom 1993). This implies that the processing cost of a referring expression must be justified in terms of the contribution it serves within a particular context by identifying the referent, adding new information, contributing to the discourse structure, or by other means (Almor 1999). Thus, by choosing the more complex indefinite noun phrase form over the simpler indefinite form, speakers are likely to: (i) establish a new referent in the common ground, and (ii) signal that the newly introduced referent will give structure to the upcoming discourse.

Summing up the major theoretical claims made with respect to the interpretation and discourse behaviour of the three indefinite markers in English, German and Romanian, it can be stated that these types of referring expressions mark referents, which are referentially stable and which give structure to the upcoming discourse in terms of referential persistence and topic shift potential. Furthermore, I showed that both (i) the referential properties and (ii) the discourse properties of the referents differ in strength cross-linguistically. Indefinite noun phrases headed by *this* in English show a stronger tendency for referential readings, but a lower discourse structuring potential compared to Romanian *pe*-marking and German *so'n*, which show reversed preferences.

Before I discuss in Section 6.3.3 the observation that many languages indicate referents associated with the two characteristics mentioned above by different morphological markers, I will address a more basic question, which pertains to the relation between the referential properties (i.e. the various instantiations of specificity) and the discourse prominence in terms of discourse structuring of a particular referent. In other words, are the referents introduced by indefinite *this*, *so'n* and *pe* discourse prominent in terms of discourse structuring potential due to their referential properties, or vice versa?

### **6.3 The relation between referential properties and discourse structuring**

The relation and connection between the referential properties and the discourse status occupied by particular referents within a discourse was largely left unaccounted in previous investigations. Ionin (2006), for example, analyses English indefinite *this* and considers that it indicates the referential intention and familiarity of the speaker with a referent, but not that of the hearer. The speaker's intention to refer is responsible for the referential stability of the referents headed by *this*. Ionin furthermore adds a felicity condition to the use of indefinite *this*, which is tight to the noteworthiness of the referent associated with the indefinite expression<sup>33</sup>. In other words and according to her analysis, indefinite *this* can precede a referent only if the speaker intends to refer to exactly one entity about which he commits himself to contribute a noteworthy property, as illustrated in (2). Despite

---

<sup>33</sup> Recall from Chapters 3 and 5 that Ionin's (2006) notion of noteworthiness comes close to, but is not to be confounded with the textual characteristic of referential persistence used in this dissertation.

formalizing Prince's (1981) investigation of indefinite *this* in English, Ionin (2006) does not comment upon the relation between the referential and discourse properties of this determiner.

(2) Ionin's analysis of English indefinite *this* (2006: 187)

A sentence of the form  $[sp \alpha] \phi$  expresses a proposition only in those utterance contexts  $c$  where the following felicity condition is fulfilled: the speaker of  $c$  intends to refer to exactly one individual  $x_c$  in  $c$ , and there exists a property  $u$  which the speaker considers noteworthy in  $c$ , and  $x_c$  is both  $\alpha$  and  $u$  in  $c$ . When this condition is fulfilled,  $[sp \alpha] \phi$  expresses that proposition which is true at an index  $i$  if  $x_c$  is  $\phi$  at  $i$  and false otherwise.

Von Heusinger (2011b) investigates the German proximal indefinite demonstratives *so'n* and *dieser* and shows that they share similar referential and discourse properties to English indefinite *this*. First, adopting a Kaplanian semantics (1977, 1989) of demonstratives, he assumes that these indefinite markers can raise the attention of the hearer towards the following discourse, in addition to the visible discourse situation or previous text. Second, the *speaker's intention to refer* is regarded as central to the interpretation of indefinite demonstratives. Third, the referential properties of the indefinites headed by *dieser* and *so'n* are accounted for in terms of specificity as referential anchoring (von Heusinger 2002). If indefinite demonstratives bear a specific reading, as in the case of the German indefinite noun phrase central to his analysis, then the speaker's referential intention is represented as an anchoring function between the speaker as an attitude holder and the intended referent. While the anchor (i.e. the speaker) is speaker-given and hearer-given, the content of the anchoring function (i.e. the intended referent) is hearer-new. Thus, upon encountering the indefinite noun phrases headed by the two German demonstrative indefinites, the hearer will establish a permanent discourse representation for an individual referent in the case of *dieser* and for the property, and therefore also for the individual referent in the case of *so'n*, as presented in (3).

(3) Readings for German indefinite *dies* and *so'n* (von Heusinger 2011b: 25)

- (a)  $[dies_{indef} N]$  = is defined only if there is an **anchoring function from the speaker** to an object such that the object is  $N$ .

- (b)  $[so'n_{indef} N] =$  is defined only if there is an **anchoring function from the speaker to a property P** such that the referent is N and that there is a referent x such that x is N and P.

Fourth, the discourse contribution of noun phrases headed by indefinite *dies* and indefinite *so'n* can be derived on the basis of their semantics summarized in (3). Summing up, this analysis offers an explanation for the semantic and discourse properties of indefinite demonstratives, assuming that the discourse contribution of the referents they are associated with is secondary to their referential properties.

In this dissertation I adopt a functionalist view on referring, which is related to that of (Givón 1981, 1984) and assume that the discourse prominence (in terms of discourse structuring potential) associated with the referents headed by English indefinite *this*, German *so'n* and Romanian *pe* is the feature that essentially characterises these referents, rather than their referential properties. While the notion of semantic specificity (or referentiality, in Givón's 1984 terminology) is necessary to describe the effects of the referents headed by the aforementioned markers in most contexts at sentence level, I argue that this notion is not sufficient to account for the whole range of data at hand. More precisely, I consider the referential properties of these indefinite markers to be derived from their discourse properties. Throughout this dissertation I provided empirical evidence that echoes this point. In the following, I briefly mention some examples that were introduced in the previous Chapters and which underline that, on the one hand, semantic specificity strongly correlates with discourse prominence, but that, on the other hand, the first is not a prerequisite for the latter.

First, recall that noun phrases headed by the three markers discussed in this dissertation do not always bear a (referential, scopal or epistemic) specific interpretation, as shown in example (4). If we assume *pe*-marking to signal (epistemic) specificity in this context, we would expect that the speaker refers to an individual the identity of whom is known to him. However, such a reading does not hold, as the sentence can be continued with a comment that inhibits the referential anchoring of the referent of the indefinite to the speaker. Despite being preceded by *pe*, the indefinite noun phrase *o asistentă* ('a nurse') does not indicate epistemic specificity. It was argued in Chapter 5 that in example (4), the speaker expresses his referential intention to introduce an entity that *will* additionally play a

prominent role in the subsequent discourse. The speaker's referential intention is mirrored by the referential persistence of the referent introduced in this way in the discourse.

- (4) Rareş a văzut-o pe o asistentă ieşind din cabinet, noi nu ştim cine este ea. Important este că ne putem furişa înăuntru acum.

'Rareş saw PE a nurse leaving the practice, we don't know who she is. What matters is that we can sneak inside now.'

Second, it was shown that some sentences containing sentence-level operators allow for non-specific interpretations of the marked indefinites. Consider example (5), in which the referent of the modified noun phrase *so'nen Prinz auf einen weißen Ross* ('such a prince on a white horse') was expected to take an existential reading in this context, due to the presence of *so'n* heading the noun phrase. This expectation remains not fulfilled in the context given below, as the existence of the referent is denied by the immediately following context.

- (5) Carrie will *so'nen/ einen Prinz auf einen weißen Ross* heiraten. Aber sie hat noch keinen kennengelernt.

'Mary wants to marry so-a/ a prince on a white horse. But she has not met one yet.'

Third, an account that assumes the discourse properties of a referent to be secondary to their referential properties has to accommodate such cases in which the referent of a noun phrase preceded by the simple indefinite article displays similar discourse properties to those of the referent preceded by the markers *pe*, *so'n* and *this*. That these are cases, which should be accounted for as well becomes evident in light of the experimental findings reported in Chapters 3, 4 and 5. The experimental evidence underlined the fact that participants chose to pick up referents introduced by the simple indefinite article as well. Furthermore, in some cases, referents associated to simple indefinites were even more likely to become the topic constituent in the ensuing discourse, compared to their marked counterparts. As mentioned earlier in this dissertation, I account for such seemingly surprising results by assuming that the indefinite article is underspecified for discourse prominence in terms of discourse structuring. In other words, I assume that English, German and Romanian use *this*, *so'n* and *pe* respectively, to indicate that the referents associated with the noun phrases they head will be discourse prominent, whereas the

simple indefinite article can also allow for discourse prominent referents in marginal cases. Summing up, an account, which derives the discourse properties of referents headed by indefinite *this*, *so'n* and *pe* from their semantic properties can neither account for cases in which non-referential but marked noun phrases are discourse prominent, nor for such cases in which non-marked and non-referential noun phrases are discourse prominent.

Overall, the theoretical and empirical findings of this dissertation underlined the strong correlation between the referential and discourse properties of the referents headed by *pe*, *so'n* and *this*. However, I argue that the distinction between discourse prominent and less prominent referents is more basic to these types of referring expressions than that of reflecting referentiality, as discourse prominence reflects the intention to give structure to the discourse at a global level. In line with Givón's considerations (1982, 1984) on other markers of indefinite noun phrases, I consider that when introducing a referent by means of the three markers central to this dissertation, speakers signal their referential intention to mention a referent that will play an important role and thus give structure to the ensuing discourse.

In the following, I bring independent evidence for the general need displayed by various languages to distinguish between prominent and non-prominent referents at the discourse level. First, I briefly discuss the grammaticalization path of the indefinite article, the possible predecessor (with respect to its function) of the three indefinite markers discussed in this dissertation. According to different studies (Givón 1981, Heine 1997, Stark 2002, among many others), the indefinite article entered different languages as a presentative marker for referents that were likely to be picked up in the following discourse. Second, I draw on data from the grammaticalization of differential object marking in Romanian and show that the *pe*-marker was used for topical referents and for referents, which showed a strong referential persistence after being introduced in the discourse. This indicates that, at least in Romanian, the function of *pe* was to mark discourse prominent referents, regardless of the type of referring expression headed by *pe*. In nowadays Romanian, however, the initial function of *pe* is visible only in such contexts that allow for a fluid alternation, that means in contexts in which the marker is optional (i.e. with definite and indefinite noun phrases). Additionally, I briefly discuss a recent study (von Heusinger and Chiriacescu 2011), which pointed out exactly this aspect, namely that *pe*-marking boosts the discourse prominence (in terms of referential persistence and topic shift potential) of definite noun

phrases as well. The picture that emerges by the end of this Chapter is that differences in discourse prominence are not a synchronic phenomenon of the three languages under investigation, but that they reside in the diachronic structure of these languages as well.

### **6.3.1 Grammaticalization of the indefinite article cross-linguistically**

In this Section, I briefly present the grammaticalization path of the indefinite article cross-linguistically to bring support for the observation that languages need to distinguish between prominent and less prominent referents. As in any other grammaticalization process, the evolution of the English, German and Romanian indefinite article systems followed several (partially overlapping) stages. Many linguists (Greenberg 1987, Givón 1981, Heine 1997, Stark 2002) agree that, in the majority of languages, the starting point for the evolution of an indefinite article was the number word for ‘one’. The next stage was the so-called ‘presentational use’ of the indefinite article that described a grammaticalization stage in which the article preceded a new and specific referent, which was necessarily picked up in the subsequent text (i.e. it was characterised by a high referential persistence). Analysing indefiniteness in Old Italian, Stark (2002: 318) makes the same observation, namely that on the way to systematic grammaticalization, various languages “mark each important, individualized referent, i.e. marking of highly ‘persistent’ (Givón 1981) elements, which remain a matter of interest (topical) for a considerable portion of text”. Heine (1997: 72) describes this second stage of grammaticalization, as one in which the newly developed article “introduces a new participant presumed to be unknown to the hearer, and this participant is taken up as definite in subsequent discourse”. These studies underline the observation that languages needed a device to mark newly introduced referents, which were also more likely to be picked up in the ensuing discourse due to their enhanced prominence.

In the course of time, however, the use of the indefinite article spreads to contexts in which it introduces all types of referents, including less prominent ones, which are characterized by their ‘specificity potential’ (Heine: 72), in order to finally reach the last stage, namely that of a generalized article (Givón 1981, Heine 1997). The five-stage model proposed by Heine (1997) for the grammaticalization of the indefinite article is given in (6) below.

## (6) Heine's model of the grammaticalization of the indefinite article (1997: 73)

numeral *one* > presentative use > specific indefinite > non-specific indefinite > generalized article

As discussed by (Givón 1981, 2001) and Borer (2005), modern Hebrew uses the marker *xad*, with the meaning 'one' to mark referents with particular referential properties and discourse characteristics. However, I pointed out in Chapters 3 to 5, when discussing the referential properties of indefinite noun phrases, that the simple indefinite article in the three languages under investigation is neutral or underspecified with respect to triggering particular referential properties of their referents. Furthermore, as it can be used with different types of nouns, the indefinite article in English, German and Romanian reached the final stage of a generalized article (Givón 1981). To conclude, the indefinite articles in the languages under consideration fail to distinguish between "presentatively" introduced referents (i.e. brand-new, specific referents, which will be mentioned again in the subsequent discourse segment) and referents that will not play a prominent role in the ensuing discourse. It can be argued that the markers *this*, *so'n* and *pe* have developed to indicate exactly this distinction in terms of prominence, which was lost, whereas in languages like modern Hebrew, in which the indefinite article did not reach the final stage of grammaticalization, it is still used to mark this distinction.

The development of the indefinite article along the five stages of grammaticalization presented in (6) above motivates on independent grounds the distinction languages used to make between prominent and non-prominent referents (in terms of referential persistence, as an indicator of discourse structuring). The evolution of the indefinite article might foreshadow the grammaticalization path of the three markers discussed here. I argued in the preceding Chapters that indefinite noun phrases headed by *this*, *so'n* and *pe* are discourse prominent in terms of their potential to give structure to the discourse, and, at the same time, that they display a preference for referential readings. Moreover, I showed in Section 6.2 above that the discourse and referential properties of the referents headed by the three markers are not identical. While *pe*-marking in Romanian and indefinite *so'n* in German allow for non-specific interpretations of their referents (especially in the so-called neutral or transparent contexts), indefinite *this* in English is more rigid, as it seems to trigger the referential reading of its associated referents in most contexts. Assuming that the special markers discussed here (will) spread along the same grammaticalization path as

the indefinite article, and that the different stages of this developmental path do not exclude each other, but often overlap, it could be assumed that the Romanian and German markers are more advanced in their grammaticalization process compared to English indefinite *this*.

After showing that the distinction between prominent and non-prominent referents constituted the first stage in the grammaticalization process of the indefinite article from the numeral word for “one”, I briefly discuss a more recent study on the distribution of Differential Object Marking in Romanian (von Heusinger and Chiriacescu 2011). This study underlines the fact that the distinction between prominent and less-prominent referents is important not only for referents realized as indefinite noun phrases, but also for referents mentioned as definite noun phrases.

### 6.3.2 DOM in Romanian as a marker of discourse prominence

In accounting for the distribution of *pe*-marking with definite unmodified noun phrases in Romanian, which represents a theoretical puzzle for different accounts on DOM in this language (Dobrovie-Sorin 1994 and others), in von Heusinger and Chiriacescu (2009, 2011) we tested the discourse behaviour of definite noun phrases in a sentence-continuation experiment that resembles the one presented in Chapter 5. This experiment aimed to find out whether direct objects realized as definite unmodified noun phrases show a preferential status within the discourse similar to that of the *pe*-marked indefinite noun phrases. The findings of the study revealed two patterns, namely (i) that *pe*-marked definite direct objects are referentially more persistent in the following discourse, and (ii) that *pe*-marked direct objects show a preference to become topics two or three sentences after being introduced in the discourse. Thus, exactly like their indefinite counterparts, referents realized as definite unmodified noun phrases headed by *pe* are more discourse prominent compared to referents introduced by the simple indefinite article. These findings did not only differentiate between the contexts in which one form of the direct object (i.e. *pe*-marked vs. non-*pe*-marked) is preferred over the other, but also pointed out the fact that the distinction between prominent and less prominent referents is not tight to the type of referring expression used. In other words, *pe*-marking is employed to signal discourse prominence, independently of the type of referring expression used. This observation is confirmed by the grammaticalization of differential object marking in Romanian. In order

to investigate the factors triggering DOM in Romanian, in Chiriacescu (2007) I performed a diachronic study, which comprised the time period 1800-2000. I coded different properties of direct object and subject referents in about 14 short stories written within the aforementioned time span of 200 years. Due to lack of space, I cannot go into all details revealed by the diachronic study, but see Chiriacescu (2007) for a detailed presentation of the results. Overall, the findings of the corpus investigation reveal the fact that *pe* started to head such direct objects that appeared in syntactic topical positions, in order to then spread along the referentiality scale to the right, to pronouns, proper names, definite and indefinite noun phrases. The part of the survey relevant for the purposes of this dissertation is the investigation of the behaviour of the referents of the *pe*-marked objects with respect to their referential persistence. The results revealed that referents headed by *pe* were more likely to be mentioned in the following five sentence-continuations, compared to their unmarked counterparts.

Summing up, I assume that DOM-marking in Romanian developed in order to underline the contrast between referents that will be referential persistent and such referents that will remain unmentioned in the following discourse, independently of the type of referring expression used for that purpose. In line with the conclusions presented in von Heusinger and Chiriacescu (2010), I assume that the initial function of *pe*-marking is visible in synchronic Romanian only in contexts in which the marker is optional, i.e. with definite and indefinite noun phrases. These considerations parallel the assumptions made in Dalrymple and Nikoleava (2011), namely that DOM can be explained in terms of information structure. In light of the findings from Romanian, I argue that the function of DOM is to signal discourse prominence in terms of high referential persistence and topic shift potential.

The more general claim I wish to make based on these findings is that in SVO languages, the direct object position is more likely to host a marker to distinguish between prominent and less prominent referents. It was noted in Chapter 2 and pointed out several times throughout this dissertation, that referents mentioned in subject position are inherently prominent, as they generally represent hearer-old and discourse-old information, as they appear in sentence-initial position, as they generally denote human referents, which are generally realized as semantic Agents (in a transitive event with semantic Agent and Patient roles), and because speakers tend to encode important information in subject

position (Givón 1983, Brennan et al. 1987, Gundel 1988, Crawley et al. 1990, Gordon et al. 1993, Lambrecht 1994, Arnold 1998, Walker et al. 1998, Ariel 2001). Referents realized in other syntactic positions, and especially in direct object position, are generally considered less prominent, as they tend to represent the amount of new information in a sentence. It is natural, then, that if a language develops a marker to distinguish between prominent and non-prominent referents, it will do so for referents that are not already prominent. The direct object position is more likely to qualify as a good host for that marker. The Romanian *pe*-marker indicates that an otherwise non-prominent referent realized in direct object position becomes a good competitor for the subject referent in terms of its discourse prominence. Accordingly, it is no surprise that *pe* in Romanian accompanies referents realized in direct object position, indicating their discourse prominence in terms of their upcoming discourse structuring potential. Referents that do not compete with the most prominent subject referents in terms of several characteristics remain unmarked.

The observation that referents in direct object position are more prone to receive a special marker is underlined by the occurrence of indefinite *this* in English, which appears in direct object position (Givón 1981). German *so'n* in its indefinite use is found in direct object position as well. It could be assumed that languages that have developed DOM, and which allow for optionality in this domain, will not develop an extra marker to distinguish between prominent and less prominent referents in direct object position, as this position is already filled. However, languages that do not have a means to indicate the alternation between prominent and less prominent referents will develop a special marker for that purpose. In the remainder of this Chapter, I discuss several reasons for why the proximal demonstrative is a good candidate to mark discourse prominent referents.

### **6.3.3 A candidate for discourse prominence: the proximal demonstrative**

In Chapters 3 and 4 I enumerated some reasons for why the proximal demonstrative makes a better candidate for indicating discourse prominence in terms of discourse structuring potential compared to other determiners, such as the indefinite article, the simple definite article or the distal demonstrative. I argued, for example, that English indefinite *this* and German *so'n* are usually indicating the speaker's referential intention to refer to an entity that will be elaborated upon in the subsequent discourse. This referential intention is

paralleled at sentence-level by the referential properties of the referents headed by these markers. In other words, indefinite *this* and *so 'n* are usually referential, whereas other types of markers can be more easily associated with a non-referential interpretation. The referents in (7a-c), which are preceded by *a*, *the* and *that* are not “directly referential” terms King (2001, 11), but rather introduce referents in the discourse that are context-independent (and thus also speaker-independent) expressions, for which the assignment of any arbitrary value to the referent in question would result in a felicitous interpretation.

- (7) (a) Mary met a boy today. I don't know who, but I heard she likes him.  
 (b) I would like to meet the fastest bike-rider in the world.  
 (c) Most avid snow skiers remember that first black diamond run they attempted to ski. (King 2001: 53)  
 (d) Most avid snow skiers remember \*this first black diamond run they attempted to ski. (adapted from King 2001: 53)

Indefinite *this*, on the contrary, is a context-dependent expression, as its felicitous use in a particular context is tight to the referential intention of the speaker, as illustrated in (7d). Another reason for the development of the proximal demonstrative to indicate referents with particular referential and discourse properties comes from the observation that only deictic *this* can be used in a “forward-pointing” way, that means for referents that represent hearer-new and discourse-new information, which will be picked up. As Fillmore (1997: 104) puts it: “I can introduce my frog act, or I can introduce my explanation of something, by saying “This is my imitation of a frog“ or “This is my explanation“, respectively; similarly, I can post-announce my frog act or my explanation by saying something like “*That was my imitation of a frog* or *That was my explanation*”. Payne and Huddleston (2002: 1509) talk about “anticipatory anaphora” when they describe the “forward-pointing” property of indefinite *this* in addition to its function of introducing new referents in the discourse.

Various factors that could have contributed to the emergence of *so 'n* in German and *this* in English as markers of discourse structuring devices were presented. While Wald (1983) and Lyons (1999) noted that they do not know any other language besides English indefinite *this* and German *dieser* that use a demonstrative as an indefinite expression, Diessel (1999) brings empirical evidence from Urim to contradict this claim. Furthermore,

Newman (2000) and Jaggard (2001) noted that, in Hausa it is, again, the speaker-*proximal* demonstrative that serves as a base for the indefinite determiner signalling discourse prominence, not the speaker-*distal* one. A closer look at other languages reveals that using a demonstrative to indicate the preferential discourse status of a referent is not uncommon. In a recent study on Finnish, a language that lacks definite or indefinite articles in its standard use, Kaiser (2011) showed that the demonstrative adjective *sallinen* is used in colloquial, spoken language for referents that signal the need to be picked up in the immediately following discourse. Further evidence comes from Polish, a West Slavic language that lacks indefinite or definite articles, but that has a rich inventory of demonstratives, distinguishing between proximal *ten/ ta/ to/ ci* ('this'), distal *tam* ('that', literally 'there' + 'this') and *taki* ('such', literally 'so', 'this way', 'this much'). In 'out-of-the-blue' or indefinite contexts as illustrated in (8) below, speakers prefer *taki*, which is similar to German indefinite *so* 'n over other demonstrative forms. It seems that *taki* can be felicitously used in this context, as further information about the referents it heads must follow.

- (8) Koło mnie mieszka taki facet. Jest całkiem przystojny.  
 Next me lives such.Masc.Nom.Sg guy.Masc.Nom is quite handsome  
 'There is this man who lives next to me. He is pretty good looking.'<sup>34</sup>

The considerations made above make it seem quite natural for a language to use a demonstrative for referents with particular referential and discourse properties. In fact, the data presented from other languages indicate that this is the case, even though these markers might not be used in the same straightforward way as English *this* and German *so* 'n.

<sup>34</sup> Thank you to Barbara Tomaszewicz for providing the translation for this and other examples on Polish and for discussing this data with me.

## 6.4 Summary

In this Chapter, I investigated the relation between the preference for referential readings and the discourse prominence in terms of discourse structuring displayed by referents associated with indefinite noun phrases. I argued that the status occupied by a particular referent in the discourse determines its referential properties. More concretely, I showed that the referential properties (i.e. wide scope, rigid reference, etc.) of the indefinite noun phrases headed by English indefinite *this*, German *so'n* and Romanian *pe*-marking can be analysed as epiphenomena of their discourse prominence as discourse structuring devices.

The article systems of the three languages under consideration here (i.e. Romanian, German and English) distinguishes between definite articles and indefinite ones, a distinction that guides the comprehender towards a referent that is known to him or not. However, the distinction between the definite and indefinite articles does not reflect whether a newly introduced referent will play a more important role in the ensuing discourse, or whether the speaker has a particular individual in mind, or not. Furthermore, as DuBois (1980) notes, the indefinite article serves to introduce both props (i.e. predicate modifiers with no referential function) and participants (actual discourse referents). I argue that the three markers investigated in this dissertation emerged to mark exactly this distinction between prominent and less prominent referents. This seems to indicate that the distinction between prominent and non-prominent discourse referents is relevant for such referents that are introduced in direct object position and not only for referents realized as indefinite noun phrases. As I argued above, direct objects that make good competitors for the subject referent in terms of discourse prominence, tend to receive more grammatical marking compared to less prominent direct object referents. The fact that prominent direct objects are distinguished from non-prominent direct objects receives support from the grammaticalization of the indefinite article.

## 7. Conclusions

The research presented in this dissertation contributes to the general understanding of indefinite noun phrases. Whereas previous linguistic investigations focused primarily on the behaviour and contribution of indefinite noun phrases at sentence level, this dissertation set out to test the import of different types of indefinite noun phrases at the discourse level as well. On the one hand, I investigated the discourse structuring potential of indefinite noun phrases and, on the other hand, I focused on how the referential properties of particular referents, which become visible at sentence level, relate to the discourse properties of these referents. More concretely, I investigated the relation between the referential properties of referents associated with different types of indefinite noun phrases in terms of discourse prominence as discourse structuring potential. The experiments presented in this dissertation revealed novel aspects of indefinite noun phrases that pertain to the larger discourse segments in which they occur. In this final Chapter, I review the main claims and findings of this dissertation and consider directions for future research.

This dissertation focuses on the discourse contribution of indefinite noun phrases in terms of giving structure to the following discourse. The general claim of the dissertation is that indefinite noun phrases differ with respect to their discourse structuring potential. I compared indefinite noun phrases introduced by means of the simple indefinite article with indefinite noun phrases headed by the markers *pe*, *so 'n* and *this* in Romanian, German and English respectively. I argued that the two types of noun phrases differ with respect to their discourse prominence. To test this discourse prominence in terms of discourse structuring potential, I used three quantifying methods, namely (i) referential persistence, (ii) topic shift potential and (iii) the type of referring expression used to pick up the referents. In light of the story-continuation experiments presented in Chapters 3-5, I showed that referents associated with these indefinite noun phrases are referentially more persistent in the subsequent discourse, and more likely to shift the current topic of the discourse. In contrast, simple indefinite noun phrases are underspecified with respect to these two characteristics. The third characteristic did not show differences between the two indefinite forms, as referents associated to both marked and unmarked indefinite noun phrases were picked up by explicit types of referring expressions (i.e. definite noun phrases).

Contrary to former approaches, my analysis integrates discourse-level properties related to different types of indefinite noun phrases and discusses these findings against prior accounts that explained the behaviour of such indefinite noun phrases in terms of (different instantiations of) specificity. More concretely, I showed that at sentence-level, the indefinite noun phrases headed by the three markers display a high tendency for referential readings (i.e. they tend to be interpreted as taking wide scope, to have fixed reference, etc.) as well. There is a striking convergence between semantic specificity and discourse prominence for the indefinite noun phrases headed by the markers central to this dissertation. Based on different examples, I showed that the stability of reference these three types of indefinite noun phrases display derives from the discourse prominence constraint associated with their referents. When using one of the three indefinite noun phrases, the speaker signals his intention to refer to a particular entity and for which the hearer is instructed to establish a permanent representation as further information about this referent will follow. The indefinite article, on the contrary, is underspecified with respect to this constraint. That means, upon using an indefinite noun phrase headed by a simple indefinite article, the speaker does not commit himself to elaborating upon the referents introduced in this way and thus, the hearer is not instructed to expect further information about this referent (though this possibility is not excluded either). In sum, returning to the relation between the referential properties and the discourse properties of the referents of the indefinite noun phrases headed by the three markers, I argue that their discourse prominence triggers particular referential properties in certain contexts, but not others. Thus, these indefinite noun phrases show an affinity for particular referential properties (e.g. fixed reference, wide scope) because of discourse prominence, while the availability of wide scope readings is not an absolute requirement.

In Chapter 6 I argued that it is not surprising that languages have developed a means to mark such referents that will play a preferential role in the ensuing discourse, as the indefinite article once fulfilled this presentative function (i.e. it introduced referents in the discourse that were important for the following discourse.) Moreover, it is not surprising that indefinite noun phrases that will play an important role in the subsequent discourse are often realised in grammatical direct object position, as it is assumed that new referents are usually mentioned in this position (Prince 1981).

As the three languages under investigation make distinctions in their article system between definite and indefinite determiners, this distinction helps the interlocutor to decide whether the discourse referent in question is known to him or not. However, the interlocutor is unable to conclude whether a newly introduced referent will be important for the development of the discourse or not. DuBois (1980) mentions that the indefinite article in English and possibly other languages can introduce both ‘props’ (predicate modifiers with no referential function) and ‘participants’ (actual discourse referents). Furthermore, in the case of definite noun phrases, the information between speaker and hearer is symmetrical, that means that both discourse partners can identify the referent. In the case of indefinite descriptions, the information both participants to the discourse have is asymmetrical, in the sense that the speaker could refer to an entity that cannot be identified by the interlocutor. I argued that in this model, the role of the noun phrase preceded by the three markers is to adjust the informational asymmetry usually encountered with indefinite descriptions. Even though the interlocutor cannot immediately identify the referent, he is nevertheless given a hint that the referent is more important and will be taken up again. In other words, the information asymmetry between speaker and hearer is temporary.

The predictions that result in light of the findings of this dissertation are, that languages with Differential Object Marking (DOM) will indicate discourse prominent referents, which are expected to be elaborated upon in the subsequent discourse, by means of these markers, whereas languages that do not show DOM will employ other means for such referents. While this study has focused on the discourse structuring potential of indefinite descriptions in Romanian, English and German, we have seen connections throughout to related areas that should be investigated more closely in future work in light of the findings presented here. Furthermore, it would be interesting to see how the implications of the present results relate to other types of indefinite noun phrases as well.

The findings of this dissertation were integrated in an expectancy-driven communication model. Recent psycholinguistic research has convincingly shown that statistical regularities are observed at different levels of linguistic output, as for example at the phonetic level (Saffran et al. 1996), at the syntactic level (Elman 1993, Levi 2008) and at the semantic level (Tabossi 1988, Bicknell et al. 2008). All these studies bring favourable evidence for the fact that hearers identify frequency patterns in order to predict what is likely to occur in the following context. The discourse prominence of the indefinite noun phrases headed by

*pe*, *so'n* and *this*, which is calculated in terms of the expectancy of referential persistence and topic shift potential, underline the fact that language users make use of regularities at the discourse level. The necessity to dissociate between local-level and discourse-level expectations was underlined throughout the dissertation. The first type of expectation, which is generally discussed in the literature on pronoun resolution in terms of accessibility, can be derived from structural factors such as a particular syntactic position and a certain argument structure. Based on such factors, hearers make probabilistic expectations about those referents that are more probable to be mentioned next and more prone to be pronominalized. The second type of expectation, which is important for the present analysis, comes about by discourse structuring devices such as the three types of indefinite noun phrases discussed in this dissertation. Such linguistic devices target larger discourse units in that they indicate which referent will be mentioned over larger discourse segments. It is important to note that the discourse structuring potential is not to be confounded with accessibility or activation, as different types of referring expressions that are associated with a high discourse structuring potential are not necessarily more activated or accessible compared to referring expressions associated that are not associated with a discourse structuring potential. Moreover, I argued that the expectancy of continuation (i.e. the forward pointing potential) is a property defining indefinite noun phrases, but that at the same time it is a fine-grained feature that differentiates between several sub-types of such expressions. In other words, a referent mentioned for the first time by means of an indefinite noun phrase is more or less prone to be mentioned again and to shift the topic in the following discourse. Moreover, I showed that the three languages under investigation do not differ with respect to the expectancy types they employ, as both local-level effects and discourse-level effects are realized. What is different is the formal marking employed by each language for each function and the relative weighting that languages assign to each of them. The main suggestion resulting from the presented evidence is that expectancy operates in parallel at different levels of representation.

This dissertation leads to several questions about the nature of discourse-level structures and relationships, about their interaction with other levels of linguistic representation, and about the types of phenomena that might be sensitive to coherence-driven factors. One of the fundamental questions underlying theories of language production and comprehension concerns referent-tracking, including what referents are preferred to be picked up in the subsequent discourse. In the course of this dissertation I discussed several factors have an

impact upon the expectancies towards the upcoming discourse to a greater or lesser degree. The experimental findings reported in this dissertation confirmed recent studies that showed that the likelihood of mention and the likelihood of pronominalization relate to two textual characteristics of a referent that should be kept apart, as they point into different directions. One path for future research would be to examine other factors that might impact or trigger one of the two expectations (i.e. referential persistence and the topic shift potential) in order to determine what these two expectancies tell us about the discourse status and accessibility of a referent.

One such path was briefly discussed in Chapters 4, 5 and 6. In light of the findings from the sentence-continuation experiment, it became clear that the factors that pertain to sentence-level phenomena (e.g. grammatical role, thematic roles) or to two-sentence contexts (parallelism, coherence relations, verb-type biases) will have to be eventually integrated into a global model of communication. It seems that lexical semantic knowledge provides information about the subsequent discourse. For example, when people talk about a hospital, they activate expectations for hearing about patients being operated on, about an operation event in general, because people know what hospitals are for. Furthermore, grammatical structure (i.e. passive vs. active sentences, etc.) can both facilitate and hinder expectations for certain concepts to occur, as they impose a certain grammatical structure which has to be respected. In addition, the wider discourse and the event schema activated at a particular point can alter and even override the expectations, as it can change the global structure of the events being described. The continuation stories provided by the participants in the three experiments discussed in this dissertation point to the fact that event schemas give rise to various probabilistic expectations on the hearer's side with respect to the subsequent discourse and who will be mentioned next. The findings of the experiment indicate that hearers activate the subevents belonging to a particular superordinate event and generally strive to fill in the missing information of this event whenever it is left unmentioned. At this point we can only speculate that event schemas impact the expectations about who will be mentioned next and have the potential to impact and even override the effects of local factors, such as the presence of different prominence markers. Future research in this area would shed light upon the interaction between various sentence-level and discourse-level factors guiding reference resolution, upon a better understanding of the organization and function of semantic memory.

## BIBLIOGRAPHY

- Abbott, B. (2006). Definiteness and indefiniteness. In Horn, L. R. and Ward, G. L., editors, *The handbook of pragmatics*. Blackwell, Oxford.
- Aissen, J. (2003). Differential object marking: Iconicity vs. economy. *Natural language and linguistic theory*, 21(3): 435–483.
- Alexiadou, A. (2001). *Functional structure in nominals: nominalization and ergativity*. John Benjamins, Amsterdam.
- Almor, A. (1999). Noun-phrase anaphora and focus. The informational load hypothesis. *Psychological review*, 106(4): 748–765.
- Anderson, S. R. and Keenan, E. L. (1985). Deixis. In Shopen, T., editor, *Language, typology, and syntactic description*, volume 3. Cambridge University Press, Cambridge.
- Ariel, M. (1988). Referring and accessibility. *Journal of linguistics*, 24(1): 65–87.
- Ariel, M. (1990). *Accessing noun-phrase antecedents*. Routledge, London; New York.
- Arnold, J. and Griffin, Z. M. (2007). The effect of additional characters on choice of referring expression: everyone counts. *Journal of memory and language*, 56(4): 521–536.
- Arnold, J. and Tanenhaus, M. (2007). Disfluency effects in comprehension: How new information can become accessible. In Gibson, E. and Perlmutter, N., editors, *The processing and acquisition of reference*. MIT Press, Cambridge, MA.
- Arnold, J., Tanenhaus, M., Altmann, M., and Fagnano, M. (2004). The old and thee, uh, new. *Psychological science*, 15: 578–582.
- Arnold, J. E. Marking salience: The similarity of topic and focus. Manuscript. University of Pennsylvania.
- Arnold, J. E. (1998). *Reference form and discourse patterns*. PhD thesis, Stanford University, Stanford, CA.
- Arnold, J. E. (2001). The effect of thematic roles on pronoun use and frequency of reference continuation. *Discourse processes*, 31(2): 137–162.
- Arnold, J. E. (2010). How speakers refer: The role of accessibility. *Language and linguistics compass*, 4(4): 187–203.
- Au, T. K.-F. (1986). A verb is worth a thousand words: The causes and consequences of inter- personal events implicit in language. *Journal of memory and language*, 25: 104–122.
- Auer, Peter, J. (1981). Referential problems in conversation. *Journal of pragmatics*, 8: 627–648.
- Bernstein, J. (1997). Demonstratives and reinforcers in Romance and Germanic languages. *Lingua*, 102: 87–113.
- Bever, T. (1970). The cognitive basis for linguistic structures. In Hayes, J. R., editor, *Cognition and the development of language*. 279–362. Wiley, New York.

- Beysade, C. (2008). *Ma fille sera religieuse ou femme de médecin*. Abstract for journées sémantique et modelisation, Toulouse.
- Bicknell, K., Elman, J. L., Hare, M., McRae, K., and Kutas, M. (2008). Online expectations for verbal arguments conditional on event knowledge. In Love, Bernd, McRae, Ken and Sloutsky, V., editors, *Proceedings of the 30th annual conference of the cognitive science society*, 2220–2225, Austin, TX. Cognitive Science Society.
- Birner, B. and Ward, G. (1998). *Information status and noncanonical word order in English*. John Benjamins, Amsterdam.
- Bolinger, D. (1972). *Degree Words*. Mouton, The Hague.
- Bolinger, D. (1977). *Meaning and Form*. Longman, New York.
- Bossong, G. (1985). *Empirische Universalienforschung. Differentielle Objektmarkierung in den neuiranischen Sprachen*. Narr, Tübingen.
- Bowdle, B. and Ward, G. (1995). Generic demonstratives. In *Proceedings of the Twenty-First Annual Meeting of the Berkeley Linguistics Society*, 32–43.
- Bransford, J. D., Barclay, J. R., and Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, 3: 193–209.
- Brasoveanu, A. and Farkas, D. F. (2010). Talk given at the International Conference Specificity from empirical and theoretical points of view. Stuttgart, August 2010.
- Brennan, S. (1995). Centering attention in discourse. *Language and cognitive processes*, 102: 137–167.
- Brennan, S., Marilyn, F., and Pollard, C. (1987). A centering approach to pronouns. Paper presented at ACL proceedings, 25th annual meeting.
- Bühler, K. (1934). *Sprachtheorie. Die Darstellungsfunktion der Sprache*. Fischer, Stuttgart.
- Burgmann, K. (1904). *Demonstrativpronomina der indogermanischen Sprachen*. Teubner, Leipzig.
- Burkhardt, A. (1982). Abtönungspartikeln als Mittel des Vollzugs präsuppositionaler Akte. Zu Dittmanns und Rombouts' Untersuchungen über die Abtönungsfunktion von auch, denn und doch. *Zeitschrift für germanistische Linguistik*, 10: 85–112.
- Caramazza, A., Grober, E., Garvey, C., and Yates, J. (1977). Comprehension of anaphoric pronouns. *Journal of verbal learning and verbal behaviour*, 16: 601–609.
- Chafe, W. (1974). Language and consciousness. *Language*, 50: 111–133.
- Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Li, C., editor, *Subject and topic*, 25–55. Academic Press, New York.
- Chafe, W. (1994). *Discourse, consciousness, and time*. Chicago University Press, Chicago.
- Chierchia, G. (1993). Questions with quantifiers. *Natural language semantics*, 1: 181–234.
- Chierchia, G. (2001). A puzzle about indefinites. In Checchetto, C., Chierchia, G., and Guasti, M.-T., editors, *Semantic interfaces: Reference, anaphora and aspect*, 51–89. CSLI Publications, Stanford, CA.
- Chiriacescu, S. (2007). *Pe-Markierung und Diskurs-Prominenz im Rumänischen*. Masterarbeit, University of Stuttgart.

- Chiriacescu, S. (2010). Functia prezentativa a lui so'n in limba germana ('the presentative function of German so'n'). In Saramandu, N., Nevaci, M., and Radu, C. I., editors, *Lucrarile celui de-al treilea Simpozion International de Lingvistic*. Bucharest UP, Bucharest.
- Chiriacescu, S. and von Heusinger, K. (2009). Pe-marking and referential persistence in Romanian. In Riester, A. and Onea, E., editors, *Focus at the Syntax-Semantics Interface*. SinSpeC - Working papers of the SFB 732, volume 3, 1–17. University of Stuttgart.
- Chomsky, N. (1995). *The minimalist program*. MIT Press, Cambridge, MA.
- Christophersen, P. (1939). *The articles: A study of their theory and use in English*. Oxford University Press, Oxford.
- Clark, H. (1996). *Using language*. Cambridge University Press, Cambridge.
- Clark, H. and Brennan, S. (1991). Grounding in communication. In Resnick, L., Levine, J., and Teasle, S., editors, *Perspectives on socially shared cognition*. APA Books, Washington.
- Clark, H. H. and Haviland, S. E. (1977). Comprehension and the given-new contract. In Freedle, R. O., editor, *Discourse Production and Comprehension*, volume 1 of *Discourse processes: advances in research and theory*, 1–40. Lawrence Erlbaum Associates, Hillsdale, NJ.
- Clark, H. H. and Marshall, C. R. (1981). Definite reference and mutual knowledge. In Joshi, A. K., Webber, B. L., and Sag, I. A., editors, *Elements of discourse understanding*, 10–63. Cambridge University Press, Cambridge.
- Clark, S. and Card, K. (1969). The role of semantics in remembering comparative sentences. *Journal of experimental psychology*, 83: 545–553.
- Comorovski, I. (1995). On quantifier strength and partitive noun phrases. In Bach, E., Jelinek, E., Kratzer, A and Partee, B. H., editors, *Quantification in natural languages*, 145–177. Kluwer, Dodrecht.
- Comrie, B. (1975). Causatives and universal grammar. *Transactions of the philological society*, 1–32.
- Cornilescu, A. (2001). Romanian nominalizations: Case and aspectual structure. *Journal of linguistics*, 37: 467–501.
- Cornish, F. (1999). *Anaphora, discourse and understanding. Evidence from English and French*. Clarendon Press, Oxford.
- Crawley, R. A., Stevenson, R., and Kleinman, D. (1990). The use of heuristic strategies in the interpretation of pronouns. *Journal of psycholinguistic research*, 19: 245–264.
- Dabir-Moghaddam, M. (1992). On the (in)dependence of syntax and pragmatics: Evidence from the postposition -ra in Persian. In Stein, D., editor, *Cooperating with written texts: the pragmatics and comprehension of written texts*, 549–573. De Gruyter, Berlin.
- Dahl, O. (1970). Some notes on indefinites. *Language*, 46: 33–41.
- Dalrymple, M. and Nikolaeva, I. (2011). *Objects and Information Structure*. Cambridge University Press, Cambridge.

- De Swart, P. (2007). Cross-linguistic variation in object marking. PhD thesis, University of Nijmegen, Nijmegen.
- Diessel, H. (1999). *Demonstratives: Form, Function and Grammaticalization*. John Benjamins, Amsterdam.
- Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cognitive linguistics*, 17(4): 463–489.
- Diver, W. (1984). *The Grammar of Modern English VII*, 39-43. Columbia University, Columbia.
- Dixon, R. (2003). Demonstratives. A cross-linguistic typology. *Studies in Language*, 27: 61–122.
- Dobrovie-Sorin, C. (1994). *The Syntax of Romanian*. De Gruyter, Berlin.
- Dobrovie-Sorin, C. (2007). Article-drop in Romanian and extended heads. In et. al., G. A., editor, *Pitar Mos: A Building with a View*, 99–107. Bucharest University Press.
- Donnellan, K. (1966). Reference and definite descriptions. *Philosophical Review*, 75: 281–304.
- Dorit Abusch (1994). The scope of indefinites. *Natural language semantics*, 2: 83–135.
- Du Bois, J. (1987). The pear story in Sacapultec: a case study of information flow in narrativ. In Furbee, L., editor, *Mayan Voices: Maya text and discourse*, Mayan Texts 4, Native American Texts Series. De Gruyter, Berlin.
- Duden (1973). *Die Grammatik*, volume 4. Dudenverlag, Mannheim.
- Ehlich, K. (1986). So - Überlegungen zum Verhältnis sprachlicher Formen und sprachlichen Handelns, allgemein und an einem widerspenstigen Beispiel. In Rosengren, I., editor, *Sprache und Pragmatik*, volume 55 of *Lunder germanistische Forschungen*, 279–298. Almqvist und Wiksell, Stockholm.
- MacLaran, R. (1982). *The Semantics and pragmatics of the English demonstratives*. PhD thesis, Cornell University.
- Eisenberg, P. (2006). *Grundriß der deutschen Grammatik*. 3., durchges. Aufl. Metzler, Stuttgart.
- Elman, J. L. (1993). Learning and development in neural networks: the importance of starting small. *Cognition*, 48: 71–99.
- Enc, M. (1991). The semantics of specificity. *Linguistic inquiry*, 22: 1–25.
- Enfield, N. (2003). Demonstratives in space and interaction: Data from Lao speakers and implications for semantic analysis. *Language*, 79: 82–117.
- Erteschik-Shir, N. (1997). *The dynamics of focus structure*. Cambridge University Press, Cambridge.
- Escandell-Vidal, V. (2009). Differential object marking and topicality: The case of Balearic Catalan. *Studies in language*, 33(4): 832–885.
- Espinal, M. T. and McNally, L. (2010). Bare singular nouns and incorporating verbs in Spanish and Catalan. *Journal of linguistics*, 45–62.
- Evans, G. (1977). Pronouns, quantifiers, relative clauses. *Canadian journal of philosophy*, 7(4): 777–797.

- Evans, G. (1980). Pronouns. *Linguistic Inquiry*, 11(2): 337–362.
- Farkas, D. (1981). Quantifier scope and syntactic islands. In et al., R. Hendrik, editor, *Papers from the Seventeenth Regional Meeting Chicago Linguistics Society (=CLS) 17*. University of Chicago.
- Farkas, D. F. (1994). Specificity and scope. *Langues et grammaires*, 1: 119–137.
- Ferreira, V. (2003). The persistence of optional complementizer production: Why saying ‘that’ is not saying ‘that’ at all. *Journal of memory and language*, 48: 379–398.
- Ferreira, V. and Dell, G. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive psychology*, 40: 296–340.
- Ferretti, T., Rohde, H., Kehler, A., and Crutchley, M. (2009). Verb aspect, event structure, and coreferential processing. *Journal of memory and language*, 16: 191–205.
- Fillmore, C. J. (1967). On the syntax of preverbs. In *Glossa*, volume 1, 91–125. Glossa Society, New Orleans.
- Fillmore, C. J. (1982). Towards a descriptive framework for spatial deixis. In Jarvella, R. J. and Klein, W., editors, *Speech, place and action*, 31–59. John Wiley and Sons Ltd., Chichester.
- Fillmore, C. J. (1997). Lectures on Deixis. Number 65 in *CSLI Lecture Notes*. CSLI Publications, Stanford, California.
- Fodor, J. and Sag, I. (1982). Referential and quantificational indefinites. *Linguistics and Philosophy*, 5: 355–398.
- Francik, E. (1985). Referential choice and focus of attention in narratives (discourse, anaphora, topic continuity, language production). Doctoral dissertation, Stanford University, Stanford, CA.
- Franco, F. (2005). Infant pointing: Harlequin, servant of two masters. In Eilan, N., Hoerl, C., McCormack, T., and Roessler, J., editors, *Joint attention: Communication and other Minds: Issues in philosophy and psychology*, 129–164. Oxford University Press, Oxford.
- Frege, G. (1892 (1980)). Über Sinn und Bedeutung. In Patzig, G., editor, *Funktion, Begriff, Bedeutung. Fünf logische Studien* [Zuerst erschienen in: *Zeitschrift für Philosophie und Philosophische Kritik*, NF 100, 25-50.], 40–65. Vandenhoeck and Ruprecht.
- Fuchs, A. (1993). *Remarks on Deixis*. Julius Groos Verlag, Heidelberg.
- Fukumura, K. and van Gompel, R. (2011). The effects of animacy in the choice of referring expressions. *Language and cognitive processes*, 3: 1–33.
- Garcia, E. (1975). *The role of theory in linguistic analysis: The Spanish pronoun system*. Benjamins, Amsterdam.
- Garrod, S. and Sanford, T. (1988). Thematic subjecthood and cognitive constraints on discourse structure. *Journal of pragmatics*, 12: 519–534.
- Garvey, C. and Caramazza, A. (1974). Implicit causality in verbs. *Linguistic inquiry*, 5: 459–464.
- Garvey, C., Caramazza, A., and Yates, J. (1975). Factors influencing assignment of pronoun antecedents. *Cognition*, 3: 227–243.

- Geist, L. (2010). Specificity types in Russian. Unpublished manuscript. University of Stuttgart.
- Gernsbacher, M. A. and Shroyer, S. (1989). The cataphoric use of the indefinite *this* in spoken narratives. *Memory and cognition*, 25: 536–540.
- Gernsbachner, M. (1990). *Language comprehension as structure building*. Erlbaum, Hillsdale, NJ.
- Gernsbachner, M. and Hargreaves, D. (1988). Accessing sentence participants: The advantage of first mention. *Journal of memory and language*, 27: 699–717.
- Geurts, B. (2002). Donkey business. *Linguistics and philosophy*, 25: 129–156.
- Gierling, D. (1997). Clitic doubling, specificity and focus in Romanian. In Black, J. and Motapanyane, V., editors, *Clitics, pronouns and movement*, 63–85. Benjamins, Amsterdam.
- Givón, T. (1981). On the development of the numeral "one" as an indefinite marker. *Folia linguistica historica*, 2: 35–53.
- Givón, T. (1983). Topic continuity in discourse: An introduction. In Givón, T., editor, *Topic continuity in discourse: A quantitative cross-language study*, 1–42. Benjamins, Amsterdam.
- Givón, T. (1984). The pragmatics of referentiality. In Shiffrin, D., editor, *Meaning, form and the use in context*. Georgetown University Press, Washington, D.C.
- Gordon, P. and Chan, D. (1995). Pronouns, passives and discourse coherence. *Journal of memory and language*, 34: 216–231.
- Gordon, P., Grosz, B., and Gilliom, L. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive science*, 17(3): 311–347.
- Gramatica Limbii Romane (2005). Editura Academiei, Bukarest.
- Greene, S., Gerrig, R., McKoon, G., and Ratcliff, R. (1994). Unheralded pronouns and management by common ground. *Journal of memory and language*, 33: 511–526.
- Greg, C., Sussman, R., Klein, N., and Tanenhaus, M. (2006). Weak definite NPs. In C. Davis, A. D. and Zabbal, Y., editors, *Proceedings of NELS 36*. UMass/Amherst, Chicago. GLSA.
- Grice, P. (1975). Logic and conversation. In Cole, P. and Morgan, J. L., editors, *Speech acts*, volume 3 of *Syntax and semantics*, 41–58. Academic Press, New York.
- Grosz, B., Joshi, A., and Weinstein, S. (1995). Centering: A framework for modelling the local coherence of discourse. *Computational linguistics*, 21: 2203–226.
- Grosz, B. and Sidner, C. (1986). Attentions, intentions, and the structure of discourse. *Computational linguistics*, 12: 175–204.
- Gundel, J. (1988). Universals of topic-comment structure. In Hammond, M., Moravcsik, E., and Wirth, J. R., editors, *Studies in syntactic typology*, 209–239. Benjamins, Amsterdam.
- Gundel, J., Hedberg, N., and Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 69: 74–307.
- Halliday, M. and Hasan, R. (1976). *Cohesion in English*. Longman, London.

- Hanks, W. (1992). The indexical ground of deictic reference. In Durant, A. and Goodwin, C., editors, *Rethinking context: Language as an interactive phenomenon*, 43–76. Cambridge University Press, Cambridge.
- Haspelmath, M. (1997). *Indefinite pronouns. Oxford studies in typology and linguistic theory*. Oxford University Press, Oxford.
- Hawkins, J. (1978). *Definiteness and indefiniteness. A study in reference and grammaticality prediction*. Croom Helm, London.
- Hawthorne, J. and Manley, D. (2012). *The reference book*. Oxford University Press, Oxford.
- Heim, I. (1982). *On the semantics of definite and indefinite noun phrases*. PhD thesis, University of Massachusetts, Amherst, MA.
- Heim, I. (1991). Artikel und definitheit. In von Stechow, A. and Wunderlich, D., editors, *Semantik: Ein internationales Handbuch der Zeitgenössischen Forschung*, 487–535. De Gruyter, Berlin.
- Heine, B. (1997). *Cognitive Foundations of Grammar*. Oxford University Press, Oxford.
- von Heusinger, K. (2002). Specificity and definiteness in sentence and discourse structure. *Journal of semantics*, 19: 245–274.
- von Heusinger, K. (2007). Accessibility and definite noun phrases. In Schwarz-Friesel, M., Consten, M., and Knees, M., editors, *Anaphors in text: Cognitive, formal and applied approaches to anaphoric reference*, 123–144. Benjamins, Amsterdam.
- von Heusinger, K. (2008). Verbal semantics and the diachronic development of differential object marking in Spanish. *Probus*, 20(1): 1–33.
- von Heusinger, K. (2011a). Specificity. In von Heusinger, K., Maienborn, C., and Portner, P., editors, *Semantics. An international handbook of natural language meaning*, volume 2, 1025–1058. De Gruyter, Berlin.
- von Heusinger, K. (2011b). Specificity, referentiality and discourse prominence: German indefinite demonstratives. In Reich, I., editor, *Proceedings of Sinn und Bedeutung 15*, 9–30. Universaar, Saarland University Press, Saarbrücken.
- von Heusinger, K. and Chiriacescu, S. (2009). Definite "bare" nouns and Pe-marking in Romanian. In Espinal, M., Leonetti, M., and McNally, L., editors, *Proceedings of the IV Nereus international workshop "Definiteness and DP structure in Romance languages"*. Arbeitspapier 124. Fachbereich Sprachwissenschaft, 63–82. Universität Konstanz.
- von Heusinger, K. and Chiriacescu, S. (2010). Definite NPs and DOM-marking in Romanian. Manuscript. University of Stuttgart.
- von Heusinger, K. and Onea, E. (2008). Triggering and blocking effects in the diachronic development of DOM in Romanian. *Probus*, 20(1): 67–110.
- Hemforth, B. and Konieczny, L. (2000). *German Sentence Processing*. Kluwer Academic Publishers, Dordrecht.
- Henn-Memmesheimer, B. (1986). *Nonstandardmuster. Ihre Beschreibung und das Problem ihrer Arealität*. Niemeyer, Tübingen.
- Hideki, M., Kaiser, L., and Masao, O. (1999). Embedded topicalization in English and Japanese. *Lingua*, 109: 1–14.

- Himmelman, N. (1996). Demonstratives in narrative discourse: a taxonomy of universal uses. In Fox, B., editor, *Studies in anaphora*, 205–254. Benjamins, Amsterdam.
- Himmelman, N. (1997). Deiktikon, Artikel, Nominalphrase. Zur Emergenz syntaktischer Struktur. Max Niemeyer Verlag, Tübingen.
- Hintikka, J. (1986). The semantics of a certain. *Linguistic inquiry*, 17(2): 331–336.
- Hole, D. and Klumpp, G. (2000). Definite type and indefinite token: The article *son* in colloquial German. *Linguistische Berichte*, 231–244.
- Hopper, P. and Thompson, S. (1980). Transitivity in grammar and discourse. *Language*, 56(2): 251–299.
- Horn, L. (1988). Pragmatics theory. In Newmeyer, F., editor, *Linguistics: The Cambridge survey*, 113–1145. Cambridge University Press, Cambridge.
- Hottenroth, P. M. (1982). The system of local deixis in Spanish. In *here and there: Cross-linguistic studies on deixis and demonstration*, 133–154. Benjamins, Amsterdam.
- Ionin, T. (2006). This is definitely specific: Specificity and definiteness in articulate systems. *Natural language semantics*, 14: 175–234.
- Jackendoff, R. (1972). *Semantic interpretation in generative grammar*. MIT Press, Cambridge, MA.
- Jäger, G. (2003). Learning constraint sub-hierarchies: The bidirectional gradual learning algorithm. In Blutner, R., editor, *Pragmatics in OT*, 251–287. Palgrave Macmillan, Houndmills.
- Jansen, T. (1993). Tenses and demonstratives: conspecific categories. In Geiger, R. and Rudzka-Ostyn, B., editors, *Conceptualizations and mental processing in language*, 741–783. De Gruyter, Berlin.
- Kaiser, E. (2000). Pronouns and demonstratives in Finnish: Indicators of referent salience. In *proceedings of the third international conference on discourse anaphora and anaphor resolution*.
- Kaiser, E. (2003). The quest for a referent: A crosslinguistic look at reference resolution (Finnish, Estonian, Dutch). PhD thesis, University of Pennsylvania, Philadelphia, PA.
- Kaiser, E. (2010). Investigating the consequences of focus on the production and comprehension of referring expressions. *International review of pragmatics*, 2: 266–297.
- Kaiser, E. (2011). Focusing on pronouns: Consequences of subjecthood, pronominalisation, and contrastive focus. *Language and cognitive processes*, 26(10): 1625–1666.
- Kaiser, E. and Trueswell, J. C. (2008). Interpreting pronouns and demonstratives in Finnish: Evidence for a form-specific approach to reference resolution. *Language and Cognitive Processes*, 23(5): 709–748.
- Kallulli, D. (2000). Direct object clitic doubling in Albanian and Greek. In Beukema, F. and den Dikken, M., editors, *Clitic phenomena in European languages*, 209–248. Benjamins, Amsterdam.
- Kameyama, M. (1996). Indefeasible semantics and defeasible pragmatics. In Kanazawa, M., Pinon, C., and de Swart, P., editors, *Quantifiers, deduction, and context*, 111–138. CSLI, Stanford, CA.

- Kamp, H. (1981). A theory of truth and semantic representation. In Groenendijk, J., Janssen, T., and Stokhof, M., editors, *Truth, interpretation and information*, 1–41. Foris, Dordrecht.
- Kamp, H. and Bende-Farkas, A. (2006). Epistemic specificity from a communication-theoretical perspective. Submitted to *Journal of semantics*.
- Karmiloff-Smith, A. (1985). Language and cognitive processes from a developmental perspective. *Language and cognitive processes*, 1: 61–85.
- Karttunen, L. (1969). Pronouns and variables. In Binnick, R., Davison, A., Green, G., and Morgan, J., editors, *Papers from the fifth regional meeting of the Chicago Linguistic Society*, 108–116. Chicago Linguistic Society, Chicago.
- Karttunen, L. (1974). Presupposition and linguistic context. *Theoretical Linguistics*, 1: 181–194.
- Karttunen, L. (1976). Discourse referents. In McCawley, J., editor, *Syntax and semantics 7: Notes from the Linguistic Underground*, 363–385. Academic Press, New York.
- Kasher, A. and Gabbay, D. M. (1976). On the semantics and pragmatics of specific and non-specific indefinite expressions. *Theoretical linguistics*, 8: 145–190.
- Keenan, E. (1976). Towards a universal definition of "subject". In Li, C., editor, *Subject and Topic*. Academic Press, New York.
- Keenan, E. and Ebert, K. (1973). A note on marking transparency and opacity. *Linguistic inquiry*, 4: 421–424.
- Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. CSLI publications, Stanford, CA.
- Kehler, A., Kertz, L., Rohde, H., and Elman, J. L. (2008). Coherence and coreference revisited. *Journal of semantics*, 25: 1–44.
- Kemmerer, D. (1999). "Near" and "far" in language and perception. *Cognition*, 73: 35–63.
- Kertz, L., Kehler, A., and Elman, J. (2006). Grammatical and coherence-based factors in pronoun interpretation. In *Proceeding of the 28th annual conference of the Cognitive Science Society*, 1605–1610, Vancouver, BC.
- Kirsner, R. (1979a). *The problem of presentative sentences in modern Dutch*. North Holland Linguistic Series. Benjamins, Amsterdam.
- Kirsner, R. S. (1979b). Deixis in discourse: An exploratory quantitative study of the modern Dutch demonstrative adjectives. In Givón, T., editor, *Discourse and syntax*, volume 12 of *syntax and semantics*, 355–375. Academic Press, New York.
- Klein, U. (2007). *Encoding of argument structure in Romanian and Siswati*. PhD thesis, King's College, London.
- Klein, W. (1982). Local deixis in route directions. In Jarvella, R. and Klein, W., editors, *Speech, Place and Action: Studies in deixis and related topics*, 161–182. Wiley, New York.
- König, E., Stark, D., and Requardt, S. (1990). *Adverbien und Partikel*. Julius Groos Verlag, Heidelberg.

- Koornneef, A. and van Berkum, J. (2005). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. *Journal of Memory and Language*, 54(4): 445–465.
- Kornfilt, J. and von Heusinger, K. (2009). Specificity and partitivity in some Altaic languages. In Shibagaki, R. and Vermeulen, R., editors, *Proceedings of the 5th Workshop on Formal Altaic Linguistics*, MIT Working Papers in Linguistics, 19–40. MIT Press.
- Kratzer, A. (1998). More structural analogies between pronouns and tenses. In Strolivich, D. and Lawson, A., editors, *SALT VIII*, Cambridge, MA. CLC Publications.
- Krifka, M. (2001). Quantifying into question acts. *Natural Language Semantics*, 9: 1–40.
- Kripke, S. (1979). Speaker's reference and semantic reference. In French, P., Uehling, T., and Wettstein, H., editors, *Contemporary perspectives in the philosophy of language*, 6–27. University of Minnesota Press, Minneapolis.
- Küntay, A. C. and Özürek, A. (2002). Joint attention and the development of the use of demonstrative pronouns in Turkish. In *Proceedings of Boston University conference in language development 26*, 336–347, Somerville, MA. Cascadilla Press.
- Laca, B. (1987). Sobre el uso del acusativo preposicional en español. In *Romanistisches Jahrbuch 38*, 290–312. De Gruyter, Berlin.
- Lakoff, R. (1974). Remarks on this and that. In *Papers from the tenth regional meeting of the Chicago Linguistic Society*, 345–356.
- Lambrecht, K. (1994). *Information structure and sentence form: topic, focus and the mental representations of discourse referents*. Cambridge University Press, Cambridge, MA.
- Lazard, G. (1984). Actance variations and categories of the object. In Plank, F., editor, *Objects: Towards a theory of grammatical relations*, 267–292. Academic Press, London.
- Legendre, G. (2000). Positioning Romanian verbal clitics at PF: An optimality theoretic analysis. In Gerlach, B. and Grijzenhout, J., editors, *Clitics from different perspectives*, 219–254. Benjamins, Amsterdam.
- Leonetti, M. (2003). Specificity and object marking: the case of Spanish *a*. In von Heusinger, K. and Kaiser, G. A., editors, *Proceedings of the workshop semantic and syntactic aspects of specificity in Romance languages*. Arbeitspapier 113, 67–101, Konstanz. Universität Konstanz.
- Levinson, S. (1977). *Social deixis in a Tamil village*. PhD thesis, University of California, Berkeley, CA.
- Levinson, S. C. (1983). *Pragmatics*. Cambridge textbooks in linguistics. Cambridge University Press, Cambridge, MA.
- Levinson, S. C. (2004). Deixis and pragmatics. In Horn, L. and Ward, G., editors, *The handbook of pragmatics*, 97–121. Blackwell, Oxford.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106: 1126–1177.
- Lewis, D. (1969). *Convention: A philosophical study*. Harvard University Press, Cambridge, MA.

- Lewis, D. (1979). Scorekeeping in a language game. *Journal of philosophical logic*, 8: 339–359.
- Ludlow, P. and Neale, S. (1991). Indefinite descriptions: In defense of Russell. *Linguistics and philosophy*, 14(2): 171–202.
- Lyons, C. (1977). *Semantics*. Cambridge University Press, Cambridge.
- Lyons, C. (1999). *Definiteness*. Cambridge University Press, Cambridge.
- MacDonald, M. and Just, M. (1989). Changes in activation levels with negation. *Journal of experimental psychology: Learning, memory, and cognition*, 15: 633–642.
- MacWhinney, B. (1997). Lexical connectionism. In Broder, P. and Murre, J., editors, *Models of language acquisition: Inductive and deductive approaches*. MIT Press, Cambridge, MA.
- Malchukov, A. and de Hoop, H. (2007). On fluid differential case marking: A bidirectional OT approach. *Lingua*, 117(9): 1636–1656.
- Mardale, A. (2002). Note despre construirea obiectului direct prepositional in Romana si Spaniola. *Studii si cercetari lingvistice*, 1-2: 77–94.
- McDonald, J. and MacWhinney, B. (1995). The time course of anaphor resolution: Effects of implicit verb causality and gender. *Journal of memory and language*, 34: 543–566.
- McElree, B. (2006). Accessing recent events. In Ross, B., editor, *The psychology of learning and motivation*, 155–200. Academic Press, San Diego, CA.
- McRae, K., Spivey-Knowlton, M., and Tanenhaus, M. (1998). Modelling thematic fit (and other constraints) within an integration competition framework. *Journal of memory and language*, 38: 283–312.
- Milsark, G. (1974). *Existential sentences in English*. PhD thesis, MIT, Cambridge, MA.
- Miltsakaki, E. (2003). *The syntax-discourse interface: Effects of the main-subordinate distinction on attention structure*. PhD thesis, University of Pennsylvania, Philadelphia, PA.
- Moore, C. and Dunham, P. (1995). *Joint attention. Its origin and role in development*. Lawrence Erlbaum, Hillsdale, NJ.
- Naess, A. (2004). What markedness marks. The markedness problem with direct objects. *Lingua*, 114: 1186–1212.
- Niculescu, A. (1965). Obiectul direct prepozitional in limbile romanice. In Nikolaeva, I., editor, *Individualitatea limbii romane intre limbile romanice*, 1–49. Editura Stiintifica, Bukarest.
- Oh, S. Y. (2001). A focus-based study of English demonstrative reference: With special reference to the genre of written advertisements. *Journal of English linguistics*, 29(2): 124–148.
- Onea, E. and Hole, D. (2010). *Differential object marking in Romanian and binding*. Manuscript at University of Stuttgart.
- Pană Dindelegan, G. (1997). Din nou despre statutul prepoziției. Cu referire specială la prepoziția PE. *LR XLVI*, 1: 27–55.
- Partee, B. (1970). Opacity, coreference, and pronouns. *Synthese*, 21: 359–385.

- Passonneau, R. (1994). Protocol for coding discourse referential noun phrases and their antecedents. Technical report, Columbia University.
- Perlman, A. (1969). "This" as a third article in American English. *American Speech*, 44(1): 76–80.
- Piwek, P., Beun, R.-J., and Cremers, A. (2008). "Proximal" and "distal" in language and cognition: Evidence from deictic demonstratives in Dutch. *Journal of pragmatics*, 40(4): 694–718.
- Poesio, M. (2004). Centering: A parametric theory and its instantiations. Technical report, University of Essex.
- Prince, E. F. (1981). Toward a taxonomy of given-new information. In Cole, P., editor, *Radical pragmatics*, 223–255. Academic Press, New York.
- Prince, E. F. (1992). The ZPG letter: Subjects, definiteness, and information-status. In Mann, W. C. and Thompson, S. A., editors, *Discourse description: Diverse linguistic analyses of a fund-raising text*, 195–325. Benjamins, Amsterdam.
- Reinhart, T. (1982). Pragmatics and linguistics: An analysis of sentence topics. *Philosophica*, 27: 53–94.
- Roberts, C. (2002). Demonstratives as definites. In van Deemeter, K. and Kibble, R., editors, *Information sharing*, 89–196. CSLI, Stanford, CA.
- Rohde, H. (2008). Coherence-driven effects in sentence and discourse processing. PhD thesis, University of California, San Diego, CA.
- Rohde, H., Kehler, A., and Elman, J. (2006). Event structure and discourse coherence biases in pronoun interpretation. In Sun, R., editor, *Proceeding of the 28th annual conference of the Cognitive Science Society*, 617–622, Mahwah, NJ. Lawrence Erlbaum.
- Russell, B. (1905). On denoting. *Mind*, 14(56): 479–493.
- Russell, B. (1919). *Introduction to Mathematical Philosophy*. Allen and Unwin, London.
- Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294): 1926–1928.
- Sharvit, Y. (1999). Functional relative clauses. *Linguistics and philosophy*, 22: 447–478.
- Sheldon, A. (1974). The role of parallel function in the acquisition of relative clauses in English. *Journal of verbal learning and verbal behavior*, 13: 272–281.
- Sidner, C. L. (1983). Focusing on the comprehension of written anaphora. In Brady, M. and Berwick, R., editors, *Computational models of discourse*, 267–330. MIT Press, Cambridge, MA.
- Silverstein, M. (1976). Hierarchy of features and ergativity. In Dixon, R., editor, *Grammatical categories in Australian languages*, 112–171. Humanities Press, New Jersey.
- Smyth, R. (1994). Grammatical determinants of ambiguous pronoun resolution. *Journal of psycholinguistic research*, 23: 197–229.
- Smyth, R. and Chambers, C. (1996). Parallelism effects on pronoun resolution in discourse context. In *Proceedings of the discourse anaphora and anaphor resolution (DAARC'96)*, 268–280, Lancaster, UK.

- Sperber, D. and Wilson, D. (1987). *Relevance. Communication and cognition*. Harvard University Press, Cambridge, MA.
- Sportiche, D. (1992). Clitic constructions. Manuscript. University of California, Los Angeles.
- Stalnaker, R. (1974). Pragmatic presuppositions. In Munitz, M. and Unger, P., editors, *Semantics and philosophy*, 197–213. New York University Press, New York.
- Stark, E. (2002). Indefinites and specificity in Old Italian texts. *Journal of semantics*, 19: 315–332.
- Stark, E. and Sanda, S. (2008). Why is there differential object marking in Romance? 29th Annual Meeting of the German Society for Linguistics in Bamberg, Germany.
- Steriade, D. (1980). Clitic doubling in the Romanian Wh-constructions and the analysis of topicalization. In *Papers from the sixteenth regional meeting of the Chicago Linguistic Society*, 282–297, Chicago, IL. University of Chicago.
- Stevenson, R., Crawley, R., and Kleinman, D. (1994). Thematic roles, focus, and the representation of events. *Language and cognitive processes*, 9: 519–548.
- Stewart, A., Pickering, M., and Sanford, A. (2000). The time course of the influence of implicit causality information: Focusing versus integration accounts. *Journal of memory and language*, 42: 423–443.
- Strauss, S. (1993). A discourse analysis of “this”, “that” and “it” and their plural forms in spontaneous spoken English. Master’s thesis, University of California, Los Angeles, CA.
- Strauss, S. (2002). This, that, and it in spoken American English: A demonstrative system of gradient focus. *Language sciences*, 24: 131–152.
- Strawson, P. F. (1950). On referring. *Mind*, 59: 320–344.
- Strohmeier, F. (1907). *Der Artikel beim Prädikatsnomen im Neufrenzösichen*. Freiburg, Bielefeld.
- Tabossi, P. (1988). Effects of context on the immediate interpretation of unambiguous nouns. *Journal of experimental psychology: Learning, memory, and cognition*, 14(1): 153–162.
- Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive science*, 2: 49–100.
- Tanenhaus, M. and Trueswell, J. (1995). Sentence comprehension. In Miller, J. and Eimas, P., editors, *The handbook of perception and cognition: Volume 11*, 217–262. Academic Press, San Diego, CA.
- Ionin, T. (2010). The Scope of indefinites: An experimental investigation. *Natural language semantics*, (18): 295–350.
- Terkel, S. (1974). *Working*. Avon, New York.
- Tomasello, M. (1995). Joint attention as social cognition. In Moore, C. and Dunham, P., editors, *Joint attention. Its origins and role in development*, 130–145. Lawrence Erlbaum, Hillsdale, NJ.
- Tomic, O. (1996). Possessive clitics in Macedonian and Bulgarian. In Halpern, A. and Zwicky, A., editors, *Approaching second: Second position clitics and related phenomena*, 511–536. CSLI Publications, Stanford, CA.

- Umbach, C. and Ebert, C. (2010). German demonstrative so - intensifying and hedging effects. To appear in: Hans Christian Schmitz and Henk Zeevat (eds), *Proceedings of the ESSLI X Workshop on formal and experimental approaches to discourse particles and modal adverbs*.
- van Peteghem, M. (1993). *La détermination de l'attribut nominal. Etude comparative de quatre langues romanes*. Brussels, Paleis der Akademie.
- van Rooy, R. (1999). *The specificity of indefinites*. Manuscript University of Amsterdam.
- Wald, B. (1983). Referents and topic within and across discourse units: Observations from current vernacular English. In Klein, F., editor, *Discourse Perspectives on Syntax*, 91–116. Academic Press, Waltham, MA.
- Walker, M. (1998). Centering, anaphora resolution, and discourse structure. In Walker, M., Joshi, A., and Prince, E., editors, *Centering theory in discourse*, 401–435. Clarendon Press, Oxford.
- Walker, M., Joshi, A., and Prince, E. (1998). Centering in naturally occurring discourse: An overview. In Walker, M. A., Joshi, A. K., and Prince, E. F., editors, *Centering theory in discourse*, 1–28. Clarendon Press, Oxford.
- Walker, M. and Prince, E. (1996). A bilateral approach to givenness: A hearer - status algorithm and a centering algorithm. In Fretheim, T. and Gundel, J., editors, *Reference and referent accessibility*, 291–306. Benjamins, Amsterdam.
- Wasow, T. (1997). End-weight from the speaker's perspective. *Journal of psycholinguistic research*, 20: 347–361.
- Wasow, T. and Arnold, J. (2003). Postverbal constituent ordering in English. In Rohdenburg, G. and Mondorf, B., editors, *Determinants of grammatical variation in English*, 120–154. De Gruyter, Berlin.
- Wespel, J. (2006). *Indefinite demonstratives*. Manuscript. University of Stuttgart.
- Wolter, L. (2006). *That's that: The semantics and pragmatics of demonstrative noun phrases*. Doctoral dissertation, University of California, Santa Cruz, CA.
- Wright, S. and Givn, T. (1987). The Pragmatics of indefinite reference: Quantified text-based studies. *Studies in Language*, 11(1): 1–33.
- Yeom, J.-I. (1998). *A presuppositional analysis of specific indefinites: Common grounds as structured information states*. Garland Publishing, New York.
- Zifonun, G., Hoffmann, L., and Strecker, B. (1997). *Grammatik der Deutschen Sprache*. De Gruyter, Berlin.



