

Reference Processing and its Universal Constraints

Reference Processing and its Universal Constraints

A dissertation presented by
António Horta Branco

DISSERTATION TITLE: **Reference Processing and its Universal Constraints**

AUTHOR: **António Manuel Horta Branco, MA**

ACADEMIC INSTITUTION: **Faculty of Sciences, University of Lisbon**

DEGREE: **Doctor of Philosophy**

COURSE: **Informatics**

ADVISORS: **Hans Uszkoreit,
Full Professor at the Department of Computational
Linguistics, University of Saarland, Germany**
**José Luiz Fiadeiro,
Associate Professor at the Department of Informatics,
University of Lisbon, Portugal**
**Palmira Marrafa,
Assistant Professor at the Department of Linguistics,
University of Lisbon, Portugal**

RESEARCH SITES: **Department of Informatics, University of Lisbon,
and Language Technology Group of DFKI-German
Research Center for Artificial Intelligence**

SUBMISSION DATE: **September 1999**

APPROVAL DATE: **December 1999**

COMMITTEE: **Ducla Soares (Chair), João Branquinho, Helder Coelho,
José Luiz Fiadeiro, Miguel Filgueiras, Palmira Marrafa,
Hans Uszkoreit**

Contents

Contents	5
Abstract	9
Resumo em Português	11
Acknowledgments	15
1 Introduction: Anaphora and Reference Processing	19
1.1 Introduction to the Dissertation	20
1.2 Anaphora	23
1.3 Reference Processing	31
1.4 Summary	35

2	Heuristics and Constraints	37
2.1	The Integrative Framework	40
2.2	Recent Developments and Lines of Research	54
2.3	Centering Attention for Anaphor Resolution	70
2.4	Summary	87
3	Empirical Universality	89
3.1	Universal Constraints	90
3.2	Split Obliqueness and Derived Constructions	102
3.3	Branching Obliqueness and Subject-orientedness	111
3.4	Summary	119
4	Symmetries and Duality	121
4.1	Long-Distance Reflexives	122
4.2	New Symmetries with Four Constraints	135
4.3	The Quantificational Structure of Binding Constraints	139
4.4	New Insights into Binding	146
4.5	Summary	146
5	Formal Semantics	146
5.1	Obliqueness Quantification in Grammar	146
5.2	Obliqueness Quantification as a Component of Meaning	146
5.3	Binding Constraints as Determinants of Meaning Composition	146
5.4	From Anaphor Resolution to Reference Processing	146
5.5	Summary	146

6	Constraint-based Specification	146
6.1	The Grammatical Knowledge and Processing Framework	146
6.2	A Module of Linguistic Knowledge at Odds in Grammar	146
6.3	Binding Constraints in a Constraint-based Framework	146
6.4	Summary	146
7	Computational Implementation	146
7.1	Implementation Formalism	146
7.2	Implemented Grammar	146
7.3	Implementation of Binding Constraints	146
7.4	Summary	146
8	Conclusions: Summary and Outlook	146
8.1	Summary of the Dissertation	146
8.2	From the Angle of Antecedents	146
8.3	Binding Constraints in Discourse	146
8.4	Gaps and Linguistic Universals	146
8.5	Generation of Anaphoric Expressions	146
8.6	Cognitive Underpinnings	146
	Annex I Implemented Grammar	146
	Annex II Examples	146
	Example 1	146
	Example 2	146
	Example 3	146

Bibliography	146
Index of Names	146
Index of Languages	146
Index of Definitions	146

Abstract

A systematic research on reference processing from the angle of binding constraints over anaphoric links is presented in this dissertation.

Empirically, binding constraints exhibit a universal dimension given that, under suitable parameterization, they appear to hold across natural languages from different language families. Conceptually, they are interrelated in a highly non-trivial, symmetric fashion, lending them a modular character as a consistent set of conditions on reference.

Inasmuch as they appear as conditions on the antecedents anaphors can take, binding constraints are of critical importance for reducing the search space of the complex procedure of anaphor resolution and enhancing its accuracy. However, attempts towards formal and computational accounts of these constraints have persistently come up against notorious resistance concerning their integration into the course of reference processing.

Our primary aim in this dissertation is thus to dissolve this gap between, on the one hand, the practical relevance, empirical adequacy and conceptual elegance of binding constraints, and on the other hand, their thorough integration into the semantic analysis of natural languages. Accordingly, we aim to clarify how binding constraints contribute to reference processing by fully integrating them into natural language grammars and by discussing how that can be practically achieved in a prominent constraint-based framework such as Head-driven Phrase Structure Grammar.

While this is the unifying goal of the present dissertation, several other contributions were also obtained. A complete list of such instrumental results and their implications is provided in the last chapter, from which we would like to emphasize the following two points:

Abstract

We argue for an innovative understanding of binding constraints, as an alternative to the mainstream view envisaging them as well-formedness conditions on syntactic representations. Under this new conception, binding conditions are the effect of phase quantifiers over reference markers arranged in obliqueness hierarchies, thus constraining the interpretation of the anaphoric expressions they happen to be associated with.

We offer an algorithm for the verification of binding constraints, as an alternative to the mainstream one, which is exponential and, moreover, has the drawback of requiring extra grammatical parsing steps. The algorithm we designed is polynomial, is completely integrated into the grammatical parsing process, was given a full-fledged constraint-based specification and it has an exemplifying implementation in a computational grammar we developed.

Resumo em Português

(Abstract in Portuguese)

Na presente dissertação procedemos a um estudo sistemático do processamento de referência do ponto de vista das restrições de ligação (binding constraints) sobre os elos anafóricos (anaphoric links).

Problema

As restrições de ligação incidem sobre as relações anafóricas e constituem um tópico central de investigação no domínio do processamento da semântica das linguagens naturais. Em termos empíricos, estas restrições tem por base generalizações empíricas com uma robustez notável, cuja validade parameterizável para as diferentes linguagens naturais lhes confere um carácter de universalidade. Em termos conceptuais, as restrições de ligação exibem entre si relações que apresentam uma certa simetria, a qual lhes confere uma natureza modular enquanto conjunto consistente de condições sobre a referência.

Nesta medida, e visto surgirem como condicionantes não triviais dos antecedentes que podem ser tomados pelas expressões anafóricas nominais – os sintacticistas costumam referir-se-lhes como princípios de ligação (binding principles) –, as restrições de ligação são da maior importância para reduzir o espaço de pesquisa no complexo processo de resolução de anáforas (anaphor resolution) e melhorar a exactidão deste.

De forma contrastante, contudo, as tentativas de tratamento formal e computacional das restrições de ligação têm revelado de forma persistente a existência de dificuldades consideráveis quando se trata de encontrar uma solução para sua cabal integração no curso do processamento de referência.

Do ponto de vista da resolução de anáforas, as restrições de ligação têm sido aleatoriamente entendidas quer como preferências quer como filtros – parecendo claramente ser este último o seu estatuto porém –, sendo que de qualquer dos modos o seu tratamento integral muito raramente tem sido alvo de atenção consequente.

No que toca à forma como tipicamente têm sido integradas na gramática, também aqui se tem levantado inúmeras dificuldades. Na formulação actualmente em voga, cujas primitivas foram estabelecidas no início dos anos oitenta, o processo de verificação das restrições de ligação sobre representações linguísticas envolve um algoritmo exponencial, o qual exige inclusive vários passos de triagem (parsing) extra-gramatical. Mais recentemente, mesmo quando se trata de enquadramentos teóricos proeminentes para a representação e processamento do conhecimento gramatical baseados em restrições (constraint-based), estes ou pressupõem extensões dedicadas do formalismo de descrição para lidar com as restrições de ligação (e.g. Gramática Lexical-Funcional – Lexical Functional Grammar), ou não proporcionam de todo qualquer integração para estas restrições no edifício gramatical (e.g. Gramática de Estruturas Sintagmáticas determinadas pelos Núcleos (GESN) – Head-driven Phrase Structure Grammar).

Objectivo

O objectivo principal subjacente à investigação apresentada nesta dissertação é pois o de se obter a eliminação do hiato entre, por um lado, a adequação empírica, a elegância conceptual, e a relevância prática das restrições de ligação, e por outro lado, a total e consistente integração destas últimas na gramática das linguagens naturais e no curso do processamento de referência.

As restrições de ligação tomam a forma de condições gramaticais sobre os possíveis elos anafóricos entre expressões anafóricas e seus potenciais antecedentes. Deste modo, encarada sob o ângulo do seu fio condutor mais genérico, a nossa investigação procura clarificar se e como podem as restrições de ligação receber uma total integração no processamento de referência por via da sua completa integração nas gramáticas das linguagens naturais. Esta clarificação é procurada através da discussão de como na prática pode tal integração ser realizada num enquadramento teórico proeminente, como é o GESN.

Contributos de maior relevo

Alcançar o objectivo genérico atrás enunciado constitui o contributo central da presente dissertação para o progresso da área do processamento da semântica das linguagens naturais, em geral, e do processamento de referência, em particular. Para o atingir, porém, vários outros resultados foram também conseguidos. Sem perder de vista o elenco pormenorizado desses resultados intermédios, cuja apresentação pode ser encontrada no último capítulo, gostaríamos de salientar à partida dois daqueles que consideramos de maior relevo.

Entre outros contributos, nesta dissertação propomos uma perspectiva radicalmente nova relativamente à compreensão das restrições de ligação, em alternativa à perspectiva em voga, a qual as entende sobretudo como condições de boa formação sobre representações sintácticas. Segundo a nova concepção que defendemos, estas restrições são entendidas como o efeito de quantificadores de fase sobre hierarquias de obliquidade (obliqueness) gramatical, os quais determinam a interpretação das expressões anafóricas às quais se encontram associados.

Discutimos também um algoritmo para a verificação das restrições de ligação, em alternativa ao algoritmo actualmente em voga, que é exponencial e que exhibe ainda o inconveniente de necessitar de uma fase de triagem extra-gramatical. O algoritmo que desenhamos é polinomial, está completamente integrado na triagem gramatical, e encontra-se operativo na gramática computacional que desenvolvemos (cf. capítulo 7 Computational Implementation, e Annex I)

Apresentação global

No capítulo 1 Introduction: Anaphora and Reference Processing, assim como no capítulo 2 Heuristics and Constraints, traçamos uma panorâmica geral do estado da arte no que diz respeito ao processamento de referência e à resolução de anáforas. Pretendemos aí tornar claro como as restrições de ligação devem ser adequadamente integrados no conjunto de factores com incidência sobre o processamento de referência.

No capítulo 3 Empirical Universality descrevemos aqueles que actualmente são tidos como os resultados centrais que qualquer abordagem das restrições de ligação tem de mostrar ser capaz de tratar. Discutimos também desafios recentes que colocam em causa a universalidade das restrições de ligação e tentamos mostrar que esses desafios não são senão o efeito de outras restrições linguísticas em interacção com as restrições de ligação propriamente ditas.

No capítulo 4 *Logical Symmetry* estudamos os pronomes reflexos de longa distância e defendemos a existência de uma quarta restrição de ligação, a qual se encontra associada a esta classe de expressões nominais. Discutimos como esta nova restrição se enquadra na teoria das restrições de ligação existente, baseada apenas em três restrições. Exploramos também as amplas consequências desta integração quer em termos de uma concepção radicalmente nova da ligação (*binding*) enquanto quantificação de obliquidade (*obliqueness quantification*), quer em termos de um novo entendimento no que diz respeito à semântica formal das expressões nominais.

No capítulo 5 *Formal Semantics* ponderamos o modo de dar enquadramento à quantificação de obliquidade no seio da gramática das linguagens naturais. Propomos também como se deve proceder quanto à forma de coordenar as restrições de ligação com o processamento de referência, em geral, e a resolução de anáforas, em particular.

No capítulo 6 *Constraint-Based Specification*, tirando partido da clarificação obtida nos capítulos anteriores, discutimos a integração das restrições de ligação no enquadramento gramatical baseados em restrições proporcionado pela GESN.

No capítulo 7 *Computational Implementation*, após uma breve sinopse dos sistemas computacionais disponíveis para a implementação de gramáticas GESN, discutimos os principais aspectos da implementação do fragmento de gramática de apoio que usámos. Em seguida relatamos como foi feita nessa gramática nuclear a integração das restrições de ligação anteriormente especificadas.

Por último, no capítulo 8 *Conclusions: Summary and Outlook* apresentamos um elenco pormenorizado dos resultados obtidos. Discutimos ainda como a investigação apresentada nesta dissertação pode eventualmente ser estendida a outros temas que parecem oferecer-se como suas continuações naturais. As nossas observações aí serão sobretudo de cariz especulativo. Destinam-se a colocar em destaque o que a priori parecem ser os mais importantes aspectos da integração das restrições de ligação em domínios de investigação sucessivamente mais amplos, em que questões práticas, factores de processamento, ou conjuntos de dados mais vastos são tomados em consideração.

Acknowledgments

Firstly, I would like to present special thanks to my advisors, José Luiz Fiadeiro, Palmira Marrafa and Hans Uszkoreit, who have been exceedingly generous with their time and comments at every stage of my research. This dissertation would not have been possible without their guidance and encouragement.

I am greatly indebted to Miguel Filgueiras, Danièle Godard, Mark Johnson and Stanley Peters, with whom I learned a considerable amount while discussing substantive aspects of my results.

I would also like to thank friends and colleagues who had the patience to read preliminary drafts. My work greatly benefited from the comments and suggestions of João Balsa, Helder Coelho and Berthold Crysmann.

I was very fortunate to have been able to undertake part of my research at the Language Technology Group, headed by Hans Uszkoreit, of the DFKI-German Research Center for Artificial Intelligence. I owe a great deal to friends and colleagues of the Saarbrücken academic community, for their help and enthusiasm, in particular to Andrew Bredenkamp, Martine Grice, Brigitte Krenn, Kordula de Kuthy, Thierry Leclerc, Klaus Netter, Karel Oliva, Wojciech Skut and Christine Thielen. Special thanks go to Gregor Erbach, who generously assisted in handling ProFIT, and to Sabine Lehmann, whose support was crucial for settling in Saarbrücken.

Partial results have been presented at various events, namely at the VI Computational Linguistics in the Netherlands Meeting (CLIN'95), University of Antwerp, Antwerp, Belgium, December 1, 1995; the Indirect Anaphora Workshop, Lancaster University, Lancaster, UK, July 18, 1996; 16th International Conference on Computational Linguistics (COLING'96), University of Copenhagen, Copenhagen, Denmark, August 5-9, 1996; 11th Pacific Asia Conference on

Acknowledgements

Language, Information and Computation (PACLIC 11), Kyung Hee University, Seoul, Korea, December 20-22, 1996; Long-distance Reflexives Workshop, LSA Linguistic Summer Institute, Cornell University, Ithaca NY, USA, July 5-6, 1997; 4th International Conference on Head-Driven Phrase Structure Grammar (HPSG'97), Cornell University, Ithaca NY, USA, July 18-20, 1997; Workshop on Quantification: Collectivity and Reciprocals, 9th European Summer School in Logic, Language and Information (IX ESSLLI), Université de Provence, Aix en Provence, France, August 11-15, 1997; Workshop sobre Linguística Computacional da Associação Portuguesa de Linguística, Faculdade de Letras de Lisboa, Lisbon, Portugal, May 25-27, 1998; 2nd Discourse, Anaphora and Resolution Colloquium (DAARC'98), Lancaster University, Lancaster, UK, August 1-4, 1998; 17th International Conference on Computational Linguistics and 36th Annual Meeting of the Association for Computational Linguistics (COLING-ACL'98), Université de Montréal, Montreal, Canada, August 10-14, 1998; Encontros de Quarta-feira, Departamento de Informática, Faculdade de Ciências da Universidade de Lisboa, Lisbon, Portugal, October 7, 1998; and Seminários de Linguística, Universidade do Algarve, Faro, Portugal, December 16, 1998.

For helpful discussion and criticism, I am grateful to the participants in these events, in particular to Peter Cole, Félix Costa, Howard Gregory, Gabriella Hermon, Arid Hestvik, Jeffrey Lidz, Sérgio Menuzzi, Haihua Pan, Fred Popowich, Ivan Sag, Pedro Santos, Philippe Schlenker, Leonie Bosveld-de Smet and Vasco Vasconcelos.

I am also grateful to my friends Maria Isabel Festas and José Leitão, who helped considerably with bibliography on cognitive studies.

At the Departamento de Informática, I benefited from the permanent and friendly patience of António Casimiro, Graça Gaspar, Antónia Lopes, João Neto and Isabel Nunes.

My research was supported by various institutions, namely Departamento de Informática da Faculdade de Ciências da Universidade de Lisboa, European Commission (project LSGRAM – Large-Scale Grammars), Fundação Oriente, Ministério da Ciência e Tecnologia (program Praxis XXI), LabMAC - Laboratório de Modelos e Arquiteturas Computacionais and Reitoria da Universidade de Lisboa.

“The variable qua variable, the variable an und für sich and par excellence [...] is the essence of the referential idiom. But it takes some distilling, for it has strong affinities with quite a variety of closely associated notions and devices.”

W. V. Quine, *The Variable*, p.272

1 Introduction:

Anaphora and Reference Processing

1.1 Introduction to the Dissertation	20
Problem	20
Goal	21
Chief contributions	21
Overview	22
1.2 Anaphora	23
Polysemic potential and context sensitivity	24
Constructive sensitivity	26
Antecedents and cospecification	27
Contexts	28
1.3 Reference Processing	31
Anaphor resolution	31
Automatic processing of reference	32
Linguistic- and cognitive-driven approaches	33
1.4 Summary	35

1.1 Introduction to the Dissertation

In the present dissertation, we carry out a systematic study of reference processing from the angle of binding constraints over anaphoric links.

Problem

Binding constraints on anaphora are a central research issue in the field of semantic processing. Empirically, these constraints stem from quite robust generalizations and exhibit a universal character, given their parameterized validity across natural languages. Conceptually, the relations between them involve a certain kind of symmetry, lending them a modular nature as a consistent set of conditions on reference.

Accordingly, inasmuch as they appear as non-trivial restrictions on the antecedents an anaphor can take – syntacticians usually refer to them as binding principles –, they are of critical importance in enhancing accuracy and reducing the search space in the complex process of anaphor resolution.

In contrast to this, however, the formal and computational handling of binding constraints has persistently presented non-negligible resistance when it comes to their integration into the course of reference processing.

From the point of view of anaphor resolution, binding constraints have been casually accepted either on the side of preferences or on the side of filters – which undoubtedly seems to be their natural place –, but in any case in a haphazard and partial fashion.

The way in which they have been typically integrated into grammar, in turn, raises also many questions. In its mainstream formulation, the basics of which date back to the early eighties, the complete process of verification of binding

constraints involves an exponential algorithm inclusively requiring a series of extra grammatical parsing steps. Furthermore, more recently, prominent constraint-based frameworks for grammatical knowledge representation and processing either require special purpose extensions of the description formalism for handling these constraints (e.g. Lexical Functional Grammar) or offer no integration at all for them into the grammatical setup (e.g. Head-driven Phrase Structure Grammar / HPSG).

Goal

The primary goal underlying the research reported in this dissertation is thus to contribute to dissolving the gap between, on the one hand, the conceptual elegance, empirical adequacy and practical relevance of binding constraints, and on the other hand, their thorough integration into grammar and the course of reference processing.

Binding constraints appear as grammatical restrictions on possible anaphoric links between anaphors and their antecedents. Accordingly, the essential thread of our research is directed towards clarifying whether and how binding constraints can be fully integrated into reference processing, by fully integrating them into natural language grammar and by discussing how this can be practically achieved in a prominent constraint-based framework such as HPSG.

Chief contributions

Meeting this goal is the primary contribution of the present dissertation to research on reference processing. In the process of meeting, however, several other results were also achieved. While a complete list of such instrumental results and their implications is provided in the last chapter, we would like to emphasize here two of them we cherish more.

In this dissertation, we argue for a radically new understanding of binding constraints, as an alternative to the mainstream view envisaging them as well-formedness conditions on syntactic representations. Under the innovative conception we argue for, binding conditions are the effect of phase quantifiers over

grammatical obliqueness hierarchies which constrain the interpretation of the anaphoric expressions with which they happen to be associated.

We also argue for an alternative algorithm for verifying binding constraints, in contrast to the currently mainstream, exponential algorithm, which, moreover, has the drawback of requiring extra grammatical parsing steps. The algorithm we designed is polynomial, is fully integrated into the grammatical parsing and has an exemplifying implementation in the computational grammar we developed.

Overview

In chapter 3 Empirical Universality, we provide an introductory description of what is at present a quite stable set of facts and central results that any account of binding constraints must be able to handle. After that, we discuss recent challenges on the universality of binding constraints and try to show that these are but the effects of the conjoined occurrence of other collateral linguistic constraints together with binding constraints proper.

In chapter 4 Symmetries and Duality, we study long-distance reflexives and argue for a formal account of a fourth binding constraint, associated with this class of nominals. We discuss how this new constraint consistently fits into the existing binding theory, based on three constraints, and explore the far-reaching consequences of this integration both for a radically new conception of binding as obliqueness quantification and for new insights into the formal semantics of nominal expressions.

In chapter 5 Formal Semantics, we elaborate on how to accommodate obliqueness quantification in grammar. We also propose how the grammatical representation of binding constraints may be coordinated with the processing of reference, in general, and the resolution of anaphors, in particular.

In chapter 6 Constraint-based Specification, building on clarification from previous chapters, we discuss the integration of binding constraints into the constraint-based framework for grammatical knowledge representation and processing provided by HPSG.

In chapter 7 Computational Implementation, after presenting a brief survey of the computational systems available for the implementation of HPSG particular grammars, we discuss the major aspects of implementing the supporting grammar fragment we used. Subsequently, we report on the integration of the previously specified binding constraints into that core grammar.

Finally, in chapter 8 Conclusions: Summary and Outlook, we make a detailed summary of the results obtained in the dissertation. We also briefly elaborate on how the research reported in this dissertation can be extended to other themes that seem to offer themselves as relevant areas for continuation. Our study here will be exploratory and will point out the seemingly important aspects of integrating binding constraints into successively broader domains, where larger sets of data, processing factors or practical issues are taken into account.

In the remainder of the present chapter and in the next chapter 2 Heuristics and Constraints, we provide an overview of the state of the art in reference processing and anaphor resolution. We aim to illustrate how binding constraints are to be properly integrated into the whole setup of factors impinging on reference processing.

In Section 1.2, we begin by introducing the phenomenon of anaphora and by underlining its basic distinctive linguistic features. In light of this characterization, Section 1.3 presents the rationale for automatic reference processing and discusses the basic goal of anaphor resolvers.

In the next chapter, Section 2.1 is devoted to presenting of a mainstream methodology for anaphor resolution and to the corresponding architecture for anaphor resolvers, while in Section 2.2 we discuss recent developments and open lines of research using this methodology. Finally, we study at length in Section 2.3 a second methodology for anaphor resolution, based on Centering Theory. We discuss also how recent research points towards a converging integration of these two major methodologies into the processing of anaphora.

1.2 Anaphora

It is an inherent feature of natural languages that their expressions may have several different semantic values. A typical example often used to illustrate this point is the expression flying planes, as it can be used either to mean planes that are moving through the air, or to mean the process of piloting planes. Accordingly, a

sensible expectation would be that natural languages are tools of reduced potential, if any, for fast continuous communicative interchanges. Assuming, just for the sake of the contrast, that each utterance has ten words on average, of which, say, four may have two different semantic values each, any speaker would then become completely confused, as there would be on average eight different interpreting possibilities per utterance.

Everyday intensive usage of a thousand natural languages around the world, however, controverts this expectation. As a matter of fact, the polysemic potential of natural languages has a corresponding twin property – the property that the meaning of a given expression can appropriately be contrived given the context in which it occurs. This has been explicitly noticed at least since Frege (1884, p.x), who in the introduction to *Grundlagen der Arithmetik* observes that “one should ask for the meaning of a word only in the context of a sentence, and not in isolation.”

This other trait of natural languages can be illustrated by simply considering the different occurrence of flying planes. In each of the contexts set up in (1), this phrase is easily associated only with one of its two possible semantic values.

- (1) a. Flying planes are complex machines.
- b. Flying planes is a difficult task.

Following the *Handbook of Logic and Language* (van Benthem and ter Meulen, 1997, p.1184), depending on the phenomenon under scrutiny, context may be taken as (i) the linguistic context of some phrase, which is just the surrounding linguistic material; (ii) the utterance context of some statement, which concerns any fact that is taken to be relevant to the utterance, and may specify the speaker, the hearer, what was said, the time of utterance, the beliefs and intentions of the conversational participants, and much more; or (iii) the common ground, which concerns the utterance context, but in the restricted sense of describing only the mutual beliefs of the conversational participants.

Polysemic potential and context sensitivity

Polysemic potential and context sensitivity are thus two cardinal features of natural languages whose formal and computational modeling have turned out to be more elusive. It would certainly be no exaggeration to state that they could be deemed as the common root of the most problematic issues to be tackled by any device aiming to achieve the automatic understanding of natural language.

Anaphora is one of such issues where the challenging effect of these two cardinal features is more blatantly evident. A first glimpse at the sort of problems at stake in dealing with anaphoric expressions can be taken if we consider in some detail the following excerpt from a real-life task-oriented dialogue, adapted from Grosz and Sidner, 1986, p.186. To obtain a realistic idea of the magnitude of the language engineering problem to be sized, we should just allow ourselves to imagine (2) as a future instructional dialogue between a human and a non-human agent.

- (2)
- H:1. First you have to remove the flywheel.
- R:2. How do I remove the flywheel?
- H:3. First, loosen the screw, then pull it off.
- R:4. OK.
5. The tool I have is awkward. Is there another tool that I could use instead?
- H:6. Show me the tool you are using.
- R:7. OK.
- H:8. Are you sure you are using the right size key?
- R:9. I'll try some others.
10. I found an angle I can get at it.
11. The screw is loose, but I'm having trouble getting the flywheel off.
- H:12. Use the wheelpuller. Do you know how to use it?
- R:13. No.
- H:14. Do you know what it looks like?
- R:15. Yes.
- H:16. Show it to me please.
- R:17. OK.
- H:18. Good. Loosen the screw in the center and place the jaws around the hub of the flywheel, then tighten the screw onto the center of the shaft. The flywheel should slide off.

Let us focus, for instance, on the two occurrences of the anaphoric expression *it*, respectively in sentences 3. and 12.. While in the former, it must be taken as referring to the screw of the flywheel, in the latter, that same expression is to be understood as referring to another entity instead, the wheelpuller. Also apparent here, the factor determining this difference in semantic value of a given expression is the context around its different occurrences. Note, however, that contrary to much

common wisdom about anaphora, the context sensitivity exhibited by anaphoric expressions is not due to the simple effect of a null or weak descriptive content, as this could be suggested by the example with it. Similarly, the semantic value of an anaphoric phrase such as the screw, a more content-loaded definite description, also shows this dependency on context. To observe this, just consider the occurrences of this phrase in sentences 3. and 18.. In the former, the screw is being used to refer to the screw of the flywheel. In the latter, it should be understood as referring to another screw, the screw in the center of the wheelpuller.

Constructive sensitivity

Although polysemic potential and context sensitivity are two crucial features involved in the problems we address in this dissertation, they do not yet render a distinctive characterization of our object of study. As a matter of fact, we are not interested here in any sort of expression with these kinds of features, but in those whose polysemic potential and context sensitivity stem from their anaphoric capacity alone. This restriction is justified both from a conceptual and a practical, language engineering point of view. It is not only the source of the polysemic potential of an anaphoric expression such as it and a non-anaphoric, ambiguous one such as flying planes that differs, but there are also distinct factors, heuristics and technical difficulties correspondingly invoked to derive their semantic values which in themselves call for quite different solutions.

In general, the polysemic potential of ambiguous phrases is limited to a restricted set of different possible semantic values. Flying planes, for instance, is ambiguous between the two readings made evident in (1). The polysemic potential of anaphoric expressions, in turn, is not restricted to any finite set of possible semantic values, as illustrated in the examples in (3), which could be indefinitely replicated.

- (3) a. John pulled off the wheel. It was heavy.
b. Mary bought a computer. It has a CD-ROM drive.
c. Peter got a ticket for the show. It is for Saturday night.
b. Claire hit the donkey. It was gray.

On a par with this difference in the dimension of the polysemic potential of anaphoric and non-anaphoric expressions, there is a different role played by the context in delimitating the relevant semantic value. As for ambiguous expressions such as flying planes, the context seems to act like a sieve, fulfilling a sort of

selectional function by helping to fix one reading among all the possible readings. That seems to be what is going on in (1)a., as the short context *are complex machines* is enough to enforce the *aircraft* interpretation to the detriment of the *piloting* interpretation.

As for anaphoric expressions such as *it*, the context seems to have more of a constructive function rather than a selective one. Here the context does not help to contrive the selection of a semantic value from a predefined set of possible values. Instead, it is involved in providing material which the semantic value of the anaphoric expression can be identified with or built upon.

It is this different importance of the context in anaphoric and non-anaphoric classes of context-sensitive expressions that may explain the contrast between the intuitions about (4)a. and (4)b..

- (4) a. Flying planes may be dangerous.
 b. It was heavy.

In both examples, the context has been reduced so that its contribution to the semantic interpretation of the context-sensitive expressions is hampered. Consequently, we cannot assign a definite semantic value to *flying planes* or to *it*. Nevertheless, we can pinpoint what are the possible semantic values that *flying planes* may assume, so that the interpretive indefiniteness of (4)a. can be dissolved. We know that (4)a. can mean either that *planes moving through the air* may be dangerous or that *piloting planes* may be dangerous. But we cannot do so with regards to *it*. It is beyond our capacity as speakers of English to present a list of the type of situations that we are entitled to mean by using (4)b.. This seems thus to require an explanation in the light of the different contributions made by the context to interpreting of the context-sensitive expressions pointed out above. While in (4)a. the context is insufficient to help select one of the two potential readings, in (4)b. it is insufficient to help construct one.

Antecedents and cospecification

This brings us to a point where we can circumscribe in more detail the specificity of the semantic polyvalence of anaphoric expressions. The constructive contribution of context to the semantics of this kind of expression is typically rendered as the requisite that the interpretation of an anaphoric phrase α is dependent on the interpretation of another phrase (or phrases) θ , usually termed as

the antecedent(s) of α . Taking an illustrative examination of this idea by Sidner (1981, p.218), one reckons that “anaphor interpretation can be studied as a computational process that uses the [...] existing specification of a phrase to find the specification for an anaphor”.

The list in (5) presents examples where anaphoric expressions establish different modes of interpretive dependency as regards their antecedents.

- (5) a. John pulled off the wheel. It was heavy.
b. The secretary introduced Monica to Bill. A few minutes later they went off for dinner.
c. The President could not take his car. The tire was flat.
d. Fewer than twenty deputies voted against the proposal. They were afraid of riots in the streets.
e. Every sailor in the Bounty had a tattoo with his mother's name on the left shoulder.

In (5)a., it has the same semantic value as its antecedent, the wheel. In (5)b., they has two antecedents, and it refers to an entity comprising the two referents of the antecedents. In (5)c., the referent of the tire is part of the referent of its antecedent in the previous sentence, his car. In (5)d., they has a so-called non-referential antecedent, fewer than twenty deputies, from which a referent is inferred to serve as the semantic value of the plural pronoun: they refers to those deputies, who are fewer than twenty in number, and who voted against the proposal. Finally, in (5)e., although one also finds a quantificational antecedent for the anaphoric expression, the relation of semantic dependency differs to the one in previous example. His mother's name does not refer to the mother of the sailors of the Bounty. It acts rather in the way of a bound variable of logical languages. For each sailor s , his mother's name refers to the mother of s .

Contexts

In each of the above examples, we observe thus that the interpretation of the anaphoric expressions is anchored in the interpretation of other linguistic expressions. Given the classification of types of contexts presented above, these are examples where the relevant context is what could be termed as the linguistic context. Nevertheless, a similar behavior of semantic deficit needing to be

supplemented may be observed even in utterances where the linguistic context is not able to play such an anchoring role.

- (6) a. They went off for lunch.
 b. The tire is flat.
 c. Richard is not coming.

Considering each of the sentences in (6) uttered in isolation, the interpretation of the expressions *they*, *the tire* or *Richard* above cannot be fully recovered from any piece of linguistic material around. However, if the utterance contexts of the sentences are suitable for their felicitous use, an adequate interpretation may be assigned to these expressions. For instance, if we are stopped in the middle of a road and a policeman comes up to us, and we simply say “The tire is flat”, the context-dependent expression *the tire* receives a perfectly defined and understandable interpretation. Or, if the boss comes into our office, doesn’t stop staring without saying a word, and we say “They went off for lunch”, we can be sure that he would assign the correct semantic value to *they*. Or still, if we were supposed to bring Richard Smith with us to the party, when we utter “Richard is not coming” to our friends on arriving at the party, our friends will realize that Richard Smith, and not Richard Nixon, Richard Montague, or any other person in the world whose name is Richard, is not coming to the party.

The dependent interpretation behavior of so-called anaphoric expressions is thus a phenomenon which is an intrinsic feature of such expressions and is invariant across different types of context. This stability of semantic dependency across contexts has been accommodated in advanced semantic frameworks, such as Discourse Representation Theory (van Eijck and Kamp, 1997), fairly simply and elegantly. First, semantic representation structures of natural language expressions are not restricted to include just the representation of semantic material corresponding to those expressions. Semantic representations may also include material inferred from other semantic representations as well as semantic material representing entities available in the utterance context, but not necessarily expressed by any linguistic phrase. Second, the notion of antecedent is given a semantic twist. It is not the linguistic expression upon which an anaphoric expression is semantically dependent, as in a first approach we have been intuitively alluding to. Rather, it is the piece of semantic representation, usually called discourse referent or reference marker, with respect to whose semantic value the semantic value of the anaphoric expression is dependent. According to Asher and Wada (1988, p.310), “the process of anaphora resolution is a process that finds the appropriate, anaphoric antecedent discourse referent to be identified or otherwise related to a discourse referent introduced by an anaphoric [expression]”. Although this is a significant distinction not to be disregarded, for the sake of making it

clearer, we will use the term 'antecedent' to refer either to the syntactic antecedent or to the semantic antecedent when no confusion may arise.

Given this quite broad sense assigned to the notion of anaphora, it may turn out to be difficult to draw a clear demarcation line between non-deictic and deictic uses of anaphoric expressions.

Some authors, such as Levinson in his handbook *Pragmatics* (Levinson, 1983, p.69), classify anaphoric expressions as non-deictic, as he considers only the linguistic context to be the relevant one to recover the antecedent of an anaphoric expression. Nevertheless, while discussing his example (40) on p.68, repeated below in (7), Levinson recognizes that the boundaries between deixis and anaphora may not lend themselves to a clear-cut formulation. He notices that "it is perfectly possible for a deictic term to be used both anaphorically and deictically", as "there refers back to whatever place London refers to, but simultaneously contrasts with here on the deictic dimension of space, locating the utterance outside London".

(7) I was born in London and have lived there ever since.

Other authors, in turn, such as Sag and Hankamer (1987), and van Eijck and Kamp (1997), openly defend the view that no definitive demarcation can be drawn between anaphora and deixis.

In this dissertation, it is not central to our goals to settle the issue concerning the differences, if any, between deixis and anaphora. As will become apparent as we proceed, it will be enough for our purpose to adopt a position close to the one advocated in Discourse Representation Theory.

A further point to be made with respect to delimiting the object of study addressed in this dissertation is that we are not going to take into account all different classes of anaphoric dependencies. All three expressions *he*, *it* and *later* below present a typical anaphoric behavior in the sense that they require their semantic values to be constructed from elements in the context.

- (8) a. Mary asked John to get down but he was not ready.
b. Bill had a heated argument with his secretary. It was a very unpleasant occasion for all of us in the office.
c. Bill had a heated argument with his secretary. Later he apologized to her.

They differ, however, as to the type of their antecedents and, consequently, to the ontological type of the entities which turn out to be their semantic values. In (8)a., *he* has a nominal phrase as antecedent, and refers to an individual. In (8)b., *it* has the previous sentence as antecedent, and refers to an event. Finally, in (8)c., *later* refers to a stretch of time and probably has a tense affix as antecedent. Given the

specific mechanisms found in natural languages to handle the expression of these three different types of entities, these anaphoric expressions present problems which are to a significant degree conceptually autonomous and whose solutions may require different approaches (vd. Webber, 1988). This dissertation will focus on anaphoric expressions of the first type.

1.3 Reference Processing

The above introductory remarks on anaphora are intended to enable us comprehend the general configuration of a central phenomenon in the semantics of natural languages. But they should also allow us to grasp the cardinal relevance of anaphora for the automatic processing of natural languages. In fact, in all areas of language engineering requiring some representation of the meaning of natural language expressions, the task of building a semantic representation for an anaphoric expression α comes down to involve the tasks of (i) finding an antecedent θ for α , and (ii) establishing a convenient relationship between the semantic representations of both.

Anaphor resolution

It is worth noting that context sensitivity is not a problem in itself for language processing in general. In a certain sense, it could be said that perfectly well-shaped programming languages with local variables have expressions that, like what as a rule happens in natural languages, are somehow anaphorically sensitive to the context. In this case, the context is defined by the beginning and the end of the relevant block to which the variables belong, and their semantic values are determined by unequivocal assignment commands. The real problems involved in processing anaphoric expressions such as it or the screw arise rather from

circumstances which, in natural languages, happen to be specifically associated with the context sensitivity exhibited by these specific expressions.

On the one hand, contrary to programming languages, the boundaries of the relevant interpretive context are not explicitly stated and must be retrieved from a superficially clueless string of words. This point can easily be made evident by taking different occurrences of syntactically similar definite descriptions. For instance, the two occurrences of the screw in sentences 3. and 18. of the dialog in (2), and the two occurrences of the flywheel, namely in the first and last utterances of that dialog. While the flywheel is assigned a single semantic value throughout the dialog, the semantics of the screw, within the same stretch of linguistic production, is in turn consecutively realized by means of two different values.

On the other hand, even if it turned out that the eventual context boundaries could be resolved, there still remains the highly non-trivial task of determining the antecedent of the expression at stake, and consequently its semantic value. For the sake of illustrating this point, consider first the three consecutive occurrences of it between sentences 12. and 16.. Here all the occurrences come after the wheelpuller in 12., and it presents an invariant semantic value as it is being used to refer to the wheelpuller in all these occurrences. Now consider it again, but this time in its two first occurrences in the dialog, in sentences 3. and 10.. Although both occurrences likewise come after the utterance of a definite description, namely the screw in 3., it does not present an invariant semantic value. In its first occurrence, in 3., although it is occurring almost immediately after the screw, it takes an antecedent occurring before that expression, in sentence 1., and refers to the flywheel. In its second occurrence, in 10., it refers however to the screw, an entity firstly evoked by the antecedent in sentence 3..

The contrast between the behavior of it in these two examples conspicuously illustrates the magnitude of the problems involved in automatically determining the antecedents of anaphoric expressions in natural languages.

Automatic processing of reference

Given the pervasiveness and intriguing properties of anaphora, reference processing and anaphor resolution have become a central research topic in natural language processing. From a broad syntactic point of view, anaphor resolution can be described as the process by which, for a given anaphoric expression *a*, the antecedent, or antecedents, of *a* is located. This is just part of the broader procedure of reference processing that, after *a* has been resolved, determines the interpretation

of a by taking into account how it relates to the meaning of the antecedent(s) of a. Reference processing, in turn, forms part of the broader procedure of semantic processing, whose expected outcome is a semantic representation of the natural language segment where a occurs.

This semantic representation is a synonymous expression of that segment, usually of a specially designed formal language, which for the sake of automating inference and interfacing with knowledge representation, is typically an empirically motivated variant of a logical language. Therefore, and now from a semantic perspective, anaphor resolution can be described as the process which, for a given anaphoric expression a and the relevant term t in the semantic representation s of a, identifies the relevant term t' in the semantic representation of the antecedent of a to which t has to be related for the interpretation of a to be accomplished. Given the specific relation between the interpretation of a and the meaning of its antecedent, the reference processing module is thus designed to ensure that terms t and t' enter the appropriate relation (identity, inclusion, is an element of, etc.).

Given the recent and converging advances both in theoretical and practical research on the semantics of natural languages, reference processing appears as one of the core issues to be dealt with by semantic processing engines, on a par with other cardinal facets, as for instance the syntax-semantics mapping of argumental structure, the underspecification of readings, controlled disambiguation or assignment of relative scope of operators.

From a strictly practical point of view, reference processing modules are thus crucial components for the quality of most medium- to high-level applications in natural language processing. They are crucial in tasks where representation of the semantic information conveyed by natural language expressions is required. Reference processing systems, in general, and anaphor resolvers, in particular, therefore become relevant components in a series of natural language applications, ranging from dialogue handling to machine translation, including, among many others, information retrieval, document sorting or automatic summarization.

Linguistic- and cognitive-driven approaches

In the next chapter, we provide an overview of the major research lines on automatic reference processing, their fundamental tenets, and the important research issues under investigation. In what follows, we group work carried out on reference processing under two main research traditions. Even though the overall

field easily accommodates and stimulates cross-fertilization of ideas, we think that these traditions ask to be conceptually individuated once they ultimately rest on different core assumptions.

The two traditions result from different views on the properties of anaphora pointed out in the present chapter. On the one hand, given their polysemic potential, anaphors admit a flexible anchoring to antecedents in the sense that a given anaphor may be alternatively resolved against one out of different possible antecedents, as illustrated in (9).

(9) John said to Peter that George doesn't believe he will be elected president.

On the other hand, the constructive context sensitivity of anaphors is expected to contribute to a resolved interpretation. Even when there is more than one appropriate antecedent available in the context, a given anaphor is nevertheless assigned an interpretation and it is resolved against one of the possible antecedents. This property is illustrated in (9) and (10).

(10) John offered Peter a book and Mary bought him a tie.

Although in (10) *him* has two possible antecedents, John and Peter, it is not left uninterpreted and it is preferably understood as referring to Peter. And given that the context of (9) may not induce a preference for one of the antecedents, the isolated utterance of (9) would certainly cause the hearer to ask for clarification from the speaker so that an interpretation could be assigned to *he* and the sentence could be properly understood.

In Sections 2.1 and 2.2 we group together proposals to a large extent sharing the view that flexible anchoring and resolved interpretation are understood primarily as a matter of linguistic relations between linguistic items. Flexible anchoring is viewed as a relation between an anaphor and its potential antecedents, which are selected from the context on the basis of linguistic properties and collected in a set of antecedent candidates. Resolved interpretation, in turn, is seen as the relation between the anaphor and one particular element of that set of antecedent candidates, which is selected primarily on the basis of considerations building again mostly on linguistic properties.

In Section 2.3 we present a different research tradition, where flexible anchoring and resolved interpretation are understood primarily as a matter of cognitive relations between the subject and entities evoked by linguistic items. Flexible anchoring is viewed as a relation between the speaker/hearer, at the time he encounters an anaphor, and some items which he may retain in his working memory and which can serve as antecedents of the anaphor. Resolved interpretation, in turn, is considered to be a relation between the speaker/hearer and

the relevant items in his memory, a relation where one of these items is selected due to the fact that it is more salient to his attention.

1.4 Summary

After introducing the dissertation, this chapter went on to clarify the notion of anaphora and reference processing.

The phenomenon of anaphora was first shown to be an extreme example of the interplay of two notorious traits of natural languages, polysemy and context sensitivity. Anaphoric expressions may thus assume many different semantic values inasmuch as their interpretation is highly dependent on the information provided by the context in which they occur. Their context sensitivity is nevertheless specific in the sense that it is constructive rather than selective. In contrast to other ambiguous expressions, whose reading in a given occurrence is selected by contextual factors from a small finite set of determined readings, the semantic value of anaphoric expressions is constructed upon the semantic value of other expressions, their antecedents. This makes it possible for anaphors in their different occurrences to assume different values, not restricted to a small predetermined set.

Given its shape, anaphora turns out to be one of the crucial phenomena that automatic processing systems of natural language should handle. Systems intended for the semantic processing of natural languages have thus to address the issue of interpreting anaphoric expressions. This implies they should be able to resolve anaphors, i.e. that they are able to find the antecedents of anaphors so that the semantics of the latter may be automatically uncovered, enabling the larger expressions in which they occur to receive full interpretation.

2 Heuristics and Constraints

2.1	The Integrative Framework	40
	Filters, preferences and weights	41
	Selecting antecedent candidates	43
	Similarity filters	43
	Binding constraints and other relative positioning filters	45
	Electing antecedents	47
	Similarity preferences	47
	Relative positioning preferences	49
	Local topic preferences	50
	Discourse domain preferences	52
	Scoring the candidates	52
2.2	Recent Developments and Lines of Research	54
	Fine-tuning the scoring schemata	55
	Evidence from corpora	56
	Evaluation	58
	Robustness	62
	Generality	64
	Multilinguality	66
	Dialogue and multimodality	68
	Identification of referents	68

Heuristics and Constraints

2.3 Centering Attention for Anaphor Resolution	70
Attentional state, discourse structure and form of expressions	70
Algorithm for anaphor resolution	73
Attentional inertia	75
Discourse locality	77
One single pronoun and non-pronominal anaphors	78
Intra-sentential anaphora	79
Correlation discourse structure vs. form of expressions	80
Correlation between attention and use of pronouns	81
Implementation of BFP algorithm	84
Converging to an integrative approach	84
2.4 Summary	87

By way of historical perspective, in the very young history of research on reference processing, 1988 can certainly be taken as a major landmark in the development of the area. As a result of independent work, two six-page papers appeared, one at the COLING conference, by Jaime Carbonell and Ralf Brown, another at the conference on Applied Natural Language Processing, by Elaine Rich and Susann LuperFoy. A third one, of a journal-article length, by Nicholas Asher and Hajime Wada, appeared in the *Journal of Semantics*. The importance of these papers is attested by the fact that, to a very significant extent, subsequent work in the area mostly extended, refined or explored issues dealt with or just indicated therein. It is not unreasonable to propose that they laid down the foundations of a major research tradition in reference processing, insofar as they set down a set of core assumptions that have since permitted a cumulative growth of results in the field.

Those core assumptions are corollaries of a central idea which resulted from a simple but crucial epistemic displacement. That displacement has its origin in the observation that “whereas the ubiquity of the [anaphora] phenomenon [...] has long been recognized, no truly comprehensive computational approaches for anaphora resolution have been proposed” (Carbonell and Brown, 1988, p.96), or in the words of Rich and LuperFoy (1988, p.18), “...although there exist many theories of anaphora resolution, no one of these theories is complete.”

As it was a fairly new theme of interdisciplinary research, researchers trying to come up with a theory of anaphora and reference processing typically primarily employed the conceptual devices they could bring from their own original disciplines. Henceforth, while the multiple aspects relevant to understanding, explaining or replicating reference processing (morphological, syntactic, semantic, pragmatic, psychological, etc.) were becoming better known, it became progressively evident that the behavior of anaphors was probably not governed by a special-purpose closed set of rules, but that it was the result of the interaction of different sorts of constraints. Some of the relevant rules may be specific for the anaphora phenomenon, but many others hold for general aspects that anaphors

share with other types of expressions, since - as with other expressions, they productively enter different kinds of linguistic constructions and bear gender, number, grammatical function, etc.

2.1 The Integrative Framework

The core tenet of those seminal papers was that reductionistic approaches aiming to account for anaphora exclusively in terms of either syntactic, semantic or pragmatic factors should be discontinued in favor of an integrative approach where all relevant aspects could be taken into account. That is what one can read from the following excerpt from Rich and LuperFoy, 1988, p.18: “There exist many partial theories each of which accounts for a subset of the phenomena that influence the use and interpretation of pronominal anaphora. These partial theories range from purely syntactic ones (for example the simple rules of number and gender agreement) to highly semantic and pragmatic ones that account for focusing phenomena. If there were a single, complete theory, then it might be appropriate to implement it. But because there are partial theories but not a complete one, we have designed an architecture [...] that allows for a loosely coupled set of modules, each of which handles a subset of discourse phenomena by implementing a specific partial theory”. The same point can be found in Carbonell and Brown, 1988, pp.96-97: “We explore the central hypothesis that anaphora resolution may be best accomplished through the combination of a set of strategies, rather by a single monolithic method [as] anaphor resolution is not a monolithic autonomous process; it requires access and integration of all the knowledge sources necessary for dialog and text interpretation”; and in Asher and Wada, 1988, p.309: “This research [on anaphora] has led to a host of constraints within different paradigms on the anaphoric process. We present a semantic framework that integrates these different constraints – syntactic, semantic, pragmatic and even “stylistic” – into a unified model of anaphora resolution”.

Filters, preferences and weights

Given this basic design decision for an integrative method, some consequences follow as corollaries. First, there is the acknowledgment that not all factors or principles encoding linguistic knowledge found out to be useful for anaphor resolution by different “partial theories”, as Rich and LuperFoy called them, have the same impact in the resolution process. Some of those rules are eliminative, in the sense that items that do not comply are discarded from the set of admissible antecedents. Some others are not eliminative, as they do not exclude items from the set of antecedent candidates, but they are still important because their observance enhances the chances for a candidate to be chosen as the antecedent of the anaphor at stake: “salience [of antecedent candidates] is the result of adding up the various preferences ... Features like recency, parallelism, reiteration, and the like each add something to the overall salience ranking, but none alone is decisive” (Asher and Wada, 1988, p.333).

This suggests that a partition of the factors should be taken into account by anaphor resolvers. The former, contributing to the selection of antecedent candidates, are usually termed in the literature as constraints or filters. The latter, contributing to discrimination among the elements of the set of eligible antecedent candidates, allowing one to be picked up as the antecedent, are usually known as preferences or heuristics.

This partitioning of the set of factors involved in the process of anaphor resolution leads to a second important corollary of the option for an integrative method. For the sake of a principled choice of the antecedent for a given anaphor, preferences have to be put to use to help ranking the eligible candidates. Compliance with preferences should be seen as enhancing the chances an antecedent candidate has to be determined as antecedent. The observance of a given preference will contribute, with a certain value, to the total score of plausibility of a given antecedent candidate to be the elected antecedent. As not all preferences have the same impact for that plausibility, different preferences should thus be expected to contribute with different weights.

On a par with the partitioning between constraints and preferences and the weighting of preferences, a third important corollary of adopting an integrative method has to do with the place of anaphor resolution devices in the overall architecture of natural language processing systems. As constraints and preferences require a considerable amount of grammatical information to be successfully activated, many of them, though not necessarily all as we will discuss in the next chapters, apply as stressed by Carbonell and Brown (1988), “post facto”. This means that an anaphor resolver should be expected to act typically as a

post-parsing module, when previous parsing modules have already elicited enough linguistic information from the input string.

Given that parsing modules, the outcome of the parsed string and the overall architecture of different Natural Language Processing systems show great variation and difference, the implementation of an integrative anaphor resolver is expected to show considerable difference from system to system. Even though, under these circumstances, no model of an anaphor resolver more specific than the methodological lines sketched above is expected or useful, some proposals have been made to at least give some conceptual organization.

That is the case of Rich and LuperFoy, 1988, where the anaphor resolver is designed with an architecture patterned in line with the idea of a blackboard system (cf. Erman et al., 1981). With such an architecture, each relevant rule or principle is modeled as a “constraint source” which takes relevant linguistic information from the central “handler” and returns to it information relevant to the location of the antecedent. The “handler”, in turn, is responsible for putting together the information received from the different “constraint sources” and deciding which candidate to chose as antecedent. More recently, an interesting development in the idea of conceiving an anaphor resolver according to a blackboard architecture was presented in Paraboni and Lima, 1998. There, the “constraint sources” are conceived as reflexive agents, which are understood as rule-based entities acting according to the perceived environment, following Russel and Norvig, 1995.

Though the gist of the integrative approach originates in the papers mentioned above, with a first non-negligible approximation to it by the definition of “context factor” in Alshawi, 1987, the list of its foundational landmarks would be somehow incomplete if we did not add to it the 25 page paper by Shalom Lappin and Herbert Leass which appeared in *Computational Linguistics* in 1994.

While the first papers from 1988 were crucial to set up the rationale for the integrative method and its basic assumptions about the distinction between constraints vs. preferences, and about preferences weighting, only in the 1994 paper by Lappin and Leass can we find a thorough discussion and report on the implementation of a resolver organized according to that method. This paper was important not only for showing the practical feasibility of the approach, but also because it presented an objective evaluation of its efficiency. The authors reported that the system they implemented resolved 86% of the anaphors correctly when run on a test set of 345 sentences randomly selected from a corpus of 48 computer manuals containing 1.25 million words.

This was a telling figure given that the most efficient algorithm known to date, by Hobbs (to be presented in detail below), applied manually over the same test set, delivered 82% of anaphors correctly resolved (cf. Lappin and Leass, 1994, p.556).

Selecting antecedent candidates

Relevant filters determining the eligibility of antecedent candidates have been repeatedly highlighted since 1988 in several different works, such as Kasper, Moens and Zeevat, 1991, Lappin and McCord, 1990a, Lappin and Leass, 1994 or Mitkov, 1997b. They lend themselves to grouping into two classes. One of the classes includes filters requiring some kind of similarity between the anaphor and the eligible antecedent candidate. The other class includes filters on the relative positioning both of the anaphor and the possible antecedent in the geometry of grammatical representation. The former are well known, and their implementation is quite straightforward. The latter, however, present a non-trivial complexity and have been one of the most intriguing and rapidly evolving research issues in natural language processing, in general, and reference processing, in particular.

In this section we are going to present a brief introduction to both classes. We will return to a more detailed presentation and inquiry into the latter in the next chapters, as they are the linking thread of the results reported throughout this dissertation.

Similarity filters

A first similarity filter concerns correspondence of morphological features of Person, Number and Gender between anaphor and antecedent, usually taken as one of the basic types of agreement (vd. Pollard and Sag, 1994, p.61). Evidence of this constraint is presented in (1) below. While in a. the subject of the first sentence can be the antecedent of she, the same does not hold in b., the reason being that John and she bear different Gender values.

- (1) a. Susan entered the room. She started playing piano.
b. John entered the room. She started playing piano.
c. John and Susan entered the room. They started playing piano.

In some circumstances, namely with plural anaphors, the satisfaction of this constraint may involve a bit more than just matching the relevant values of the anaphor and the antecedent. Such kind of circumstance is illustrated in (1)c., with the interpretation where they refer to John and Susan. It exemplifies a case of so

called split anaphora, where the anaphor has more than one antecedent. In such cases, each of the antecedents may not bear all the morphological feature values of the anaphor. In Portuguese, for instance, if the anaphor is plural, each one of the antecedents may not be plural. If the anaphor is plural masculine, at least one of the antecedents is masculine; and if the anaphor is plural first or second person, each antecedent may not be first or second person.

Also, in some other languages, such as English, the strict morphological features may not be the only issue for the agreement between anaphors and their antecedents.

- (2) a. The committee decided that they would meet on Fridays.
b. Memory holds the data after it is input to the system but before it is processed.

As illustrated in (2)a. with the anaphoric link between they and the committee, a plural anaphor may have a singular antecedent when it is a collective head noun. The inverse case is also possible and is illustrated in (2)b., which repeats an example from Mitkov, 1997b, p.18.

A second similarity filter concerns the sharing of semantic type, understood as the information required in the general process of semantic selection of an argument by a predicator. The existence of this constraint can be made evident by sets of examples such as the following:

- (3) a. Insert the diskette in the computer's drive A and switch it on.
b. Insert the diskette in the computer's drive A and copy it to the hard disk.

Though the context preceding it in (3)a. and (3)b. is the same, if the anaphor it is to be resolved against a possible antecedent in that context, it has to receive different solutions in each of the examples. In (3)a., it cannot be taken as referring to the diskette, while in (3)b. it cannot be interpreted as referring to the computer. The reason for the pronoun having only as antecedent the computer in (3)a., and the diskette in (3)b., is to be found in the fact that it is the complement of switch on in the first, and the complement of copy in the latter. This implies that in the first the pronoun has a semantic type, induced by the semantic selection by the predicator switch on, that matches the semantic type of the computer but not the semantic type of the diskette, as one can switch on an object, such as a computer, but not a diskette. The same reasoning gives us, *mutatis mutandis*, the explanation for the facts observed in (3)b..

Binding constraints and other relative positioning filters

Turning now to binding constraints, we will provide here a very preliminary introduction to their basic characteristics, as they will be discussed at length in the remainder of the dissertation.

Binding constraints have been presented as the most notorious filters on the relative positioning in the grammatical geometry between anaphors and corresponding antecedents. The intriguing trait of anaphora illustrated by the data in (4) is that, though occurring in the same context, different types of anaphoric expressions admit different sets of antecedent candidates. In each example there are three phrases that can be selected to enter the set of antecedent candidates, namely John, John's elder brother, and Bill. However, in each example not all phrases are selected as antecedent candidates in spite of the fact that, for every phrase, the two previously mentioned similarity filters are complied with.

- (4) a. John's elder brother said that Bill hit himself.
b. John's elder brother said that Bill hit him.
c. John's elder brother said that Bill hit the little boy.

In (4)a. himself can only take Bill as antecedent. In (4)b. the set of antecedent candidates for him does not include Bill, but includes two other expressions, viz. John's elder brother and John. And in (4)c., the definite description the little boy admits only John as possible antecedent out of the three expressions available in the linguistic context.

Contrasts like those depicted in (4) have proven very robust in the sense that they have been observed and replicated in many different natural languages from different language families. This has given support to the hypothesis that constraints of this type, concerning the relative positioning of anaphor and antecedent in the grammatical geometry, are a universal, hard-wired component of linguistic knowledge and processing.

It is worth noting that binding constraints have been uncovered basically at the level of intra-sentential anaphora, that is anaphora where the anaphor and its antecedent belong to the same sentence. Very recent research, however, as we will discuss in detail in chapter 8 Conclusions: Summary and Outlook, has raised the hypothesis – yet waiting for full exploitation of its promise – that these constraints may also hold at the inter-sentential level.

Another relative positioning filter that has deserved some debate, especially in the formal semantics literature, was observed in the context of the research on so-called “donkey” anaphora. Though to a certain extent this is also a constraint on relative positioning of anaphors and their antecedents, it has some crucial differences as regards the binding constraints mentioned above. First, it has a much more restricted character as it concerns only anaphors with indefinite antecedents. Second, it concerns only inter-sentential anaphoric links, not intra-sentential ones. Third, it does not seem to be as empirically robust as the previous ones given the diverging results shown by repeated empirical testing.

A nice presentation of this filter can be found in Chierchia, 1995, pp.7-10, from where the examples below were taken.

- (5) a. John has a donkey. It is not happy
- b. # John does not have a donkey. It is not happy.
- c. # Every farmer who has a donkey beats it. It is not happy.
- d. # If John has a donkey, he beats it. It is not happy.

- (6) a. Either Morrill Hall does not have a bathroom or it is in a funny place.
- b. Every chess set comes with a spare pawn. It is taped under the box.
- c. It is not the case that John does not have a car. It is parked outside.

The contrast between (5)a. and the other examples in (5) have been taken as lending support to the hypothesis that when an indefinite is in the scope of negation, of universal quantification or of the antecedent of a conditional, it cannot serve as the antecedent of a pronoun appearing in subsequent sentences. In fact, while in (5)a. the pronoun *it* in the second sentence can have the indefinite *a donkey* as antecedent, that is not the case in the other examples in (5), where the indefinite occurs in the scope of those types of contexts.

Data like (6), however, have led researchers to cast doubts on the nature and robustness of the constraint involved in the impossibility of the anaphoric link between the pronoun and the indefinite in examples like (5)b-d.. The examples in (6) seem to be, from a syntactic point of view, “structurally isomorphic”, as Chierchia (1995, p.9) put it, with respect to their counterparts in (5). Nevertheless, the relevant anaphoric links here are perfectly acceptable.

Contrasts such as those between (5) and (6) led Dowty (1994, p.116) to hypothesize that the relevant distinction between the two sets of examples should be found not in syntax but in semantics. In particular, he suggested that the relevant factor at stake may be the property of downwards monotonicity. Following that suggestion, the indefinite being unable to serve as antecedent of a pronoun in a subsequent

sentence, as in (5)b.-d., should be correlated to the fact that it is in the scope of a downwards monotone context.

Electing antecedents

Although establishing the relevant preferences for electing the antecedent out of the set of antecedent candidates is a crucial issue in anaphor resolution, there are not many papers espousing the integrative method that clearly indicates the preferences adopted. From those, only a few discuss and justify each of the preferences found to be useful. Given this, the primary and more comprehensive sources of information about preferences for anaphor resolution are again Carbonell and Brown, 1988, Rich and LuperFoy, 1988, Asher and Wada, 1988, and Lappin and Leass, 1994, to which we should add the more recent Mitkov, 1997b, 1998a.

As with filters, preferences can be grouped into different classes according to their conceptual proximity. An initial class of preferences has to do with similarity between anaphor and antecedent with regard to some kind of linguistic information they exhibit.

Similarity preferences

A first similarity preference, known in the literature at least since Akmajian and Jackendoff, 1970, as grammatical function parallelism effect, is illustrated by the following examples.

- (7) a. Mary gave a diskette to Susan and then John gave her a letter.
b. Mary gave a diskette to Susan and then she gave John a letter.

Both in (7)a. and b., the pronoun in the second sentence can be interpreted as having as antecedent Mary or Susan, both from the previous sentence. However, there is an observed tendency to prefer, respectively, Susan and Mary, as antecedent by default, respectively, in (7)a. and b., if no other contextual information coerces the option for the alternative antecedent candidate. As everything else remains identical in the two examples and only the grammatical function of pronouns varies, this

difference seems to be due to the fact that in such contexts, a pronoun is preferably resolved against an antecedent with identical grammatical function.

Psycholinguistic experiments were undertaken, among others, by Smyth and Chambers (1996) to check this preference. The authors reported that subjects who were submitted to a judgment task experiment demonstrated *inter alia* “a tendency to interpret object pronouns as coreferential with object antecedents, even when the subject of the preceding clause had been topicalized in the immediately preceding discourse” (p. 273). In a second experiment with self-paced reading, the fact that “reading times for sentences with pronouns were speeded in the parallelism condition relative to the nonparallelism condition” (p. 275) bears as additional empirical evidence for the preference for identity of grammatical function between anaphor and antecedent at least in contexts such as those depicted in (7).

A stricter variant of this preference is proposed in Dagan and Itai (1990) and Mitkov, Belguith and Stys (1998). These authors assume that there is a preference for antecedent candidates that have the same grammatical function of the anaphor, and moreover, are complements of a verb of the same syntactic-semantic class as the verb that takes the anaphor as complement.

A second similarity preference is also a parallelism preference. But here what tends to be preferred is not similarity of grammatical function but similarity of semantic function (or thematic role, in other terminologies) between anaphor and antecedent. This preference can be observed from the data in (8), taken from Carbonell and Brown, 1988, p.97.

- (8) a. Mary drove from the park to the club. Peter went there too.
b. Mary drove from the park to the club. Peter left there too.

There is an anaphor here as well, *viz.* *there*, which accepts two antecedent candidates from the previous sentence, *viz.* *the park* and *the club*. And again there is a difference between the two examples (8)a. and b. concerning the preferred candidate. In (8)a., *there* is preferably resolved as *the club*, and the second sentence is thus preferably interpreted as conveying the information that Peter also went to the club. In (8)b., however, *there* is preferably resolved as *the park*, and the second sentence is thus preferably interpreted as conveying the information that Peter also left the park. As the two contexts are identical except with respect to the semantic function of *there* – in (8)a. it has a Destination function, while in (8)b. it has an Origin function – the relevant preference is apparently found in that difference. Data such as (8) can thus be taken as evidence that, at least in this type of contexts, there is a semantic role parallelism preference which, in the absence of other coercing information, resolves an anaphor against an antecedent candidate with identical semantic function.

Semantic similarity is the basis of yet a third similarity preference. As explored in Gaizauskas et al., 1996, this preference requires the use of a hierarchical ontology whose nodes correspond to semantic classes to which anaphors and antecedent candidates belong. Ontological similarity between an anaphor and its potential antecedent is then calculated using the inverse of the length of the path, measured in terms of number of nodes, between the items being compared. The less the “semantic” distance and greater the semantic similarity of the antecedent candidate to the anaphor, the higher the preference for it being chosen as antecedent.

Finally, a fourth preference which may also be included in the class of similarity preferences is concerned with what Mitkov (1997b) loosely terms as lexical reiteration. In simple terms, this preference takes into account that the repetition of the head noun of an anaphoric definite description by the antecedent can be taken as a clue to resolve the anaphoric link at stake. Although this is a very simplistic way of articulating the complex details of the anaphoric behavior of definite descriptions (vd. among many others Garnham, 1989, and O’Brien et al., 1997), it can still be of practical importance, once the antecedents of definite descriptions are far from being restricted to those with a head Noun identical to the head Noun of the anaphoric description.

We will return to this point below when dealing with recent developments aiming to enlarge the coverage and generality of the integrative method for anaphor resolution, especially regarding anaphoric definite descriptions.

Relative positioning preferences

As with the second class of filters discussed above, a second class of preferences includes a preference concerning the relative positioning of anaphor and antecedent. Although this preference is frequently pointed out as one of the most important, if not the most important, preference for anaphor resolution (e.g. Lappin and Leass, 1994, p.55), its particulars are incomparably less complex than the specifics of the filters concerning relative positioning.

This preference on relative positioning, also referred to in the literature as recency preference, accommodates the simple observation – corroborated by psycholinguistic experiments (vd. among others, O’Brien et al., 1990) – that anaphors are preferably interpreted against antecedents that occur in preceding sentences that are sequentially closer in the discourse to the sentence where the anaphor occurs. The following example is a discourse where that relative positioning preference can be detected.

- (9) Mary saw John in the supermarket. She also met Peter. He was wearing a blue coat.

The discourse in (9) is a sequence of three sentences in which the last one has the pronoun *he*, which can be resolved, resorting only to that context, against either John or Peter. The preferred antecedent, however, is Peter. The significant difference between John and Peter seems thus to be that the first occurs in the first of the three sentences and the later in the second sentence. This apparently shows that the preference at stake selects the candidate occurring closer to the sentence in which the anaphoric expression occurs.

In Lappin and Leass, 1994, p.544, this preference is taken to its extreme. These authors propose another relative positioning preference requiring the intra-sentential antecedent candidates, i.e. candidates supporting an anaphoric link which does not cross sentence boundaries, to be explicitly preferred to inter-sentential ones.

Local topic preferences

A third class of preferences includes what could be called, for want of a better term, topic preferences. Under these preferences, the antecedent is given prominence because it is the candidate referring to the topic or theme to which the sentence supposedly aims to add information.

The topic of a sentence may be marked by means of specific syntactic constructions, such as cleft constructions, topicalizations, etc. If there is no special purpose syntactic device being used to mark the topic of the sentence, then in configurational languages such as English or Portuguese, the phrase with the less oblique grammatical functions, typically the subject, is usually taken to be the topic of the sentence.

Examples are given in Carbonell and Brown, 1988, p.98, and repeated in (10), where a topic preference can be observed.

- (10) a. It was Mary who told Jane to go to New York. Why did she do it?
b. It was Jane who went to New York at Mary's bidding. Why did she do it?

In each of the two sentence discourses of (10), the pronoun *she* occurring in the interrogative has two eligible candidates, Mary and Jane, made available by the first sentence. However, in (10)a. *she* is preferably interpreted as having Mary as antecedent, while in (10)b. *she* is preferably interpreted as having Jane as

antecedent. Given the syntactic structure of the sentences, this difference seems to originate in the fact that Mary is the marked topic in (10)a., while the marked topic in (10)b. is Jane.

In some papers, such as Strube and Hahn, 1996 and Strube, 1998, we find a simplified version of this topic preference. There, potential antecedents are given prominence simply on the basis of their part of speech classification. Typically, proper nouns are preferred as antecedent candidates over definite descriptions, which in turn are preferred over pronouns.

Another variant of the topic preference is the one based on the hierarchy of grammatical roles. This preference is illustrated by examples such as the following.

(11) John was introducing Peter to Bill, when Mary called him.

Although the main clause makes available three eligible antecedent candidates for him, there seems to be a natural tendency for the pronoun to be preferably interpreted against John, which is the subject, and not against Peter or Bill, which have more oblique grammatical functions.

Sometimes, this preference may not be immediately recognizable given its formulation. This is the case in Mitkov, 1997a, where the preference for antecedents that are “non-prepositional” covers in part the effects of the hierarchy of grammatical functions preference. This is so because more oblique functions in languages such as English are marked by prepositions.

Another point worth noting concerns the possible divergent effects of this preference based in the hierarchy of grammatical functions, discussed here, and the parallelism of grammatical function preference discussed above in the class of similarity preferences. For an anaphor with object function that has to be resolved either against a subject or an object, the first preference will select the subject while the latter will select the object. This is a very interesting case of the interaction of preferences where fine tuning of the conditions of their application seems to be asking for further research. It also clearly illustrates the need for an algorithm to combine all the preferences to work together towards the goal of discriminating the elements of the set of antecedent candidates, a point which we will address below in detail.

Discourse domain preferences

Coming now to a last class of preferences, we find a set of preferences that may be termed as discourse domain preferences. These preferences try to accommodate the fact that, given a specific domain of discourse in which an anaphor occurs, e.g. aeronautics, finance, sports, etc., some concepts are more likely to be the subject matter which the discourse is about. Accordingly, phrases expressing, or related to phrases that express those concepts are expected to be good antecedent candidates in the anaphoric links occurring in that discourse. To a certain extent, given that those phrases are expected to occur more frequently, these preferences can also be seen as probability of occurrence preferences, in the absence of objective measures of the probability of occurrence of phrases in a discourse.

In Mitkov, 1997b, three discourse domain preferences are proposed. The “domain concepts” preference gives prominence to the NPs expressing domain concepts. A second preference gives prominence to the objects of the class of selected verbs expressing typical relations of the domain discourse. And the third preference concerns expressions occurring in headings: “if an NP occurs in the head of the section, part of which is the current sentence, then consider it as the candidate likeliest to be the antecedent” (p.17).

Scoring the candidates

Having gathered the major preferences for discriminating antecedent candidates pointed out in the literature, we turn now to the issue of how to coordinate the multiple contributions of the different preferences.

Coordination of preferences aims to assign to each preference its relative importance in the global pool of preferences, in a way that their conjoined effects may model as closely as possible the human selection of an antecedent out of the set of antecedent candidates. Typically, this is done by establishing a scale and assigning values from that scale to each preference. The difference in the value assigned to each preference is thus intended to reflect the different impact of the preferences in choosing the antecedent. On the other hand, an antecedent candidate gets a score which reflects “how much it is preferred”. For each candidate this is done by using the value of every preference that applies to it in order to calculate its score. Different scores for different candidates are then used to discriminate among candidates and are expected to reflect the different status of the candidates in

terms of their suitability to be selected as the antecedent out of the set of candidate antecedents.

Given this setup, the endeavor of putting preferences to good use then comes down to the simpler, though highly interrelated, tasks of: (i) establishing a scale of evaluation for the relative importance of preferences; (ii) deciding on the values from that scale to be assigned to each preference; and (iii) coming up with an algorithm to calculate the score for each candidate using the values assigned to the preferences applying to it.

In order to get an idea of the type of answers these questions have received in the literature, we can look at some illustrative examples. Focusing first on the scale of evaluation and the weights assigned to preferences, we notice, for instance, that Asher and Wada (1988) adopted a scale using the numbers in the range from 0 to 4, while the scale of Rich and LuperFoy (1988) ranges from -5 to +5 (though no information was given here about the specific value assigned to each preference).

In Lappin and Leass, 1994, a different proposal is put forward. They adopt a scale from 0 to 100, and a specific preference, for example, antecedent candidate recency, receives the maximum value of 100, while the preference concerning the hierarchy of grammatical functions, as another example, receives 80 when applying to a subject candidate.

An example of a different scale can also be found in Mitkov, 1998a. Preferences are weighted there according to one of four values, -1, 0, 1 or 2. Sentence recency preference, for instance, contributes with, respectively, 2, 1 and 0, for candidates situated, respectively, 1, 2 and 3 sentences back from the sentence containing the anaphor.

Turning now to the algorithm for obtaining the aggregate score of each candidate, the first reference goes to the simplest one, found in Asher and Wada, 1988 and Mitkov, 1998a. Each candidate gets a total aggregate score resulting from the simple sum of the values given to the preferences applying to it.

In Lappin and Leass, 1994, addition is also used as the basis of the scoring algorithm, though a degrading factor is also taken into account. The aggregate score of an antecedent candidate is progressively degraded, depending on how far back the sentence where it occurs is with respect to the sentence of the anaphor. After the aggregate score has been computed, there is still a weight threshold filter which eliminates candidates scoring below a given value. The idea of a filtering threshold is also used in Mitkov, 1995.

Rico Pérez (1994) introduces some more sophistication to the procedure for using preferences to select a candidate. Each antecedent candidate is associated with a vector whose coordinates bear the values assigned to the candidates by the different preferences. The anaphor is also associated with a vector of the relevant feature

space. Comparison between the anaphor and each candidate is done via the scalar product of the corresponding vectors, taken as a measure of distance between vectors. The result is a list of candidates ordered according to the closeness of their vector to the vector of the anaphor in the feature space. The first ranked candidate(s) is(are) taken as the antecedent(s). An approach based on vectors representing candidates in a feature space is also adopted in Connolly et al., 1995.

Another type of sophistication can be found in Rich and LuperFoy 1988. A given preference contributes to the aggregate score of a candidate, not only with a value but also with a measure of the confidence of the evaluation it performs. The final scoring of n preferences for discriminating among candidates is obtained by the function

$$\text{final score} = \frac{\sum_{i=1}^n \text{score}(i) \times \text{confidence}(i)}{\sum_{i=1}^n \text{confidence}(i)},$$

which computes an average weighted not by the number of distinct scores, but by the total confidence expressed for the scores. In this respect, it is worth noting that Rich and LuperFoy do not make any distinction between constraints and preferences. In their approach there are only preferences, in that some of the preferences have weights and confidence values that make them have the filtering behavior of what other authors define as filters.

Mitkov (1997b) reports on a comparative study of these two approaches, i.e. constraints plus preferences vs. just preferences. The author concluded that the use of constraints as preliminary filters before the application of the scoring algorithm involving preferences “could make the procedure faster and more accurate” (p.20).

2.2 Recent Developments and Lines of Research

Given the plurality of approaches exemplified above, the question which naturally arises is thus how to opt between them. Maybe a more embracing formulation would rather be how to weight the effective contribution of preferences to

discriminating the antecedent candidates. This undoubtedly hinges upon the general issue of evaluation and comparison of anaphor resolvers, which we will address below. But it is also in its own right a highly challenging research issue in anaphor resolution, yet to be thoroughly addressed, as authors tend to devote just a few words if any in their papers on why they opted for that specific scoring scheme .

Fine-tuning the scoring schemata

Two first attempts to tackle this problem as a research issue can be found in Lappin and Leass, 1994, and Gaizauskas and Humphreys, *forth.*. Their simple experiments, though far from exhaustive investigations into the architecture of scoring schemes, represent a first step in the direction of fine-tuning anaphor resolvers. Basically, they experimented with deactivating each of the preferences and then compared the performance of their weakened resolvers with that of the original resolvers with all preferences active. As expected, they found that the deactivation of certain preferences has a more drastic effect in the decrease of the performance than others. The more crucial factor was found to be the recency preference.

Further work shedding light on how to optimize the scoring scheme was presented in Mitkov, Belguith and Stys, 1998. These authors define what they call the “discriminative power” of a preference P as “the ratio number of successful antecedent identifications when P was applied/number of applications of P”. They find that the preference with more discriminative power (given their scoring scheme and the kind of text they were using to run the resolver) is the one concerned with parallelism of grammatical function in coordination contexts (100%). Their recency preference appeared in last position (34.4%).

They also measured the “frequency of use” of preferences, defined as “number of applications/number of anaphors”. The most frequent preference turned out to be, as expected, antecedent candidate recency (98.9%), and the least used was the preference on parallelism of grammatical function in its stricter version, where identity of verb type is also considered (11.1%).

However informative these figures may be, they were not used by the authors in any reported attempt to optimize the scoring scheme and performance of their resolver.

Evidence from corpora

Other paths of inquiry have been pursued with the aim of improving the empirical support of the scoring scheme.

In this connection, it is worth mentioning the exemplary work by Dagan and Itai (1990), where statistical methods were used with the aim of improving the performance of anaphor resolvers. The rationale behind their approach is as follows. Frequency of cooccurrence of verbs and their complements in a corpus may be taken as reflecting semantic selectional constraints of the predicators as regards their arguments. Selectional constraints can be used to discriminate among antecedent candidates. When a given anaphor, which is the complement with grammatical function GF of the verb V, has more than one eligible antecedent candidate, it chooses as antecedent the candidate that is more “strongly” selected by V in the same circumstance of the anaphor, i.e. the one which more frequently occurs as GF of V in the corpus.

With this methodology set up, an experiment was performed to check their utility. First, frequency figures concerning co-occurrence of pairs verb V plus head noun of a given complement GF of V were collected. Second, the authors focused on resolving ambiguous anaphoric occurrences of the pronoun *it*, i.e. in occurrences where the pronoun had more than one eligible antecedent candidate. Using a corpus of 28 million words and a threshold of 5 occurrences of V+noun in corpus as the significant number of times for a phrase to be admitted as antecedent candidate, they found that only in 64% of those cases of ambiguous pronouns in the antecedent candidates set contained at least one candidate passing the threshold. They also found that only for 55% of ambiguous pronouns were their antecedents correctly selected using this method.

Building on these results and on the subsequent study reported in previous versions of Dagan et al., 1995, Lappin and Leass (1994) discussed in detail on how to coordinate this corpus-based methodology with the integrative method underlying their anaphor resolver. They concluded that “statistically measured [selectional patterns] alone provide a far less efficient basis for anaphora resolution than an algorithm based on [preferences]” (p.554). They also conclude that incorporating statistical information in a preference-based scoring method “can yield a modest improvement in performance” of the resolver, given the way they envisaged intertwining the two approaches. This was done by using the statistics-based procedure to confirm or invert the ranking of antecedent candidates proposed by the method based on preferences. For each pair (a_1 , a_2) of antecedent candidates for a given anaphor which is a complement GF of a verb V, where a_1 has a higher aggregate score than a_2 , their relative ranking inverts if the frequency of

occurrence of a_2 as GF of V exceeds the corresponding frequency of a_1 by a “parametrical specified value” (p.553).

Another very promising attempt to improve the empirical support of the scoring scheme, also using statistical methods, can be found in the more recent Ge et al., 1998.

The methodology followed here is also integrative in its basic setup. For each pronoun, a set of antecedent candidates is collected, which are ranked taking into account four factors: Similarity of morphological features and semantic type (filters) and “recency of mention” and “mention count” (relative positioning preferences). What is innovative is that for a given antecedent candidate, the weights assigned according to each of these factors are rendered in terms of the probability that that candidate has to be picked as the antecedent when considered from the perspective of the relevant factor. The four probabilities are then multiplied together and the antecedent selected is the candidate with the highest combined probability.

It would be beyond the scope of this dissertation to examine all the details of how the authors propose to obtain the probabilities for each different factor and different candidate. Taking the recency of mention preference as an illustration, it can be stated that obtaining the corresponding probability for a given candidate requires first determining the distance of the candidate from the corresponding pronoun (according to a certain metric taking into account the syntactic structure along the lines of Hobbs’ algorithm). As the candidate is in the i th position in its candidate set, the annotated corpus is searched to determine the number C_i of candidates in the i th position in their candidate sets that are correct antecedents of their corresponding pronouns. The corpus is also used to determine the number A of all correct antecedents it contains. Finally the relevant probability is given by $\frac{C_i}{A}$.

The authors report that in their experiment they obtained 84.2% accuracy for their pronoun resolution system in a corpus of 93,931 words, 3975 sentences, and 2477 pronouns, manually tagged with referent indices and the numbers for repetition of antecedents, where 10% of the corpus was reserved for testing and the remaining 90% for training.

Works like the ones described above highlight the increasing importance of using corpora to support research and development in the field of natural language processing, in general, and anaphor resolution, in particular. The use of corpora presupposes the existence of annotation tags and procedures, whose definition, however, in the case of anaphora has proven no simple task. Even though rough and simple annotation schemata have been adopted in a first practical approximation, such as those used in UCREL, 1998, or in ARPA, 1996, a thorough and comprehensive annotation of anaphoric links raises a series of yet to be answered questions both in

theoretical and practical terms, as reported in the pioneering studies by Rocha (1998b), and Poesio and Vieira (forth.).

Evaluation

The endeavor to use corpora to extract information for improving anaphor resolvers is intimately related to the issue of evaluating and comparing their performance. Without a commonly agreed set of documents over which to run resolvers under evaluation and a standard annotation scheme for anaphoric links in the documents, that desideratum lacks its essential basis to be further pursued.

A first attempt to originate the creation of such a canonical measuring procedure was made in the context of the Sixth Message Understanding Conference, held in 1995 (vd. ARPA, 1996, and Grishman and Sundheim, 1996). One of the aims of this Conference was to compare the performance of different systems in what was called the Coreference Task. For this purpose, the same corpus of documents was used by all the systems, which were then evaluated by an evaluation algorithm (cf. Vilain et al., 1995). Systems functionality was restricted to the resolution of coreferential anaphoric links.

The measures aimed at by the evaluation algorithm are variants of the standard Recall and Precision measures used in the evaluation of information systems (see, e.g., Salton, 1989). Recall is a measure of how many of what a system should find it actually found. Precision is a measure of how many of what the system found it was meant to find. For example, if in a given task there were 1,000 items to be retrieved and the system retrieved 750 of which only 500 are correct, the system performance is measured with a Recall of 50% ($500/1,000$) and a Precision of 66.6% ($500/750$).

The need for a variant of these measures for evaluating anaphor resolution hinges upon the fact that it is not just anaphoric links between two expressions that need to be measured, but possibly chains of anaphoric links between several coreferential expressions. In fact, if there is a chain in the text comprising links X-Y and Y-Z, but the system “discovers” links X-Z and Y-Z, it should be evaluated as having the correct chain. Furthermore, if the system does not respond with enough information to identify the whole chain, for instance by “discovering” X-Z but not Y-Z, this should not be counted as a complete failure. The intuitive assumption in this example is that the system should be measured with Precision 100% and Recall 50%.

Vilain et al. (1995) proposed a scoring algorithm providing results that are close to this intuition (it was used for the Coreference Task at the Sixth Message

Understanding Conference). Their algorithm computes Recall by taking into account each equivalence class induced by the relation of coreference and by determining, with respect to the output, the number of missing coreference links that would be needed to place all the relevant entities into the equivalence classes. It is worth taking a look at the details.

Let K_i be the correct coreference chains and R_i the response the coreference chains output by the resolver. Let K_i be an equivalence class induced by the key. Accordingly, the items of K_i that are each coreferent with the others. Let R_1, \dots, R_n be all the equivalence classes induced by the coreference relation as defined by the response. The auxiliary functions over K_i are defined as follows.

$p(K_i)$ is a partition of K_i relative to the response such that each subset of K_i in $p(K_i)$ is formed by the intersection of K_i and any R_i . Singleton sets in $p(K_i)$ contain elements that are not mentioned in the response.

$c(K_i)$ is the minimal number of links between the elements of K_i . It is given by $c(K_i) = |K_i| - 1$.

$m(K_i)$ is the number of links missing in the response as regards K_i . This is one fewer than the number of sets in the partition $p(K_i)$, $m(K_i) = |p(K_i)| - 1$.

Recall for K_i is given by

$$\text{Recall}_{K_i} = \frac{c(K_i) - m(K_i)}{c(K_i)}, \text{ or equivalently } \frac{|K_i| - |p(K_i)|}{|K_i| - 1}.$$

Extending this measure from K_i to all K_1, \dots, K_n induced by the key, the global Recall of the resolver is given by

$$\text{Recall}_T = \frac{\sum (|K_i| - |p(K_i)|)}{\sum (|K_i| - 1)}$$

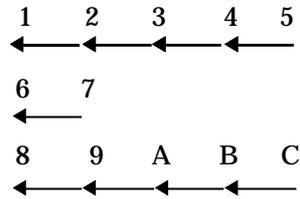
Precision is computed by replacing the number of correct links by the number of obtained links in the above formulation,

$$\text{Precision}_T = \frac{\sum (|K_i| - |p(K_i)|)}{\sum (|R_i| - 1)}.$$

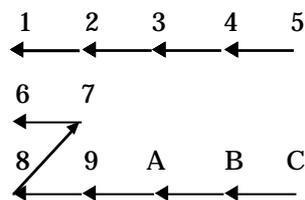
In Bagga and Baldwin, 1998a, an improvement to this algorithm was proposed. This improvement aims to bring the evaluation algorithm even closer to the intuition as to what the performance of an anaphor resolver is and how it should be scored. The authors observe that although it is the anaphoric links that count, under Vilain et al.'s algorithm all the errors are considered to be equal in terms of their negative impact in the performance of the resolver. Their observation that not all errors have identical impact is illustrated by the case-study sketched below (cf. Bagga and Baldwin, 1998a, figs.1-3):

Heuristics and Constraints

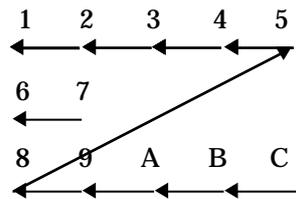
(12) a. Key



b. Response 1



c. Response 2



Taking (12)a. as the key, admittedly the error displayed in response (12)b. is less damaging for the resolver performance than the error displayed in response (12)c.. For both cases, however, the algorithm presented above will assign 100% Recall and 90% Precision. The intuition is thus that Precision should measure differently in each case, with a greater value for (12)b..

To accommodate this improvement, Baldwin and Bagga (1998a) set up what they called the B-cubed algorithm. Informally, instead of looking at the coreferential links output by the resolver, this algorithm looks at the coreferential items relative to each of the items to be anaphorically linked by the resolver. Let us take a look at its formal details.

$p(K_i)$ is defined as above. Hence let $p(K_i) = \{P_1, P_2, \dots, P_m\}$.

$m_j(K_i)$ renders the number of elements of K_i missing in P_j ,

$$m_j(K_i) = (|K_i| - |P_j|).$$

The total number of correctly coreferring items missing with respect to every item e in P_j is given by

2.2 Recent Developments and Open Lines of Research

$$m_{P_j} = \sum_{\text{for each } e \in P_j} m_j(K_i).$$

And the number of missing coreferring items with respect to each item e in S_i is given by

$$m_{S_i} = \sum_{j=1}^m \sum_{\text{for each } e \in P_j} m_j(K_i).$$

Given that the recall error for each item e is the number of missing items coreferring with e divided by the number of items of the corresponding equivalence class, Recall for K_i is the average of recall errors of each e in K_i , given by

$$\text{Recall}_{K_i} = \frac{1}{|K_i|} \sum_{j=1}^m \sum_{\text{for each } e \in P_j} \frac{m_j(K_i)}{|K_i|}.$$

The final Recall for the n equivalence classes induced by the key will thus be

$$\text{Recall}_T = \frac{1}{n} \sum_{i=1}^n \text{Recall}_{K_i}.$$

Precision in turn will be

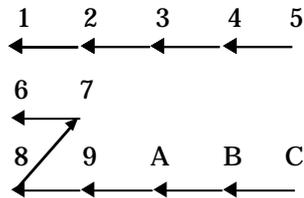
$$\text{Precision}_T = \frac{1}{n} \sum_{i=1}^n \text{Precision}_{K_i},$$

where Precision for each K_i is given by

$$\text{Precision}_{K_i} = \frac{1}{|K_i|} \sum_{j=1}^m \sum_{\text{for each } e \in P_j} \frac{|K_i| - m_j(K_i)}{|R_e|}.$$

This new scoring algorithm effectively differentiates between the Precision measures of the outputs sketched in diagrams (12)b./c., repeated below, where the measures yielded by the algorithm devised by Vilain et al. (1996) and the b-cubed algorithm are compared.

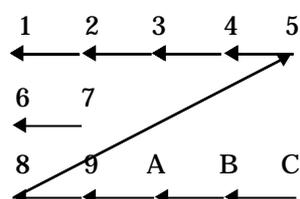
(13) a. Response 1



Precision (Vilain et al., 1996) 90%

Precision (b-cubed) 76%

b. Response 2



Precision (Vilain et al., 1996) 90%

Precision (b-cubed) 58%

We should finally mention Mitkov (1998b), who proposed a third parameter of evaluation as a scoring scheme for anaphor resolvers. This author proposes what he terms as a “critical success rate”, which is intended to measure how good the performance of a resolver is in handling the critical cases where the set of antecedent candidates has more than one element. The critical success rate is then the Recall measure of the subset of anaphors which are under such circumstances.

Robustness

Progress in research on anaphor resolution has led not only to improved understanding of anaphora, but also to the increasing awareness that developing an anaphor resolution system presents a conundrum typical of engineering problems. In order to obtain better resolution systems, research has focused on finding out more about anaphora. But the more one knows about anaphora, the more information is required from other NLP components to feed anaphor resolvers. When more NLP components are involved, it is more likely that the whole process will slow down and that certain parts of the process will fail to complete their subtask. Therefore, with current technology, the more one tries to increase the accuracy of anaphor resolvers, the more one decreases their speed and robustness. This brings us to the trade-off problem typically presented in engineering, i.e. finding the convenient balance between accuracy and utility.

Since the integrative approach to anaphor resolution was shown to be practically feasible by Lappin and Leass in 1994, there have been several studies aiming to resolve this issue. Although the goal of evaluating anaphor resolution in itself has only recently become the focus of active research, as can be seen from the discussion above, and no shared evaluation metric has been used, some studies in this area may nevertheless provide valuable preliminary indications about progressing on the basis of lean linguistic knowledge for faster resolvers.

Kennedy and Boguraev (1996) were the first to scrutinize how the integrative approach of Lappin and Leass (1994) can be evolved by scaling down some of its complexity. For the sake of speed and robustness, they dispensed with configurational syntactic information so that their resolver could avoid the eventual pitfalls of the syntactic parser outcome. They worked with “a flat morpho-syntactic analysis of the sentences” (p.113) which nevertheless assigned grammatical functions to complements, and reported a Recall of 75% for the resolution of pronouns in their test corpus.

Both Williams et al. (1996) and later Mitkov (1997a) pursued this line of inquiry and have tried to streamline further still the outcome of the syntactic analysis for anaphor resolvers. These authors experimented with syntactic representations resulting purely from part of speech tagging and parsing NPs. While Mitkov (1997a) reported a Recall of 80% for personal pronouns resolution with his test corpus, Williams et al. (1996) reported 61% for personal and demonstrative pronouns in their test corpus with a set of constraints and preferences more streamlined than the one used by Mitkov.

Baldwin (1997) introduced yet another restriction. His system was designed to attempt to resolve only anaphors whose set of antecedent candidates is a singleton. Basically, his system, under a given recency window, was thus running with filters alone. The system was said to have a 60% Recall and 90% Precision with the chosen test corpus - a poorer score for Recall than the systems referred to above, as expected.

A resolver using even more streamlined syntactic information is reported in Kameyama, 1997. This system used the output of a finite state parser which delivered “a set of (often overlapping or discontinuous) finite state approximations of sentence parts” (p.47). The resolver had a 59% Recall trying to resolve pronouns, definite descriptions and proper nouns, using only a recency preference.

When discussing anaphor resolvers on the basis of lean linguistic information, special mention is due to Hobbs (1978). The resolver reported in his work implements an algorithm that, contrarily to integrative algorithms, only uses a single type of linguistic information, without integrating different sorts of preferences. This resolver of pronouns relies only on information on the configurational syntax of the context surrounding the relevant anaphor, obtained from a rule-based syntactic parser – which the more recent works referred to above have tried to avoid, in the belief that this type of information may be more damaging for the robustness of the resolution process.

Hobbs' algorithm consists basically of an examination procedure in a tree comprising a left-to-right depth first search outside of a minimal path to the pronoun. This path satisfies certain configurational constraints to ensure that expressions are excluded from the search which, broadly speaking, are clause

mates of the pronoun being resolved. The branches searched for are to the left of the pronoun and the first suitable item encountered is taken as the antecedent. Lappin and Leass (1994) manually tested this algorithm against their test data – under the assumption thus that the resolver implementing it would work with a fully reliable parser. They measured 81% Recall, which is lower than the 86% measured by the system they had implemented, yet nevertheless a very noteworthy result for an algorithm relying purely on a single type of information.

Generality

Encouraged by the results reported in the literature on the practical feasibility of the integrative approach, another complementary trend in research on anaphor resolution has been consistently pursued since then. In the works referred to above, the goal was to find out how far one can go in terms of increasing the robustness of anaphor resolution with acceptable accuracy results. Other works, however, have focused more on understanding how far the generality of the integrative approach can be progressed. These works have addressed the question of whether and how the methods for resolving third person personal pronouns, devised by Rich and LuperFoy (1988) and Carbonell and Brown (1988) and thoroughly implemented by Lappin and Leass (1994) can be used for resolving all types of anaphoric expressions.

As already mentioned above when we addressed the issue of preferences, in order to cope with nominal expressions not only with anaphoric force, such as pronouns, but also with discriminative content, such as definite descriptions or proper nouns, Gaizauskas et al. (1996) used the notion of semantic distance in an ontology.

Gaizauskas et al. (1996) describes the LaSIE system, which was used in the 6th Message Understanding Conference (MUC-6) where ontological similarity preference was implemented. In the official texts of the MUC-6, the system is reported to have achieved 51% Recall and 71% Precision.

Poesio, Vieira and Teufel (1997) report on another experiment attempting to refine the similarity preference based on hierarchical ontologies for resolving definite descriptions. They focused on the problem they called bridging references, roughly on anaphoric definite descriptions which do not have the same head noun of the antecedent. The occurrences of this type of definite description amounted to 20% of the total number of occurrences of definite descriptions in their corpus and involved different types of semantic relations between the anaphoric expressions

and their antecedents. Their system was designed to solve bridging anaphora where the relation was of coreferentiality (a car...the vehicle), hypernymy/hyponymy (the television show...the program), meronymy (house...the chimney), names (Bach...the composer) and between compounds and non compounds (discount packages...the discounts).

The central component of the system was WordNet, a large-scale lexical semantic network for English (cf. Miller et al., 1993). The results ranged between 35% and 56% for Recall, and 71% and 94% for Precision.

Later, Poesio et al. (1998) again attacked the issue of bridging anaphora, but this time the semantic similarity was calculated without resorting to any pre-defined ontological hierarchy. The basic idea was to calculate the degree of semantic association just by using frequency data on the cooccurrence of words in corpora.

This work was inspired in the model of lexical clustering proposed by Lund et al. (1995). This model assumes an n -dimensional space. The dimensions are the n more frequent words in a given corpus. A vector $C(w)$ is associated with a given word w , where a cell i is meant to record how frequently w occurred close to w_i . The cells of $C(w)$ are calculated by specifying an inspection window with a length of k words before and k' words after w , and by scanning the corpus for occurrences w for successively incrementing their values. The cell i is incremented in a given occurrence of w , if in that occurrence w_i occurs in the inspection window. The factor by which the cell i is incremented depends on the distance of w_i to w . The distance between two words in this n -dimensional space is thus given by the distance between their vectors.

Poesio et al. (1998) experimented with different sizes for the inspection window and different methods to find the distance between two vectors. They did not use any kind of filter for anaphor resolution. From the expressions found in the five previous sentences, the antecedent for a given definite expression was then taken as the one that was closer to it in the n -dimensional space. The authors reported that "the best parameter configuration ...only resulted in appropriate antecedents for 23.6%" of the bridging anaphors they were concerned with (p.86). In a detailed evaluation of the results, they observe that "worse results were obtained for those inferential descriptions whose relationship to the antecedent is based on synonymy, hypernymy/hyponymy or meronymy than had been obtained with WordNet (22.2% instead of 39.5%), but better results were obtained for all other cases (22.8% ... opposed to 9.6% with WordNet)" (p.87).

With respect to generalizing anaphor resolution techniques to types of anaphoric expressions other than pronouns, mention should finally be made of recent efforts to cope with event anaphora as well. Humphreys and Gaizauskas (1996) report on a first attempt to look at clauses as definite descriptions of events that can entertain semantic relations among them. Their work assumed a hierarchical ontology for

events, which are described by templates organizing the different entities involved in the events. Antecedent candidates were passed through a similarity filter that checked whether events had their entities involved in the same roles. After passing this constraint, the ontology was consulted to determine what kind of relation existed between the events under scrutiny. This resolver had a very specific domain of application: texts with news about management successions in companies.

Multilinguality

As in most realms of natural language knowledge and processing, English has been by far the most widely studied language with respect to anaphor resolution. Almost all the results reported in the papers mentioned above, were obtained with systems accepting English texts as input. Therefore, when inquiring about the generality of the integrative method of anaphor resolution, one should look not only at how it performs and how it can be extended when handling different kinds of anaphora. It is also of interest to study the issue of multilinguality, i.e. how it performs when anaphor resolution is to be performed in natural languages other than English.

A couple of papers have already addressed this question. Aone and McKee (1993) and Popescu-Belis (1998), for instance, contributed with a conceptual clarification of the general architecture a language independent semantic processor should have and how the anaphor resolver fits into it. The suggestion is that the component for anaphora handling should be assigned a central role in any semantic processor and that the integrative method of anaphor resolution, given its flexibility, is in principle able to cope with multilinguality.

Authors such as Azzam et al. (1998), and Mitkov et al. (1998) tried a more concrete experiment. They extended and/or adapted anaphor resolvers originally developed for English to handle texts also in other languages, such as French, Arabic or Polish. The general conclusion is that the eventual differences in performance found between the resolvers is due to the difference in performance of other natural language processing components whose outcome anaphor resolvers must work with (e.g. the syntactic and semantic parsers), rather than to eventual inadequacies of the integrative method in coping with anaphora in different languages.

Also in this vein of multilinguality, other set of works worth mentioning are those like Mitkov, 1996c, and Mitkov and Schmidt, forth., concerned with the

relevance of anaphor resolution for machine translation. For the sake of efficiency, the rationale behind translating anaphoric expressions is that resolution should be avoided as much as possible. This means that if w_s is the anaphoric expression of source language to be translated, it should be translated to a w_t of the target language, with the same anaphoric properties with respect to the filters and preferences to be observed.

- (14) a. John said to Robert that Peter was talking to him.
the John said to_the Robert that the Peter was to talking with him
- b. O John disse ao Robert que o Peter estava a falar com ele.
the John said to_the Robert that the Peter was to talking with him
- b'. O John disse ao Robert que o Peter estava a falar consigo próprio.
the John said to_the Robert that the Peter was to talking with himself

Taking (14) as an example, we notice that a suitable translation of (14)a. is b., not b', though both b. and b' result from the translation of him by Portuguese anaphoric expressions. The point is that him in (14)a. can at least refer to John or Robert, but not to Peter. What is then important is to translate him by an expression with identical anaphoric capacity, independently of the exact antecedent of him turns out to be. This will ensure that the translation of him will have the same anaphoric capacity, i.e. that it will be able to refer to the translation of any of the possible antecedents of him.

Languages, however, have different systems of anaphoric expressions and it may happen that for a given anaphoric expression w_s , no anaphoric expression in the target language can be found with identical anaphoric capacity. If it is known that the target language has an anaphoric expression w_t whose anaphoric capacity subsumes the anaphoric capacity of w_s , i.e. w_t may have at least all the possible antecedents in any grammatical construction that w_s may have, then the costly process of anaphor resolution can still be avoided, and w_t can be used as the translation of w_s .

Anaphor resolution, however, cannot be avoided if that circumstance does not hold. If in the set of anaphoric expressions of the target language one can only find expressions whose anaphoric capacity only properly overlaps the anaphoric capacity of w_s then the exact antecedent of w_s has to be found so that an appropriate translation can be given. To translate w_s we will then choose one of the anaphoric expressions of the target language that can have the translation of the antecedent of w_s as antecedent in the translated text.

This situation is illustrated in (15) (Xue et al. (1994, ex.(3))).

- (15) a. Zhangsan zhidao Lisi renwei Wangwu zui xihuan ziji.
Zhangsan know Lisi think Wangwu most like self

- b. Zhangsan knows that Lisi thinks that Wangwu likes himself most.
- b'. Zhangsan knows that Lisi thinks that Wangwu likes him most.

The Chinese sentence in (15)a. can be translated into English as (15)b. or b'.. The difference has thus to do with how the long-distance reflexive *ziji* is resolved, a type of anaphoric expression that does not exist in English. Depending on how it is resolved, it can therefore be translated into English either as *him* or *himself*.

This shows that, for machine translation as well, anaphor resolution is an important component when resolving anaphors is unavoidable in order to calculate the translation of anaphoric expressions.

Dialogue and multimodality

Anaphor resolution is considered by some to be of major relevance for dialogue processing as well.

Building on previous observations by Francez and Berg (1994) and Groenendijk et al. (1997), Poesio (1998) is a first attempt to inquire about what kind of possibly specific filters may arise with cross-speaker anaphora. The most likely path of research to be explored, as dialogue corpora become more easily available, is to use known techniques from the integrative framework to address the resolution of anaphoric links between anaphors occurring in utterances of a given participant in a dialogue vis-à-vis their antecedents in the utterances of other participant

The same strategy of extending the integrative model for linguistic anaphors resolution has been adopted to handle dependent reference in multimodal contexts. Huls and Bos (1995) report on the implementation of a anaphor resolver for a multimodal interface that handles gestures, pointing, etc. as relevant elements in the pool of preferences for choosing the antecedent.

Identification of referents

A final word is due in this section to one of the more recent and promising research paths in reference processing. The problem to be addressed by an anaphor resolver can be said to typically consist in identifying the antecedents of anaphoric

expressions. It is often the case that antecedents of anaphors are themselves anaphors that have their own antecedents, which gives rise to chains of anaphoric dependencies.

In this regard, the interesting point made by Bagga and Baldwin (1998b) is that different entities may bear the same proper name. While we are interested, as a first step, in uncovering the relevant anaphoric chains, in a second step, we are interested in identifying the different chains whose items are referring to different entities, although the same proper name may be involved in those chains. Accordingly, in this second step, the central task is no longer to identify antecedents but to identify anaphoric chains. Now one is interested in identifying the different referents that may be referred to with type-identical proper nouns, occurring in different anaphoric chains.

The practical importance of this task of distinguishing anaphoric chains with identical proper nouns but with different referents is immediately apparent when we think of tasks such as information retrieval or multiple texts abstracting. When looking in a textual database or in the world wide web for a person called John Smith, we would like the texts about Poca Hontas' boyfriend to be clustered in a different set to that of the texts on the former Leader of the Labor Party in Britain.

Another rudimentary, but often very useful way of constructing abstracts about John Smith from a textual database is to concatenate all the sentences in all documents containing items referring to Poca Hontas' boyfriend, on one side, and all sentences with items referring to the other more recent John Smith.

The method used by Bagga and Baldwin (1998) supposes the following steps. First, in different documents anaphoric chains are uncovered for a given proper name PN. Second, for each document, a "summary" is built by collecting all the sentences where noun phrases referring to the entity referred by PN occur. Third, a vector space model is used to compute the similarity between these summaries. The summaries extracted are stored as vectors and if they have a similarity above a certain threshold they are considered to concern the same entity. The model used is the standard vector space model widely used in information retrieval (cf. Bagga and Baldwin, 1998, p.81 for more information).

2.3 Centering Attention for Anaphor Resolution

On a par with its development in the integrative framework, the study of reference processing and anaphor resolution has also been pursued along another major line of research, the framework of Centering Theory. As a preliminary very rough distinction, it can be said that the first addresses the goal of anaphor resolution from a more linguistically-driven perspective, and the latter takes a more cognitively-driven stance. According to Sidner (1981, p.218), what is specific of Centering Theory in terms of anaphor resolution is the fact that “antecedence is a kind of cognitive pointing, the kind of pointing that causes [anaphor] and [antecedent] to point (somehow) to the same class of entities in one’s mind”.

What came to be known as Centering Theory (CT) is the result of converging original lines of research by Joshi and Kuhn (1979) and Joshi and Weinstein (1981) on inference processing and discourse structure, by Sidner (1981) on resolution of reference and focus of attention, and by Grosz (1977) on dialogue understanding and centering of attention. A unified account of these approaches would appear in Grosz and Sidner (1986), which reached a wider audience later with the version by Grosz, Joshi and Weinstein (1995). An updated and concise overview of CT can be found in Walker, Joshi and Prince, 1998, and Grosz and Sidner, 1998.

Attentional state, discourse structure and form of expressions

The rationale behind CT is that there exists a correlation between the speaker’s attentional state, on the one hand, and discourse structure, inference processing time and the form of anaphoric expressions, on the other. CT assumes that

discourse is organized as a hierarchy of discourse segments such that in each segment the attention of speakers is centered in some entity that the segment is about. Concomitantly, the theory assumes that once the attention of speakers is centered by virtue of the interpretation of previous utterances in a given entity, then subsequent reference to that entity in the same discourse segment should be done by means of informationally weaker expressions.

This implies that CT assumes that expressions have different degrees of informativeness. For instance, a pronoun is less informative than a proper noun. This implies also that CT is assuming a minimization imperative that demotes the use of more informative items when their reference is already the center of attention, which is meant to reflect the fact that more processing time is required in these circumstances.

Possible relevant empirical evidence supporting CT is due inter alia to Hudson-D'Zmura and Tannenhaus (1998). They undertook experiments which revealed that subjects take less time to read discourse (16)a., 2,158 msec, than (16)b., 2,475 msec (p.205).

- (16) a. Jack apologized profusely to Josh.
He had been rude to Josh yesterday.
- b. Jack apologized profusely to Josh.
Jack had been rude to Josh yesterday.

In the second sentence of the first discourse, a pronoun is used to refer to Jack, the supposed center of attention of the previous utterance, and of the whole two sentence discourse. In the second sentence of the second discourse, a proper noun is used instead. This is taken as supporting the correlation between keeping attention centered on a given entity and using expressions with weaker informative potential.

This basic correlation by CT is formalized on the assumption that there are discourse entities called centers, a set of three Constraints on these discourse entities, and a set of two Rules, one about the form of referring expressions, and the other about different processing times in different types of changes in terms of attentional state.

Centers are “semantic entities that are part of the discourse model” (Walker, Joshi and Prince, 1998, p.3) and serve “to link an utterance to other utterances in the discourse segment that contains it” (Grosz, Joshi and Weinstein, 1995, p.208). In spite of its quite informal characterization in CT literature and its intended cognitive mood, it certainly will not be incorrect (following, among others, Webber, 1982) to say that, in a model-theoretic setup, centers are terms denoting entities that are the reference of the expression to which centers correspond. Their close

similarity to reference markers from Discourse Representation Theory is brought to light by Roberts (1998).

Following the exposition of the theory by Walker, Joshi and Prince (1998), centers come in two sorts, forward-looking and backward-looking. The set of forward looking centers of utterance U_i in a discourse segment D , $Cf(U_i, D)$, is “the set of centers evoked by an utterance U_i in a discourse segment D ” (p.3). The backward-looking center is just a distinguished member of the set of forward-looking centers, which is organized in a partial order.

The set of three Constraints of CT specify how centers are in correspondence with relevant utterances and other centers (p. 3):

(17) Constraints of Centering Theory

For each utterance U_i in a discourse segment D consisting of utterances U_1, \dots, U_m :

1. There is precisely one backward-looking center $Cb(U_i, D)$.
2. Every element of the forward-looking centers list, $Cf(U_i, D)$, must be realized in U_i .
3. The center $Cb(U_i, D)$ is the highest ranked element of $Cf(U_{i-1}, D)$ that is realized in U_i .

On the other hand, what are termed as Rules of CT (p. 4) specify how processing load is related to transitions of attentional state (Rule 2), and how the form of referring expressions is related to the center of attention (Rule 1).

(18) Rules of Centering Theory

For each U_i in discourse segment D consisting of utterances U_1, \dots, U_m :

1. If some element of $Cf(U_i, D)$ is realized as a pronoun in U_{i+1} , then so is $Cb(U_{i+1}, D)$,
2. Transition states are ordered. The Continue transition is preferred (less processing time) to the Retain transition, which is preferred to the Smooth-shift transition, which is preferred to the Rough-shift transition.

Rule 2 supposes a definition of different types of attentional state transitions. The typology of transitions is based on two factors:

(i) whether the backward-looking center is preserved from the previous utterance, i.e. from U_{i-1} to U_i , in case U_{i-1} exists;

(ii) whether the backward-looking center of U_i is likely to be preserved in the next utterance, i.e. if the backward-looking center of U_i is also the preferred forward-looking center of U_i , $Cp(U_i, D)$, where $Cp(U_i, D)$ is “the highest ranked

member of the set of forward-looking centers” (p. 3). The typology of transitions can thus be displayed in the following table:

(19) Typology of transitions in Centering Theory

		Cb is preserved from U_i to U_{i+1}	
		yes	no
In U_i Cb is the C_p	yes	Continue	Smooth-shift
	no	Retain	Rough-shift

Rule 2 predicts that more time will be required for processing utterances that do not preserve the Cb from the previous utterance, and also for processing utterances whose Cb is not the highest ranked Cf, i.e. utterances that make it hard for the next utterance to have the same Cb as the previous one.

Although not formalized in any Constraint or Rule, the literature on CT is clear about the fact that it is expected that more time-demanding transitions in attentional state will occur across structurally wider boundaries between discourse segments (cf. Grosz and Sidner, 1998).

As to Rule 1, it states that if at least one pronoun of U_{i+1} picks its antecedent from the immediately preceding utterance U_i , then at least one pronoun of U_{i+1} refers to the highest ranked forward-looking center of the previous utterance.

Algorithm for anaphor resolution

Given how CT is formalized, Rule 1 turns out to be the central principle provided by CT which any CT-oriented algorithm for anaphor resolution can be based upon. Note, however, that before Rule 1 – which interacts with Rule 2 and the Constraints – could be put to use with such desideratum, it is still necessary a preliminary clarification on (i) how to rank forward-looking centers, and on (ii) how to ascertain discourse segments boundaries.

The paper by Brennan, Friedman and Pollard (1987) is the most prominent attempt to design an anaphor resolution algorithm on the basis of the constructs of CT. In that paper, a practical solution was proposed both for (i) and (ii) above. With

respect to ranking the forward-looking centers of the utterance preceding the one in which the pronoun occurs, the paper proposes “to rank the items in Cf by obliqueness of grammatical relation of the subcategorized functions of the main verb” (p.156). As for the problem of finding rigorous criteria to ascertain discourse segments, that question was somehow avoided by pushing it back. In fact, the proposed algorithm was designed with no a priori knowledge about discourse segments boundaries. Rather, if a pronoun happens to have more than one antecedent candidate, the preferred candidate is the one whose choice will induce the lowest ranked transition of attentional state, i.e. the one that supposedly requires less processing time. Consequently, it was implicitly assumed that the algorithm applies to sentences in the same minimal discourse segment (see below for further discussion on this issue).

The proposed algorithm basically comprises the following steps (vd. Brennan et al. (1987, p.159) for a fully detailed account):

BFP Algorithm

In an utterance U_i :

1. Construct a list of relevant expressions of U_i ordered according to obliqueness of grammatical function

e_1, \dots, e_n

2. Construct a list of possible solutions for anaphora. Each possible solution results from replacing expressions in the list of 1. by centers that can be their antecedent:

c_{11}, \dots, c_{n1}

c_{21}, \dots, c_{n2}

...

3. Construct a list of possible backward-centers. This is a list of the forward-looking centers of the previous utterance, U_{i-1} , plus Nil, a value accounting for the absence of C_b in initial discourse utterances.

$C_{b1}, \dots, C_{bk}, Nil$

4. Find the cross-product of lists 2. and 3.:

$(C_{b1}, (c_{11}, \dots, c_{n1}))$

$(C_{b2}, (c_{11}, \dots, c_{n1}))$

...

$(C_{bk}, (c_{11}, \dots, c_{n1}))$

(Nil, (c₁₁, ..., c_{n1}))

(Cb₁, (c₂₁, ..., c_{n2}))

....

(Cb_k, (c₂₁, ..., c_{n2}))

...

5. Apply Constraint 3: In list 4., scan each sublist headed by the consecutive backward-centers [(Cb₁, ...) ... (Cb₁, ...) ... [(Cb_k, ...) ...]. When scanning a given sublist from top to bottom, keep the lines where the Cb_i is identical to the first element of the associated Cf(U_{i-1}) list in the sublist headed by Cb_i. Eliminate the other lines of the sublist.

6. Apply Rule 1. In the list resulting from step 5., eliminate the lines where at least one element of the Cf list corresponds to a pronoun of U_i and the associated Cb_i is not identical to any of those pronominal elements.

7. Assign to each line of the list resulting from step 6. a classification in terms of the typology of centering transitions.

8. Apply Rule 2: Order the list resulting from step 7. according to the preference for those supposedly requiring less processing time.

9. The first line in the list resulting from step 8. is the proposed solution for anaphora in U_i.

Attentional inertia

It has been widely acknowledged in the literature that this algorithm for anaphor resolution by Brennan, Friedman and Pollard (1987) (BFP algorithm), which implements the Rules and Constraints of Centering Theory, reflects quite faithfully the essential tenets of the theory. Accordingly, the algorithm has been submitted to intensive scrutiny both in terms of its empirical adequacy and in terms of its practical importance. It has been found to be problematic on several counts.

One of the major problematic issues worth pointing out has to do with the fact that the BFP algorithm converts a hierarchy of duration times of attentional transitions into a hierarchy of preferences for the use of referring expressions. In fact, it is somewhat contentious to suppose that people organize their utterances not according to what they intend to express and the many different topics they may want to address, but according to how much they can minimize the processing time of their

discourses. Nevertheless, when the BFP algorithm in line 8 uses indicators of processing time to rank the preferred solutions for anaphora, it is assuming that there is a sort of decisive attentional inertia causing speakers to somehow strive to avoid changing the topic they are talking about.

One would actually expect a quite different use of the attentional transition hierarchy of CT in the process of anaphor resolution. First of all, one would expect, on the basis of empirical evidence independent from the process of anaphor resolution, the type of attentional transition to be identified. Secondly, one would expect, given this information and the information on the form of the relevant anaphoric expressions, that it would be possible to identify their antecedents in view of the attentional transition occurring.

There is a smooth and cheap way that suggests itself and allows to overcome this drawback. It consists simply in restraining the coverage of the algorithm. In particular, it consists in assuming that the algorithm is supposed to apply only within minimal discourse segments, being that a minimal discourse segment is circumscribed by virtue of the fact that its utterances are about one and the same topic. This restriction to local centering, as it is known in the CT literature, is however a weak solution. In fact, to claim that BFP algorithm applies locally is just to push the problem of CT-oriented anaphor resolution back and reverting it to finding objective and anaphora independent criteria to identify discourse segment boundaries. And this is a highly problematic issue, as we will discuss below.

Building on these insufficiencies, it is quite straightforward to construct simple samples of acceptable discourses that CT predicts as not being possible. Consider, for instance, the discourse comprising the two sentences below.

- (20) a. It was Peter who in the morning asked Jane to come with her son to the party.
b. But it was Peter's wife who in the afternoon eventually picked him up at School because Peter and Jane were caught in a traffic jam.

According to Constraint 3 of CT, the backward-looking center of sentence (20)b. is the highest ranked element of the set of forward-looking centers of (20)a. realized in (20)b.. The highest ranked element of forward-looking centers of (20)a. is the center associated with Peter, C_{Peter} . It is focused in (20)a. with the help of a cleft construction and happens to be realized in (20)b.. Accordingly, the backward-looking center of sentence (20)b. is C_{Peter} .

Now, coming to Rule 1 of CT, one is told that if some element of the set of forward-looking centers of (20)a. is realized in (20)b. as a pronoun, then the backward-looking center of the latter, C_{Peter} in the present case, has to be realized as a pronoun.

In fact, there is only one pronoun in (20)b., namely him. Since it refers to Jane's son by having as antecedent her son of (20)a., it realizes a forward-looking center of (20)a.. Consequently, CT predicts that the backward-looking center of (20)b. has to be realized also as a pronoun. As there is only the pronoun him in (20)b., that pronoun is predicted by CT to realize C_{Peter} . This implies that CT, and a fortiori the BFP algorithm, makes the wrong prediction that him refers to Peter and not to Jane's son, as it can only be the case in (20)b..

Discourse locality

Be that as it may as regards the possibility of ensuring a reliable discourse segmentation independent of anaphora, the BFP algorithm also presents non-negligible difficulties concerning anaphors with antecedents not occurring in the immediately preceding utterance. According to an estimate by Hitzeman and Poesio (1998), using the corpus collected for the LILEX project, the anaphoric expressions in a given utterance U_i with their antecedent occurring in an utterance preceding U_{i-1} are around 8% of the total amount of anaphors. Hahn and Strube (1997), in turn, in the corpus they used detected 30% of anaphoric expressions whose antecedents were in the conditions described. Given its primitives, it is easy to understand that the BFP algorithm is not suited to providing any outcome to these types of anaphoric link. But perhaps the more crucial point to make in this connection relies in the observation that the insufficiencies of the BFP algorithm do not stem from partial implementation of CT primitives, but rather reflects essential limitations of CT itself in handling non-locality.

This issue of so-called non-local pronominal anaphora has been addressed in some recent papers, where it is possible to devise two basic lines of solution. On the one hand, papers such as Hahn and Strube, 1997, Hitzeman and Poesio, 1998, and Cristea, Ide and Romary, 1998, tend to adhere to the core notions, constraints and rules of CT, and extend the range of action of an algorithm originally prepared for local anaphora. This is done by trying to devise suitable structure for discourse and notions of proximity between discourse segments. The rationale of this strategy is to preserve the intuition that, given that working memory has a short capacity, the center of attention cannot be focused to entities out of it. Therefore, cases of non-local anaphora are expected to be somehow reduced to legitimate instances of local anaphora across discourse segments with some appropriate vicinity relation between them.

On the other hand, papers such as Walker, 1996, 1998, follow a different line of solution. They elaborate on the notion of working memory and its articulation with long term memory, rather than on discourse structure and eventual legitimate local “shortcuts” across discourse segments boundaries. The basic idea is that items may be transferred between the two types of memory, which would then enable one to understand why non-local anaphora is possible in the framework of CT.

These proposals are very recent and their insights are tentative in nature, awaiting further elaboration. Nevertheless, certain fundamental problems can already be spotted. Approaches that rely on discourse segmentation have to cope precisely with the problem of discourse segmentation according to objective criteria which support anaphora independent segmentation procedures. As stated above, we are some distance away from finding a practically acceptable solution to this issue.

As for Walker’s proposals, although they may enhance practical solutions concerning anaphor resolution systems, as we will see below, they rely on barely formalizable primitives. It is really hard to devise any sort of viable or productive formalization of the relations between cognitive constructs such as working memory and long term memory on the sole basis of language analysis, as is required by automatic language processing devices.

One single pronoun and non-pronominal anaphors

Coming back to (20) as a case study for the empirical adequacy of CT, the discourse presented there is also useful to bring some other limitations of the theory to our attention. For instance, it became clear that, when applying Rule 1 of CT, we had no problem with the fact that this Rule is not explicit about which pronoun realizes the C_b, if there is more than one pronoun. As there is only one pronoun in (20)b., we had no problem in applying Rule 1. However, that exercise made us aware of the potential problem we would encounter if the sentence at stake had more than one pronominal expression to be resolved using CT. The fact is that CT and a fortiori the BFP algorithm have nothing to say about how the remaining pronouns are or should preferably be interpreted.

Interestingly, Gundel (1998) though elaborating on a different issue, namely on the possible confluence of the proposal by Gundel, Hedberg and Zacharski (1988) and the tenets of CT, diagnoses as the possible source of this drawback the fact that “the constraint against more than one C_b per clause may be too strong” (p.195). In fact,

psycholinguistic experiments like Greene et al., 1992, p.266, have brought to light evidence that resolution processes “may not always identify a unique referent for a pronoun”.

Nevertheless, difficulties would arise if there were more than one backward-looking center. It would strongly contend with some core aspects of CT. In particular, it would involve the self-defeating assumption for CT that attention may be equally focused on several different entities. In this case, several mental entities evoked by eligible antecedents could be the actual and concomitant centers of attention, which would erase any attentional clue as to which pronoun refers to which entity and defeat the assumption that minimal discourse segments organize around a given center of attention.

By the same token, it almost goes without saying that when the resolution of non-pronominal anaphoric expressions is addressed, the insufficiencies of BFP algorithm become even more evident.

It should also be noted that line 8. makes use of the different transitions between attentional states proposed in CT. These transitions, in turn, are based only on the fact that the backward-looking centers of the current and previous utterances are identical (Continue and Retain) or not (Smooth-shift and Rough-shift), and on the fact that the backward-looking center and the preferred forward-looking center of the current utterance are identical (Continue and Smooth-shift) or not (Retain and Rough-shift). This implies that, with all the anaphoric expressions that do not realize a backward-center or a preferred forward-center, CT does not provide a solution about how to find their antecedent.

Intra-sentential anaphora

Another problematic point with BFP algorithm has to do with so-called intra-sentential anaphora, i.e. with anaphoric expressions whose antecedents occur in the same sentence. According to an estimate by Kameyama (1998, p.97), using a sample from the Brown corpus and covering only third person pronouns, around 60% of the pronouns have their antecedents in the same sentence. Despite the significant number of intra-sentential anaphoric expressions, the BFP algorithm has nothing to offer for the resolution of this type of anaphors. It offers predictions only for inter-sentential anaphora.

This shortcoming has been addressed by a number of papers such as Azzam, 1996, Okumura and Tamoura, 1996, Strube, 1996, or Kameyama, 1998. However their possible different details may be concerning how to envisage the internal

structure of sentences for the purpose of applying a CT-based algorithm for anaphor resolution, they share a common strategy. These proposals seek to keep unchanged the basics of CT and BFP algorithm, and extend their coverage from inter-sentential to intra-sentential anaphoric links by assigning a discourse-like structure to multi-clausal sentences. Under this point of view, the basic discourse units are not full sentences, but minimal subsentential clauses if they happen to be syntactically grouped into complex sentences. Accordingly, the proposed solutions for rescuing the BFP algorithm for intra-sentential anaphora bear upon a more fine grained analysis of discourse.

Even if we disregard the issue of the real adequacy of adopting this type of elementary unit for discourse and how these units are possibly gathered together in a discourse, this solution for intra-sentential anaphora still does not offer any answers for anaphoric relations within the same minimal clause, as in Bill still loves all his ex-girlfriends or Bill shaved himself. More seriously than that, however, is the fact that while aiming to extend coverage for intra-sentential anaphora, this solution restricts the previously ensured coverage concerning inter-sentential anaphora. As a matter of fact, given the “local” capacity of the BFP algorithm, as we have just discussed above, all the problems related to the inherent circumscription of its predictive power to anaphoric relations across adjacent discourse segments will now re-emerge when discussing anaphoric relations across the boundaries of multiclausal sentences which happen to be adjacent in discourse.

Correlation discourse structure vs. form of expressions

All the problems referred to above have in common the fact that there are difficulties involving the development of an anaphor resolution algorithm based in CT. We would like to turn now to more critical issues as they concern the cogency of core tenets on which CT is fundamentally based.

One of such tenets is that discourse structure, inasmuch as it is supposed to reflect changes in attentional state, plays a decisive role in the admissibility of certain forms of anaphoric expressions. Boundaries between discourse segments elaborating on different topics, correlate to corresponding changes in the centering of attention. On the other hand, still following the basic assumptions of CT, anaphoric expressions with weaker descriptive content have to be interpreted against antecedents that are associated with more attentionally prominent entities.

Accordingly, pronouns are expected to show severe restrictions in allowing for anaphoric relations which cross discourse segment boundaries.

Building on dispersed but confluent counterexamples to this expectation in earlier literature on CT (Litman, 1985, Passoneau and Litman, 1994 *inter alia*), an investigation was undertaken by Walker (1998) on the possibility that anaphoric relations entertained by pronouns with their antecedents cross discourse segment boundaries in naturally occurring discourses. To show this, she produced a systematic listing of the possible structural configurations between linearly adjacent discourse segments. She also scanned existing corpora looking for instances of discourse where these different types of boundaries could be found.

Examining the entire list she produced and the several examples she found illustrating each case would be going into too much detail for our purposes and not in line with the central aim of the present chapter. However, the conclusion she drew from her study should be mentioned. Given that for every type of boundary or structural relation between discourse segments, examples were found with pronoun-antecedent anaphoric relations across that boundary, she observes that “the data suggests that intentional [discourse] structure does not define a rule that directly predicts whether a discourse entity will be realized as a full NP or as a pronoun across a segment boundary” (p.426).

Correlation between attention and use of pronouns

On a par with the specific correlation discourse/intentional structure vs. form of referring expression just assessed, another cornerstone of CT is the correlation between attentional structure vs. form of referring expression. In this respect as well, the empirical grounding of CT appears considerably problematic.

For instance, building on the primitives of CT in her paper of 1990, Di Eugenio predicts that null subjects in Italian, even more than overt pronouns, are expected to be resolved against the backward-looking center of the corresponding sentence, and to materialize the centering transition consuming the less processing time – the Continue transition – given their null overt form. In fact, according to CT, if the information deficit of referring expressions with weaker semantic content is to be supplemented by anaphoric resolution against the entities that may constitute the center of attention, then null subjects, more than any other sort of phrase, are expected to be anaphorically interpreted in such a way.

These expectations, however, were not met, as the author herself subsequently reported, taking into account a corpus research. In Di Eugenio, 1996, we are taught that null subjects appear in only 58% of Continue transitions. Besides, more crucially for the empirical grounding of CT, only 70% of null subjects are used in Continue transitions, which means that the use of the least contentful phrase is not restricted to the least time consuming attention transition.

Interestingly, other data raising even more serious doubts on the adequacy of the correlation between center of attention and realization by a pronoun can be found in Hudson-D'Zmura and Tanenhaus, 1998, precisely in the context of psycholinguistic research looking for empirical validation of some of the CT basic assumptions.

Experiment 2 in that paper recorded the time of decision latency concerning judgments of coherence between two adjacent sentences. In the second sentence, a subject pronoun occurs whose antecedent is to be found in the previous sentence. The pronoun is not ambiguous, i.e. the linguistic context is such that its antecedent candidates set is a singleton. What varies in the different stimuli for this experiment is the grammatical function of the antecedent. It appears either as the subject or the object of the previous sentence.

For the subject-antecedent, 2,537 msec was the decision latency time, while for object antecedent the decision latency time was 2,504 msec. The difference between the times for the two types of antecedents is thus negligible, with a very slight increase in decision time with the subject-antecedent.

Now, the interesting point is that CT predicts that a difference should nevertheless exist because different centering transitions are at stake. In the object-antecedent case, since the backward-looking center of the second sentence, realized as the subject-pronoun, is different from the preferred forward-looking center of the previous sentence, there is a Retain or a Rough-shift transition. In the subject-antecedent case, in contrast, as there is no such difference, there is a Continue or Smooth-shift transition. Either of the latter two are predicted to take less processing time than either of the first two, respectively.

Smyth and Chambers (1996), while experimenting with other types of subject response, reached a possibly worse result as regards the empirical adequacy of CT predictions. In their Experiment 2, these authors did not measure times of latency decision, but times of self paced reading of the same type of sentences as described in the previous experiment of Hudson-D'Zmura and Tanenhaus (1998). Besides, they also experimented with sentences where the non-ambiguous pronoun was not in subject but in object position.

A difference in reading times both for sentences with subject-pronouns and object-pronouns was found with respect to different types of antecedents. In contrast to the results of Hudson-D'Zmura and Tanenhaus (1998), sentences with

subject-pronouns whose antecedents are also subjects require less processing time, 1,870 msec, than those whose antecedents are objects, 2,250 msec.

Sentences with object-pronouns have also led to different reading times according to the type of antecedent, with an inversion of the inequality. Sentences with object-antecedents are in this case read faster, 1,860 msec, than those with subject-antecedents, 2,100 msec.

Consequently, the results of Smyth and Chambers (1996):Experiment 2 seem to add to the suggestion drawn from the results of Hudson-D'Zmura and Tanenhaus (1998):Experiment 2. Contrary to predictions, in the sentences with an object-pronoun the antecedent involving less processing time is not the subject-antecedent, but the object-antecedent. In contrast, in the sentences with subject-pronouns, the item requiring less processing time is the subject-antecedent.

As stressed by Smyth and Chambers (1996), their experiment seems to bring to light the fact that there is a parallelism of grammatical function effect impinging on the processing of anaphoric relations, rather than an effect of attention centering. Pronouns having the same grammatical function as their antecedents seem to require less processing time than those whose antecedent have a different grammatical function.

Other experiments were also undertaken by these groups of authors, this time with ambiguous pronouns. In their Experiment 1, Hudson-D'Zmura and Tanenhaus (1998) measured reading times of sentences with a subject pronoun which could be resolved against either the subject or the object of the previous sentence. Sentences where the ambiguous subject-pronoun was resolved against the subject-antecedent were read faster, 2,158 msec, than those where the ambiguous subject-pronoun was resolved against the object-antecedent, 2,644 msec.

Again, in Smyth and Chambers (1996) we can find parallel experiments involving not only ambiguous subject-pronouns but also ambiguous object-pronouns. In their Experiment 1, they gathered judgments on the preferred antecedent of ambiguous pronouns. As to the ambiguous subject-pronoun, 85% of the subjects resolved it against the subject of the previous sentence. In contrast, for the ambiguous object-pronoun, 75% of the subjects resolved it against the object of the previous sentence.

This again is confluent with the results of Hudson-D'Zmura and Tanenhaus (1998), and reinforces, contra CT, the suggestion that it is not the center of attention, but rather a parallelism of grammatical function preference that constrains the resolution of pronouns in the reported experiments.

Implementation of BFP algorithm

Given its cognitive appeal, CT and CT-based algorithms for anaphor resolution have been one of the most active research topics in the literature on reference processing. Nevertheless, the implementation of such algorithms and its practical evaluation have not deserved scrutiny to a similar extent. An exception to this is the work reported in Azzam, Humphreys, and Gaizauskas, 1998a.

In this paper, we find a report on the implementation of the CT-based algorithm proposed by Azzam (1996), which is a variant of the BFP algorithm aiming to account for intra-sentential anaphora, where simple clauses are taken as basic discourse units. According to Azzam, Humphreys, and Gaizauskas (1998a, p.77), Recall was 53.3%, and Precision 69.7%. This result deserves the remark by the authors that “while demonstrating that the centering-based algorithm is applicable to real world text, it does question whether the more complex algorithm [based on CT] has any real advantage over ... our original approach”, which follows the hints of the integrative method. Furthermore, these results also led the authors to make some remarks confluent with the issues discussed just above. Bearing on practical results, they notice that “examining performance on the real world data also raises questions about the theoretical assumptions of [CT-based] approaches, in particular whether the center of attention is always a favored antecedent” (p. 78).

Viewed thus from different angles, inclusively from a practical perspective, the CT-based approaches to reference processing, and the CT itself, seem to have severe and deeply rooted drawbacks in terms of empirical adequacy and, consequently, also in terms of their performance in effective anaphor resolution systems.

Converging to an integrative approach

As evidence regarding its inadequacies has been piling up, recent attempts have been made with the aim of ridding CT and associated algorithms for reference processing from such inadequacies. Given that the source of most of them is deeply rooted in the primitives of the theory, these attempts had to assume, at least to a certain extent, some kind of reshuffling of the basic architecture of CT.

That is the case with the proposal put forward by Walker (1996), further elaborated in Walker, 1998. The central claim here is that the model for the working memory should be redefined. According to CT, the working memory has the structure of a stack, which allows one to understand the isomorphism between

discourse intentional structure and the cognitive attentional structure. Centers are pushed into and popped from the working memory stack according to the deployment of the intentional structure of the discourse. According to Walker (1998), however, the working memory should rather be conceived under the model of a cache memory.

Under this model, the centers of attention are retrieved from the cache and sent to the main memory, back and forth, which “does not posit that attentional state is isomorphic to intentional structure” (p.406). Therefore, by breaking the supposed correlation between discourse intention and attention, all the problems related to this get a solution. Nevertheless, this does not mean that all the difficulties had been removed. In fact, the residue of the problems detected as being related to the connection between intentional and attentional structure, reappears now as the problem of determining which, when and how the centers are brought back and forth between main or long-term memory, and cache memory.

In this respect, one is told that one should use “the pronominal features and the content of the utterance as retrieval cues for retrieving matching discourse entities from main memory” (p. 409). Then, for a given anaphor and once the cache memory is complete with antecedent candidates, one should derive the ranking of the centers in it “from a combination of syntactic, semantic, and discourse features associated with [the centers]” (p. 409).

Under our understanding, what we have here, modulo terminological renaming, resultant from reshuffling CT to rid it of its empirical inadequacies, is merely a coarse redescription of the basic architecture of the integrative method. The cache memory plays the very same role as the antecedent candidates set, where the filters to candidate eligibility are meant under the wordy description “the pronominal features and the content of the utterance as retrieval cues”. The election of the antecedent is achieved via ranking the elements of that set, where the highly ranked is the preferred candidate. This ranking is done with the help of preferences, here only vaguely referred to as “a combination of syntactic, semantic, and discourse features associated with [the centers]”.

Interestingly enough, the paper by Strube (1998) sets itself a similar goal of overcoming some well known difficulties of centering-based algorithm for anaphor resolution and, in our understanding, also lends itself to the same kind of analysis. While attempting to remedy some defects of CT, the essentials of the proposed alternative model become strikingly similar to the basic assumptions of the integrative method.

In Strube’s (1998) algorithm, the notion of working memory is replaced by “the list of salient discourse entities (S-list)” (p.1252). And the notions of “the backward-looking center and the centering transitions by the order among the elements of the list of salient discourse entities (S-list)” (p.1251). Apparently, there

is no resort to filters, as the S-list is said “to contain some (not necessary all) discourse entities which are realized in the current and the previous utterance” (p.1252), without further ado. As for preferences, there is however the use of a topic preference based on the morphological type of the candidates.

We think that these two recent papers by Walker (1998) and Strube (1998) are likely to suggest a trend in the evolution of future research on CT-based algorithms for reference processing. We believe that given the substantial difficulties of CT concerning its empirical adequacy, we will see in the future a progressive integration of the more productive insights of CT in systems developed along the lines of the integrative methodology. It is likely that eventual objective attentional factors will be counted and weighted on a par with other non-cognitive-driven preferences in the ranking of antecedent candidates. Moreover, it may turn out that we end up finding compelling to envisage the ranking of antecedent candidates in the integrative method as determining the degree of attentional focusing of each antecedent candidate by the speakers/hearers (apropos vd. Garnham, 1989, or Greene et al., 1992).

That is also the view expressed by Asher and Wada (1988, p.331):

“We do not think that the notion of the focus, the topic or the backward center of the sentences in the discourse prior to the one containing the pronoun is the right one for the analysis of anaphora. It is not clear to us that for every clause there is just one focus or perhaps any focus at all, although it is undeniable that something like a focus exists in many discourses. Unfortunately, test for local focus do not seem to capture a clearly defined notion, and recent attempts to clarify the notion of focus tend to make predictions that go beyond any supporting data from speaker’s intuitions. One simple modification to the concept of focus, however, that seems to solve at least some of the difficulties is to abandon the idea of one local focus in favor of a degree of focus or a degree of how much more salient one discourse referent is than another among the set of syntactically and semantically acceptable, potential antecedents.”

2.4 Summary

This chapter provided an overview of the state of the art in automatic reference processing and anaphor resolution.

Multidisciplinary research on anaphora has unveiled the high complexity of the interpretative process of anaphoric expressions. In order to accommodate the plethora of different factors conditioning this process - morphological, syntactic, semantic, pragmatic, etc. -, an integrative methodology for anaphor resolution was conceived and has received a wide consensus in the literature. Under this methodology, no single type of factor is expected to provide the solution for the resolution of anaphors. Rather, every relevant factor is taken into account and integrated into one of two classes. One of the classes gathers factors that act like filters inasmuch as any antecedent candidate must comply with them for qualifying itself as admissible antecedent. The other class collects factors whose compliance with a given antecedent candidate enhances the chances of this candidate being chosen as the actual antecedent.

Binding constraints, which will continue as the linking thread of the study reported in the remainder of the dissertation, clearly appear on the side of the filters. Given they are conditions on the relative positioning in the grammatical geometry between anaphors and their antecedents, they are thus crucial for sorting out what expressions can enter the set of admissible antecedent candidates of a given anaphor.

The exploitation of the potential of this integrative approach opened many active lines of research. How preferences contribute to the discrimination of antecedent candidates and which scoring scheme is more suited for that purpose; how to evaluate the performance of anaphor resolvers; how far can one go in extending the generality of the approach - to other types of anaphors, to multilinguality, to multimodality, etc.; how much can one improve speed of processing by neglecting some hard to process factors, while ensuring acceptable quality of outcome; and how corpora and statistics-driven approaches may be put to use for enhancing the quality of anaphor resolution: these are just some of the many interesting research issues

motivated by the integrative methodology for anaphor resolution and reviewed in this chapter.

A second major framework for anaphor resolution, based on Centering Theory, was also discussed. In this theory, discourse is seen as comprising segments which differ from each other because, while centering the attention of speakers/hearers on different entities, they are about different topics. Moreover, this theory claims that reference to entities on which attention is focused is expected to be made with weakly informative expressions, such as pronouns.

Detailed research on the implications of the basic tenets of Centering Theory and associated algorithm for anaphor resolution has nevertheless revealed a very weak empirical adequacy and considerable difficulty for this framework. It is hard to conceive how it may be used to resolve anaphoric expressions other than pronouns, which has been deemed as a lack of generality of the theory. Non-local anaphoric links, in which the antecedent occurs in a previous sentence not immediately preceding the one where the relevant pronoun occurs, or intra-sentential anaphoric links, where the pronoun and its antecedent occur in the same sentence, are yet further examples to which the centering-based algorithm has proven difficult to apply.

While trying to overcome these and other difficulties, the literature on centering-based anaphor resolution has put forward some alternative views vis-à-vis the basics of the framework. Interestingly, these innovative steps seem to exhibit a tendency to foster an approximation to the basic assumptions of the integrative framework, which indicates a convergent trend for the integration of these two major methodologies.

3 Empirical Universality

3.1 Universal Constraints	90
Complementary distribution	91
Parameterizable locality	91
Linear order, recess and command	92
Command and grammatical obliqueness	93
Deixis, exemption and point of view	94
Plural and split antecedents, and agreement	96
Reciprocals	97
Definites and functional symmetry	98
Empirical generalizations	99
Universality of constraints	100
Binding gaps	101
3.2 Split Obliqueness and Derived Constructions	102
Toba Batak objective voice	103
Japanese causatives	108
Balinese passive	109
3.3 Branching Obliqueness and Subject-orientedness	111
Non-linear obliqueness	112
Chinese long-distance reflexive	113
Norwegian anaphors	114
Reflexives in Portuguese oblique complements	115
Identity <i>be</i>	117
3.4 Summary	119

As follows from previous chapter, relative positioning constraints on anaphoric relations between nominals and their antecedents, usually known as binding constraints, form one of the most prominent set of filters for anaphor resolution. Given the complexity of the empirical evidence to be tackled and the highly non-trivial puzzles involved, these constraints have been at the center of one of the most active fields of research in the last three decades. This research has been sustained by successive results that delivered increasingly wider empirical generalizations and have disentangled better abstractions for the understanding of the facts. The difficulties that had to be overcome as well as the progress made in the area can be followed *inter alia* in Perlmutter and Soames, 1979, Kuno, 1987, chap.2, Lasnik 1989, chap.1, and in Pollard and Sag, 1994, chap.6.

Our purpose with the present chapter is twofold. On the one hand, we aim to provide an introductory description of what is at present a quite stable set of facts and central results that any definition of binding constraints must be able to handle. On the other hand, we aim to discuss some recent challenges on the universality of binding constraints and to show that they are but the effects of the conjoined occurrence of other collateral linguistic phenomena or constraints together with binding constraints proper.

3.1 Universal Constraints

Complementary distribution

A first central observation with respect to binding constraints is one which brings to light the apparent complementary distribution between the set of admissible antecedents of reflexive pronouns, such as himself, and the set of admissible antecedents of non-reflexive pronouns, such as him. As illustrated below, the anaphoric capacity of reflexives observes a locality constraint inasmuch as reflexives accept as antecedents NPs that are within a certain grammatical circumscription. In contrast, non-reflexives also observe, but in negative terms, such a locality requirement given that they admit antecedents which lie outside this circumscription (coindexation represents the intended anaphoric link):

- (1) a. John_i said [Peter_j shaved himself_{#i/j}].
 a'. John_i said [Peter_j shaved him_{i/#j}].
 b. John_i described [Peter_j's shaving of himself_{#i/j}].
 b'. John_i described [Peter_j's shaving of him_{i/#j}].

Parameterizable locality

The scope of the relevant local circumscription for binding constraints may vary across languages, as discussed, for instance in Manzini and Wexler, 1987, and Dalrymple, 1993. The set of contrasts below illustrates this issue. As shown in (2), in English the antecedents of reflexives occurring as complements of a verbal predicator cannot appear outside the clause boundary which is minimal with respect to the reflexive, i.e. the clause boundary – tensed or not – such that no other clause boundary intervenes between it and the reflexive:

- (2) a. Lee_i said that [John_j ordered [Peter_k to shave himself_{#i/#j/k}].
 b. Lee_i said that [John_j ordered [Peter_k to shave him_{i/j/#k}].

In contrast, as shown in (3) – taken from Manzini and Wexler, 1987, exs.(12),(17), and Hellan, 1991, ex.(12) –, in Icelandic, in order to count as minimal for the sake of defining the local circumscription for binding, the clause boundary has to correspond to a tensed clause:

- (3) a. Haraldur_i segir að [Jói skipaði [Olafu_k að raka si_{#i/j/k}]].
Harald says that Jon ordered Olaf to shave self
- b. Haraldur_i segir að [Jói skipaði [Olafu_k að raka hann_{i/#j/#k}]].
Harald says that Jon ordered Olaf to shave him

Nevertheless, as can be observed from the contrasts between a. and b. in (2) and (3), whatever the range of relevant local grammatical circumscription, the symmetry between reflexives and non-reflexives in terms of admissible antecedents is consistently observed.

Linear order, recess and command

As illustrated with the next two sets of examples, relative linear order between either reflexive or non-reflexive pronouns and their antecedents is not a constraining factor for the possibility of anaphoric links. In examples a., the antecedent precedes the anaphor in terms of linear order, while in examples b., the antecedents are preceded by the anaphors:

- (4) a. Peter_i said that John_j shaved himself_{#i/j}.
b. Peter_i said that, himself_{#i/j}, John_j shaved.
- (5) a. If John_i is around, he_i will do it.
b. If he_i is around, John_i will do it.

Although relative linear order is not determinant as such for the possibility of anaphoric links, not all nominals inside the relevant local circumscription qualify as grammatically admissible antecedents for the reflexives, nor as non-admissible antecedents for non-reflexives. As suggested by the following data, for a nominal to qualify as admissible antecedent of a reflexive, it cannot appear in a given “recess” of the grammatical geometry with respect to the position of the reflexive in that geometry.

- (6) a. Peter_i shaved himself_i.
 b. [Peter_i's] son_j shaved himself_{#i/j}.
 c. [The boy [who knows Peter_i]]_j shaved himself_{#i/j}.

Turning it into positive terms, antecedent and reflexive must be in what was termed in the literature as a command relation, which we will characterize in more detail below.

Before concentrating on the command relation, it is worth noticing that the symmetry between reflexives and non-reflexives in terms of possible antecedents is preserved in these “recesses” of grammatical geometry. This is shown with the constructions below, which involve non-reflexive pronouns, and where the binding possibilities of the examples above with reflexives are inverted:

- (7) a. Peter_i shaved him_{#i}.
 b. [Peter_i's] son_j shaved him_{i/#j}.
 c. [The boy [who knows Peter_i]]_j shaved him_{i/#j}.

Command and grammatical obliqueness

Different definitions of the command relation have been proposed as effective for the purpose of defining relevant recesses in grammatical geometry and adequately constraining binding possibilities.

Some hypotheses concerning the definition of command were based purely on syntactic notions (e.g. Reinhart, 1981, Chomsky, 1981), whereas others were set up on purely semantic terms (e.g. Jackendoff, 1972, Bach and Partee, 1980). All of them have been shown to be empirically inadequate in several places in the literature, and in particular in Pollard and Sag, 1992.

- (8) a. Himself_i, John_i shaved.
 b. John_i shaved himself_i.
 b'. John_i was shaved by himself_i.

Contra the hypothesis that the relevant command relation is based in configurational notions, the example a. above shows that a reflexive may have an

antecedent that does not c-command it. Contra the hypothesis that the relevant command relation is based in a hierarchy of thematic roles, the sentences in b. and b'. (cf. Pollard and Sag, 1994, chap.6, ex.(111)) display grammatical anaphoric links where the reflexive is commanded (b.) and not commanded (b'.) by its antecedent in such hierarchy.

In this dissertation, we adopt a command relation based on the obliqueness hierarchy of grammatical functions, as proposed in Pollard and Sag, 1994, p.253:

- (9) Obliqueness-command (O-command)
 Y locally o-commands Z iff Y is less oblique than Z.
 Y o-commands Z iff Y locally o-commands X dominating Z.

The obliqueness hierarchy of grammatical functions is depicted below, and a discussion on the empirical evidence concerning its existence, as well as on the criteria to identify different grammatical functions and establish their hierarchy can be found in Pollard and Sag, 1987, pp.117ff,175, and in the references therein:

- (10) Obliqueness Hierarchy
 Subject < Object < Indirect Object < Obliques

With such a command relation in place, we can now come back to the examples in (6) and (7). The complementary distribution of the admissible antecedents of reflexives and non-reflexives is thus based both on the notion of locality and the notion of command. A given nominal is thus a grammatically admissible antecedent of a reflexive iff both it is in the relevant local grammatical context of the reflexive and it o-commands the reflexive. Symmetrically, a given nominal is a grammatically admissible antecedent for a non-reflexive iff it is not the case that it is in the relevant local grammatical context of the non-reflexive and o-commands it. These two requirements together thus give shape to what, for short, is usually referred to as the (parameterizable) local domain for binding.

Deixis, exemption and point of view

Parallel to the complementary distribution of corresponding admissible antecedents, there are some other differences with respect to anaphoric properties of reflexives and non-reflexives. One such difference is the fact that, given reflexives cannot take antecedents outside their local domain, they cannot have a deictic value, contrarily to what happens with non-reflexives:

(11)a. # Mary likes himself.

b. Mary likes him.

Given that, on the one hand, admissible antecedents of a reflexive obey a locality requirement, and on the other hand, have to o-command it, an interesting occurrence of reflexives to address is the one where they occur at the bottom of the obliqueness hierarchy. In such cases, there cannot be any antecedent available that complies with these two requirements. The usage of reflexives in these contexts, known as exemption positions, however, is not completely disallowed, as shown by the example below (from Kuno, 1987, ex.(32)a.):

(12) John_i knew that there was [a picture of himself_i] in the post office.

Accordingly, the requirement that the antecedent of a reflexive locally o-commands it – as in (11)a. above – holds iff the reflexive does not occur at the bottom of the relevant obliqueness hierarchy.

Interestingly, in exemption positions, reflexives can then take split antecedents (Zribi-Hertz(89):(42)a.), a possibility disallowed in remaining positions, as it will be discussed in the next subsection:

(13) Mary_i eventually convinced her sister Susan_j that John had better pay visits to everybody except themselves_{i+j}.

Still, even for exempt reflexives, certain requirements other than locality and o-command seem to impinge on their admissible antecedents. These requirements have been discussed in the literature at length (e.g. by Kuno (1987), Sells (1987), Zribi-Hertz (1989), or Safir (1992)), and have been typically identified as non-grammatical factors, having to do mainly with discourse issues and in some cases with processing factors. The most significant ones are compiled in Pollard and Sag, 1994, pp.266ff.

One of the most notorious constraints in this respect seems to be related to what is known as “point of view”, and it is illustrated with the following contrast, taken from Kuno, 1987, ex.(31):

(14) John_i told Mary that [as for himself_i], he wouldn't have to move.

John_i heard from Mary that [as for himself_i] he wouldn't have to move.

Apparently, a reflexive in an exemption position may be assigned an antecedent if the referent of the antecedent is the individual whose perspective is being described in the text.

Since in exemption positions, reflexives are allowed to get a non-local antecedent, in those occurrences they can be replaced by non-reflexives with the same antecedent. Exemption positions are thus positions where the general symmetry between reflexives vs. non-reflexives is exceptionally broken. Accordingly, additional evidence for the relevance of the point of view factor for exempt reflexives can be found, for instance, in semantic contrasts such as the one displayed below, taken from Cantral, 1974, and referred to in Zribi-Hertz, 1989, ex.(28):

(15) The women_i were standing in the background, with the children behind them_i.

The women_i were standing in the background, with the children behind themselves_i.

Both constructions, with the reflexive or with the non-reflexive, are grammatical but are said to possibly describe different situations in accordance with different relative locations between the women and the children. In sentence b., the children are located behind the women from the point of view of the women, that is, behind the women's backs and consequently in the foreground of the picture. In sentence a., by contrast, the children may be located behind the women from the point of view of the speaker, i.e. in the background of the picture.

This interpretive contrast has been taken as further evidence for the fact that exempt reflexives induce readings where the relevant event is described under the point of view of the referents of their antecedents.

Plural and split antecedents, and agreement

Another difference in terms of anaphoric behavior between reflexives and non-reflexives can be found in the antecedents their plural forms can take. As illustrated below, while a plural non-reflexive may be dependent from more than one nominal, which need not be conjoined, a plural reflexive cannot split its antecedents:

(16) John_i told Mary_j they_{i+j} should leave.

John_i told Mary_j about themselves_{i+j}.

As discussed at length in Higginbotham, 1983, when plural non-reflexives take split antecedents, each one of the different nominals acting as antecedents must observe the non-locality requirement impinging over the antecedents of non-reflexives:

(17) # John_i said [Mary_j like them_{i+j}].

Given that the prototypical case of anaphoric relation is the one by means of which coreference is established between the anaphoric items and their antecedents – for the different semantics possibly involved in anaphoric relations see chapter 5, Formal Semantics –, an anaphor typically exhibits the value for person, gender, and number also exhibited by its antecedent. This has led to the widespread assumption that there is morpho-syntactic agreement between anaphors and antecedents. This seems however not to be case, as illustrated by the data below ((18)a. is taken from Pollard and Sag, 1994, ex.(27)):

- (18) a. The Faculty_i voted themselves_i a new raise.
 b. Each senator_i voted against the proposal. They_i were afraid of riots in the streets.

The appearance of agreement therefore seems to result more from semantic similarity between the reference of anaphor and the reference of the antecedent than from a strict morpho-syntactic process, as seems to be the case between, for instance, subject and verb in Portuguese or English.

Reciprocals

Reciprocal expressions, such as each other, have been quite widely assumed to pattern like reflexives with respect to their binding constraints. Starting with Lebeaux, 1983, evidence has been accumulated in the literature, however, which casts serious doubts on that generalization.

In previous works of ours – Branco, 1992, 1993 –, developed mainly with data from Romance languages, we argued at length that reciprocals are not a multiword syntactic unit, but rather a multiclausal construction. Moreover, we argued that this multiclausal construction is a specific elliptical variant of a type of constructions we termed as zoom constructions. In the present dissertation, we build on these results, and consequently exclude reciprocals from any empirical test concerning the anaphoric properties of reflexives.

Definites and functional symmetry

Reflexive and non-reflexive pronouns present, in general, symmetric behavior with respect to the admissible antecedents they can take. In spite of this symmetry, these two classes do not exhaust all the different types of binding constraints that may occur. At least another class of nominals has been identified as showing a third type of behavior in terms of binding constraints. This is the class of non-pronominal definite NPs, which covers proper nouns and definite descriptions, and is usually referred to in the syntactic literature as r-expressions, for referential expressions (a designation that may underlie, it should be noted, the controversial assumption that pronouns do not refer). As illustrated below, in terms of binding capacity, definites do not pattern like reflexives (cf. contrast a./a'.) nor like non-reflexive pronouns (cf. contrast b./b'.).

- (19) a. Peter_i shaved himself_i.
a'.# Peter_i shaved your friend_i.
b. Peter_i said that [Mary shaved him_i].
b'.# Peter_i said that [Mary shaved your friend_i].
c. [The barber [who shaved Peter_i]] said that Mary loves your friend_i.
d. [The girl [who shaved Peter_i]] loves your friend_i.

Taken together, the examples and contrasts above illustrate the fact that a grammatically admissible antecedent of a definite has both to lie outside the relevant local domain of the anaphoric expression, and not to o-command it.

The existence of this third class of nominals somehow contends with the elegance of the symmetry between reflexives and non-reflexives in terms of binding properties. In fact, symmetry is not only an appealing fact in itself in scientific generalizations, but in this specific case, also allows to entertain further analysis based on quite appealing functional arguments, which, more often than not, are not fully spelled out in the literature. In particular, the existence of two symmetric classes of anaphors would lend support, in one way or another, to the view that the ultimate reason for the observed binding behavior would have to do with deep processing and cognitive issues, namely with the fact that the items in each class would evoke cognitive representations arguably located in different types of memory state, or under different types of attentional focus.

Given this, and aiming to preserve the elegance of that bipolar symmetry, some attempts have been put forward to envisage the binding constraint associated with

definites as a derivative or side effect of some other independent principles, typically pragmatic in nature. This would allow one to keep the simpler view that there are only two (symmetric) grammatically determined binding constraints (cf. Reinhart, 1983), or even only one (cf. Levinson, 1991).

In the present dissertation, we follow the work of authors such as Lasnik (1986) and do not force any reduction of the set of binding constraints. As we will argue at length in the next chapter, not only the set of binding constraints is not restricted to the two “symmetric” ones, as while considering a fourth one, the symmetry between all the constraints – not just between two of them – reappears in a completely new light. Moreover, in the last chapter 8 Conclusions: Summary and Outlook, we will speculate on how far linguistic facts and generalizations admit cognitive or functional explanations for binding.

Empirical generalizations

Since Chomsky, 1981, the empirical generalizations that capture binding constraints have been termed, mainly in syntactic literature, as binding principles and collected in what is usually referred to as binding theory. The definition of these principles supposes three accessory devices: (i) the classification of anaphoric nominals into three distinct classes, according to the set of admissible antecedents they can take; (ii) the definition both of command and local command relations; (iii) the assignment of indices to NPs.

As can be seen from the discussion above, the three relevant classes uncovered to date collect reflexives (reflexive pronouns), pronouns (non-reflexive pronouns), and non-pronouns (definite nominals), respectively.

As we follow, as a starting point, the binding theory as defined by Pollard and Sag (1994, chap.6), the command relation we adopt is based on the obliqueness hierarchy and receive the definition stated above in (9). The definition of local command is tuned for languages such as English or Portuguese. It will be adapted in a parametric fashion for languages with a different scope for the locality requirement.

The indices are used as a convention to represent the fact that, if two NPs are coindexed then they are anaphorically related.

For ease of reference to the circumstance where two different NPs are both coindexed and lie in the command relation, the auxiliary notion of o-binding is introduced:

- (20) O-binding (O-freeness)
Y (locally) o-binds Z just in case Y and Z are coindexed and Y (locally) o-commands Z. If Z is not (locally) o-bound, then it is said to be (locally) o-free.

With this in place, empirical generalizations are captured in the so-called binding theory as follows (cf. Pollard and Sag, 1994, p.254):

- (21) Binding Theory
Principle A
A locally o-commanded reflexive must be locally o-bound.
Principle B
A pronoun must be locally o-free.
Principle C
A non-pronoun must be o-free.

Universality of constraints

One of the most notorious characteristics of binding constraints relies on the fact that they have been observed to hold in many different languages even from very disparate language families. This does not mean that continued research on different languages and on an ever increasing set of data does not occasionally bring to light evidence that, at first sight, appears to defy such universality of binding constraints. However, upon further analysis of apparently eccentric new data, it has turned out that binding principles have either been amenable to reshuffling or parameterization so that its empirical adequacy has been continuously widened. Given the successful capacity of binding principles for being upgraded and their high degree of empirical generalization, eventual data that defy their predictions have been typically taken more as challenges for the improvement of binding theory and its universal character than as bold falsifications of its validity.

Given its hypothesized universality, binding theory has thus been viewed also from a different perspective, not as a set of generalizations, but as a set of criteria to identify types of nominals. Under this perspective, it is very interesting to note that binding behavior turns out to be the central criteria to identify, for instance, pronouns in spite of the many diverse forms in which they can surface in different languages. In this respect, the cross-linguistic research conducted by Bresnan

(1998) is quite revealing in the sense that she reports that, taking into account the world languages, a pronoun can surface at least in one of five different forms. It can have a null morphological structure (zero), be a morphological inflection (bound), have a specialized syntactic position and be phonologically bound to a host (clitic), be morphologically free form but atonic (weak), and be free and possibly receive primary sentence accents (NP). In all cases, “what universally characterizes a pronoun is its referential role [...] not its syntactic category” (p.3).

Binding gaps

The fact that binding principles are hypothetically universal does not necessarily mean that every known principle holds in every language. Levinson (1991, pp.133-140) discusses languages that have no items complying with principle A. And take for instance the case of proper nouns in Vietnamese (cf. Lasnik, 1986, pp.153ff):

- (22) a. John_i tin John_i sé tha^vn'g.
John thinks John will win
- b. # John_i thu'óng John_i.
John likes John

The examples above suggest that proper nouns in Vietnamese are not ruled by principle C but rather by principle B.

This should not appear as a surprise and is in line with what is known about cross linguistic implicative generalizations.

For instance, it is known that there are languages that admit relativization only in subject position. But it is also known that if a language admits relativization in object position then it admits relativization also in subject position. Moreover, if a given language admits relativization in indirect object position, then it admits relativization also in object and subject position, and so on and so forth. And similar hierarchies of possibilities seem to exist as regards other linguistic phenomena (cf. the Accessibility Hierarchy of Keenan and Comrie, 1977).

Although, as far as we know, this idea has not been explored in the literature with respect to binding principles, it seems perfectly admissible that there may be a hierarchy of activated binding principles such that if a given principle holds in a certain language then all the other principles that are behind it in that possible hierarchy also hold in that language.

The data presented above on Vietnamese suggest thus that in this language principle C might not be active. If this hypothesis will turn out to be confirmed by further research on Vietnamese, this could count as evidence for the fact that principle C is not in the bottom of the eventual cross-linguistic hierarchy of binding principles.

We will return to this issue in the next chapter, 4 Symmetries and Duality, and in the last chapter, 8 Conclusions: Summary and Outlook.

3.2 Split Obliqueness and Derived Constructions

Given that the definition of binding principles relies on the definition of o-command, which in turn is based on the obliqueness hierarchy, the encoding of information about this hierarchy in the grammatical geometry is crucial for the set up of binding theory. In the HPSG framework, as defined by Pollard and Sag (1994) and presented in detail below in Section 7.1, the hierarchy of grammatical functions associated with each predicator is encoded as the value of the SUBCAT attribute.

The SUBCAT feature has been shown to be a critical point of articulation between highly autonomous principle-based syntax and semantics components (vd. among others Frank and Reyle, 1995). On the one hand, the list in SUBCAT value records information on strict subcategorization properties of the corresponding predicator in what respects the nature and number of its elements. On the other hand, by means of the linear order hierarchy assigned to these elements, syntactic generalizations concerned with word order, binding, alternations, etc. are also registered (Pollard and Sag, 1987, Ch. 5).

Recent research by Iida, Manning, O'Neil and Sag (1994) and Sag and Godard (1994), brought to light evidence that, contrary to what was originally assumed in Pollard and Sag, 1987, 1994, these different syntactic generalizations may not be encoded by one and the same ordering of the subcategorized elements. This issue

was directly addressed by Manning and Sag (1995), who proposed to recast the SUBCAT information into two different lists.

One of the lists displays the subcategorized elements according to an order relevant to their linear surface concatenation. This "valence" list results from appending SUBJ, SPEC and COMPS lists, which are but the result of a previous segmentation of the SUBCAT list proposed by Borsley (1987) and taken up by Pollard and Sag (1994, Ch. 9).

The other list, in turn, orders the subcategorized elements according to a hierarchy relevant at least to set up the binding relations between them. This "argument" list is the value of the new ARG-S feature.

In this connection, a crucial point we want to argue for in the present chapter is that, in order to increase the empirical adequacy of binding theory, the reshuffling of the "original" SUBCAT list must be further extended. In particular, on a par with its segmentation into sublists and its splitting into possibly different obliqueness hierarchies, a branching obliqueness ordering should also be admitted.

Before addressing the issue of branching obliqueness in the next section, however, we will first review the main arguments of Manning and Sag (1995) for splitting the obliqueness hierarchy. While doing so, we intend to discuss how the proposal put forward by these authors can be qualified in order to guarantee that the split obliqueness hierarchies function as a locus for cross-linguistic parameterization for binding theory.

Toba Batak objective voice

The main arguments presented in Manning and Sag (1995) for splitting obliqueness rest on the analysis of two linguistic phenomena: reflexives in objective voice constructions of Toba Batak, a western Austronesian language, and reflexives in Japanese causative constructions.

The pair of sentences in (23) from Toba Batak illustrates the distinction between the objective voice in a., expressed by the *di-* verbal prefix, and the active voice counterpart in b., expressed by the *mang-* prefix (Manning and Sag, 1995, ex.(16)).

- (23) a. *di-ida si Torus si Ria.*
 _{OV-see PM Torus PM Ria}
 Torus sees/saw Ria.

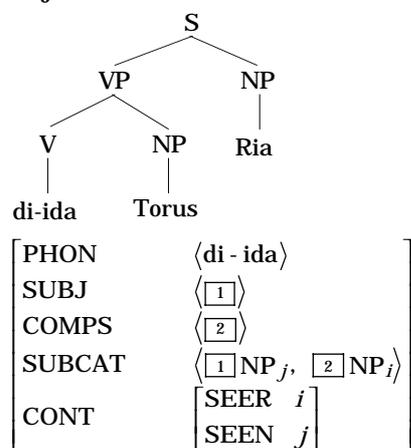
b. mang-ida si Ria si Torus.

AV-see PM Ria PM Torus

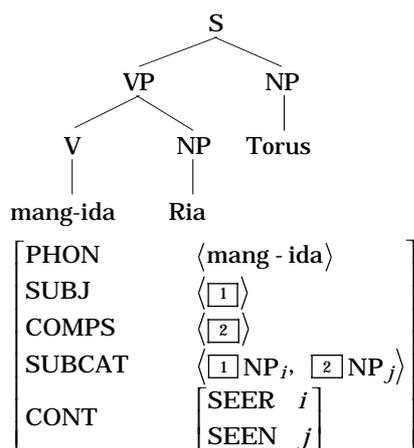
Torus sees/saw Ria.

There is strong evidence (vd. Manning, 1995, pp.2ff, and Manning and Sag, 1995, pp.8ff) that in transitive constructions of Toba Batak, a verb and the following NP form a VP constituent regardless of the voice chosen. Therefore, the constituent structure of (23)a. and b. is presented, respectively, in (24)a. and b., together with the corresponding lexical entry of the verb (the feature-based representation is quite self-explanatory, but for an explicit presentation of the AVM description language, and linguistic significance of the different attributes, see Section 7.1 below).

(24) a. Objective voice: Torus saw Ria.



b. Active voice: Torus saw Ria.



Now, the examples in (25) show all the possible occurrences of one reflexive NP in the basic transitive structures illustrated in (23). In (25)a. and a', the reflexive occurs in objective constructions, respectively, as an immediate constituent of VP and as an immediate constituent of S. The corresponding active constructions are displayed in (25)b. and b'. (Manning and Sag 1995, exs.(22), (23)).

(25) a. # **di-ida diri-na si John.**

[saw himself]_{VP} John

* **himself saw John**

PHON	⟨di - ida⟩				
SUBCAT	⟨NP: <i>npro</i> _j , NP: <i>ana</i> _i ⟩				
CONT	<table style="border: none;"> <tr> <td style="border: none;">SEER</td> <td style="border: none;"><i>i</i></td> </tr> <tr> <td style="border: none;">SEEN</td> <td style="border: none;"><i>j</i></td> </tr> </table>	SEER	<i>i</i>	SEEN	<i>j</i>
SEER	<i>i</i>				
SEEN	<i>j</i>				

a'. **di-ida si John diri-na.**

[saw John]_{VP} himself

John saw himself.

PHON	⟨di - ida⟩				
SUBCAT	⟨NP: <i>ana</i> _j , NP: <i>npro</i> _i ⟩				
CONT	<table style="border: none;"> <tr> <td style="border: none;">SEER</td> <td style="border: none;"><i>i</i></td> </tr> <tr> <td style="border: none;">SEEN</td> <td style="border: none;"><i>j</i></td> </tr> </table>	SEER	<i>i</i>	SEEN	<i>j</i>
SEER	<i>i</i>				
SEEN	<i>j</i>				

b. **mang-ida diri-na si John.**

[saw himself]_{VP} John

John saw himself.

PHON	⟨mang - ida⟩				
SUBCAT	⟨NP: <i>npro</i> _i , NP: <i>ana</i> _j ⟩				
CONT	<table style="border: none;"> <tr> <td style="border: none;">SEER</td> <td style="border: none;"><i>i</i></td> </tr> <tr> <td style="border: none;">SEEN</td> <td style="border: none;"><i>j</i></td> </tr> </table>	SEER	<i>i</i>	SEEN	<i>j</i>
SEER	<i>i</i>				
SEEN	<i>j</i>				

b'.# **mang-ida si John diri-na.**

[saw John]_{VP} himself

* **himself saw John**

PHON	⟨mang - ida⟩				
SUBCAT	⟨NP: <i>ana</i> _j , NP: <i>npro</i> _j ⟩				
CONT	<table style="border: none;"> <tr> <td style="border: none;">SEER</td> <td style="border: none;"><i>i</i></td> </tr> <tr> <td style="border: none;">SEEN</td> <td style="border: none;"><i>j</i></td> </tr> </table>	SEER	<i>i</i>	SEEN	<i>j</i>
SEER	<i>i</i>				
SEEN	<i>j</i>				

The pair of grammatical constructions (25)a'/(25)b. confirms that binding principles cannot be defined in terms of linear word order or a configuration-based command relation. In (25)a'. the antecedent precedes the reflexive, but in (25)b. it is the reflexive that precedes the antecedent; in (25)b. the antecedent c-commands the reflexive, but in (25)a'. it is the other way around.

However, contrary to the assumptions of the binding theory of Pollard and Sag (1994), the definition of binding constraints cannot be based simply on the SUBCAT order. This is made evident by (25)a. and (25)a'., whose grammatical status is not correctly predicted. In (25)a. the reflexive is bound by a less oblique element in the SUBCAT list, in accordance with principle A, but the construction is not acceptable. In (25)a'., the reflexive is bound by a more oblique element in the SUBCAT list, in violation of principle A, but the construction is acceptable.

The solution adopted by Manning and Sag (1995) consists of a three step move: (i) to keep the binding theory unchanged; (ii) to create a new list of subcategorized elements, which is named ARG-S (from argument structure); (iii) to define o-command relations on the basis of the obliqueness hierarchy established on this new list, which may be different from the obliqueness hierarchy established in the SUBCAT list.

Let us then see how this solution works for the problematic examples in (25). In (26) we find the lexical entries of (25) after being reshuffled according to Manning and Sag's proposal (for the sake of readability, the representation of SUBJ and COMPS features is omitted).

(26) a.

PHON	⟨di - ida⟩				
SUBCAT	⟨ 1 <i>j</i> , 2 <i>i</i> ⟩				
ARG - S	⟨ 2 NP: <i>ana</i> , 1 NP: <i>npro</i> ⟩				
CONT	<table style="border-collapse: collapse; margin-left: 5px;"> <tr> <td style="border-right: 1px solid black; padding-right: 5px;">SEER</td> <td style="padding-left: 5px;"><i>i</i></td> </tr> <tr> <td style="border-right: 1px solid black; padding-right: 5px;">SEEN</td> <td style="padding-left: 5px;"><i>j</i></td> </tr> </table>	SEER	<i>i</i>	SEEN	<i>j</i>
SEER	<i>i</i>				
SEEN	<i>j</i>				

a'.

PHON	⟨di - ida⟩				
SUBCAT	⟨ 1 <i>j</i> , 2 <i>i</i> ⟩				
ARG - S	⟨ 2 NP: <i>npro</i> , 1 NP: <i>ana</i> ⟩				
CONT	<table style="border-collapse: collapse; margin-left: 5px;"> <tr> <td style="border-right: 1px solid black; padding-right: 5px;">SEER</td> <td style="padding-left: 5px;"><i>i</i></td> </tr> <tr> <td style="border-right: 1px solid black; padding-right: 5px;">SEEN</td> <td style="padding-left: 5px;"><i>j</i></td> </tr> </table>	SEER	<i>i</i>	SEEN	<i>j</i>
SEER	<i>i</i>				
SEEN	<i>j</i>				

b.

PHON	⟨mang - ida⟩				
SUBCAT	⟨ $\boxed{1}_i, \boxed{2}_j$ ⟩				
ARG - S	⟨ $\boxed{1}$ NP: <i>npro</i> , $\boxed{2}$ NP: <i>ana</i> ⟩				
CONT	<table style="border-collapse: collapse; border: none;"> <tr> <td style="border-right: 1px solid black; padding-right: 5px;">SEER</td> <td style="padding-left: 5px;"><i>i</i></td> </tr> <tr> <td style="border-right: 1px solid black; padding-right: 5px;">SEEN</td> <td style="padding-left: 5px;"><i>j</i></td> </tr> </table>	SEER	<i>i</i>	SEEN	<i>j</i>
SEER	<i>i</i>				
SEEN	<i>j</i>				

b'.

PHON	⟨mang - ida⟩				
SUBCAT	⟨ $\boxed{1}_i, \boxed{2}_j$ ⟩				
ARG - S	⟨ $\boxed{1}$ NP: <i>ana</i> , $\boxed{2}$ NP: <i>npro</i> ⟩				
CONT	<table style="border-collapse: collapse; border: none;"> <tr> <td style="border-right: 1px solid black; padding-right: 5px;">SEER</td> <td style="padding-left: 5px;"><i>i</i></td> </tr> <tr> <td style="border-right: 1px solid black; padding-right: 5px;">SEEN</td> <td style="padding-left: 5px;"><i>j</i></td> </tr> </table>	SEER	<i>i</i>	SEEN	<i>j</i>
SEER	<i>i</i>				
SEEN	<i>j</i>				

It is easy to check that the correct predictions hold if the relevant o-command relations are established on the ARG-S list: the reflexive is now coindexed with a more oblique element in (25)a./ (26)a., and with a less oblique antecedent in (25)a'./ (26)a'.

An interesting perspective under which the otherwise quite stipulative Manning and Sag's solution could be envisaged would stress the fact that, in spite of lexical "processes" for the formation of alternations in Toba Batak (v.z. objective voice constructions), it would be the obliqueness hierarchy of the underlying "pre-processed" predicator that counts for binding.

Under this hypothesis the apparently puzzling data on Icelandic (Sells, 1987, exs.(17)/(18)) involving the active and the passive voice will also receive an immediate account:

- (27) a. Hann_i sagd'í [ad' sig_i vantad'í haefileika].
 he said that self lacked ability
 He_i said that he_i lacked ability.
- b. # Honum_i var sagt [ad' sig_i vantad'í haefileika].
 he was told that self lacked ability
 He_i was told that he_i lacked ability.

In Icelandic the nominal *sig* is a subject-oriented long-distance reflexive (see Section 4.1 for a full discussion on this type of anaphors), i.e. its antecedent can only be in subject position. Nevertheless, in spite of the fact that in both constructions above the antecedents of that nominal are in subject position in the matrix clause, only (27)a., the active voice construction is grammatical. Accordingly, we seem to have here a phenomena possibly similar to the one discussed above concerning the alternation active vs. objective voice in Toba

Batak. As in Toba Batak, in Icelandic the relevant obliqueness hierarchy for binding seems to be the one associated with the “pre-processed” active morphological form of the relevant predicator. This would then explain why the subject of the passive construction cannot be the antecedent of the subject-oriented sig, as that complement is the indirect object, not the subject, in the active voice counterpart.

Japanese causatives

Other linguistic evidence put forward to support the obliqueness split is the behavior of reflexives in Japanese causative constructions, as originally argued by Iida, Manning, O’Neil and Sag (1994).

The analysis of case marking, agreement and word order phenomena in Japanese causatives reveals that this type of construction exhibits properties of a single clause sentence. As to the Japanese reflexive *zibun*, as with English reflexives, it must be locally *o*-bound, with some particulars, as for instance its being subject-oriented, such as the Icelandic sig.

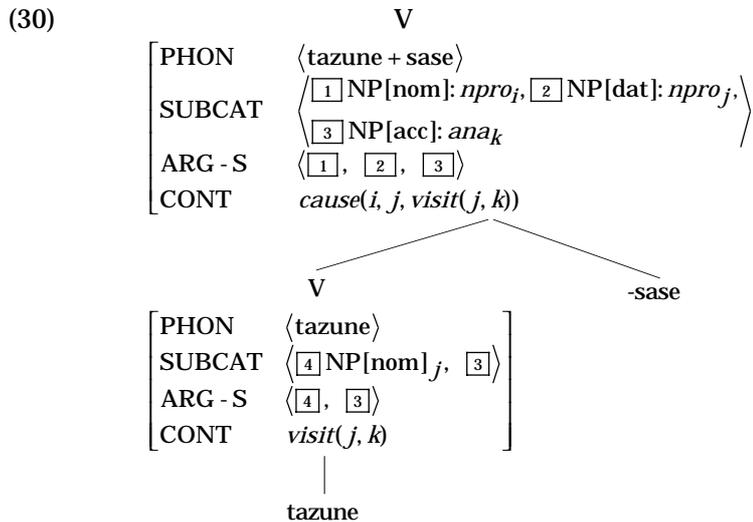
Now, the example of (28) illustrates that, in the context of causatives, *zibun* is not restricted to being bound by the subject of its clause (Manning and Sag, 1994, ex.(44)):

- (28) Taroo_i ga Ziroo_j ni aete *zibun-zisin*_{i/j} o hihans-ase-ta.
Taroo NOM Ziroo DAT purposefully self ACC criticize-CAUS-PAST
Taroo_i purposefully made Ziroo_j criticize himself_{i/j}.

Pronouns also exhibit special behavior in the context of causatives. Contrary to the requirements of principle B, in such contexts pronouns may be bound by an antecedent occurring in the same clause, but only if it is the subject of the causative construction. This is illustrated in (28) (Iida et al., 1994, ex.(17)).

- (29) Taroo_i wa Ziroo_j ni kare_{i/#j} o bengos-ase-ta.
Taroo_i TOP Ziroo_j DAT he_i ACC defend-CAUS-PAST
Taroo_i made Ziroo_j defend him_{i/#j}.

The solution proposed in Iida et al., 1994, to account for the apparently peculiar properties of binding constraints in causatives relies on the assumption that the derived lexical representation of a causative verb, such as *tazune-sase* (Engl.: made somebody visit), has the form sketched in (30), where *tazune* is the verb to visit and *-sase* the causative suffix (Iida et al., 1994, ex.(25)).



Consequently, this solution relies also on the three basic assumptions adopted for the analysis of Toba Batak reflexives: (i) the principles of Binding Theory remain invariant; (ii) a new list of subcategorized elements, termed ARG-S, is adopted; (iii) o-command relations can be defined on the basis of the obliqueness hierarchy established in this new list. Moreover, there is a fourth assumption which proposes that principles A and B should be validated in at least one of the two ARG-S features occurring in the derived lexical entry of a causative verbal form.

Looking at the lexical representation of a causative verb in (30) and the examples (28) and (29), it is easy to check that principle A is satisfied in the lower ARG-S list for the binding Ziroo/himself, where Ziroo is the subject, and in the upper ARG-S for the binding Taroo/himself, where Taroo is now the subject. As to the contrast in (29), principle B is satisfied in the lower ARG-S list, where the pronoun is locally o-free.

Balinese passive

The two solutions above, by Manning and Sag (1995) and Iida et al. (1994) respectively, concerning anaphora in the input and output counterparts of lexical processes may be seen as sharing the common hypothesis that the obliqueness hierarchy is split into an obliqueness hierarchy associated with the input variant of

the relevant predicator and an obliqueness hierarchy associated with the output variant. Nevertheless, these two proposals together are not consistent.

If one adopts the solution proposed by Iida et al. (1994) – binding principles validated in at least one of the two hierarchies –, then incorrect predictions will be made with respect to Toba Batak. In particular, in the objective voice constructions represented in (25)a./a'., contrary to fact, both examples should be expected to be grammatical inasmuch as in construction a'., principle A for reflexives would be validated in the input obliqueness hierarchy (in ARG-S value), and in construction a., it would be validated in the output obliqueness hierarchy (in SUBCAT value).

An interesting way of rescuing the appealing idea of splitting obliqueness to account for data such as those associated with Japanese causatives or Toba Batak active voice constructions is to envisage the obliqueness hierarchy or hierarchies where binding principles are expected to be validated – the lexically underived and/or the lexically derived one – as being subject to parameterization. Under this hypothesis, Toba Batak would be a language where binding constraints are active in the underived obliqueness hierarchy (even if both hierarchies happen to exist and be different, as in objective voice), while Japanese would be a language where binding constraints can be validated in at least one of the two obliqueness hierarchies, (if both happen to exist and be different, as in causative constructions).

An important way of empirically lending support to this hypothesis will be to find a language that exhibits the third logical possibility of parameterization in this respect, namely a language where binding constraints are validated only in the derived obliqueness hierarchy (when it exists).

We believe the data independently discussed by Arka and Wechsler (1996) allow us to conclude that Balinese, an Austronesian language from the Indonesian archipelago, is parameterized for that third type of validation of binding constraints.

The following example (Arka and Wechsler, 1996, ex.(29)b.) is a passive sentence in Balinese. The reflexive in object position can take an antecedent in the subject position of the passive:

- (31) *Ida_i ka-edengin ragane_i ring potreka-ne antuk ajin ida-ne.*
 3sg PASS-show self at picture-DEF by father 3-DEF
 (s)he_i was shown himself/herself_i in the picture by his/her father

However, in a similar passive sentence, the object reflexive cannot take an antecedent in the Agent of passive (by-phrase) position (Arka and Wechsler, 1996, ex.(29)c.):

- (32) ?/# Tiang ka-edengin raganei ring potrekae antuk idai.
 1sg PASS-show self at picture-DEF by 3sg
 I was shown himself/herself_i by him/her_i in the picture

If Balinese was like Toba Batak we would expect (32) to be grammatical because the antecedent locally *o*-commands the reflexive in the underived obliqueness hierarchy (ARG-S value). On the other hand, if Balinese was like Japanese, we would also expect (32) to be grammatical because the reflexive is locally *o*-commanded and the non-reflexive is not locally *o*-commanded in at least one of the two obliqueness hierarchies, specifically in the underived one (ARG-S value). However, the sentence is not grammatical. This can be taken as meaning that in Balinese only the derived obliqueness hierarchy (when it exists) is relevant for validating binding constraints. Accordingly, in (32) the object of the passive is less oblique than the agent of the passive (by-phrase) and the reflexive turns out not to be locally *o*-bound, and the pronoun not to be locally *o*-free, which explains the ungrammaticality of this example.

3.3 Branching Obliqueness and Subject-orientedness

Following our aim to address recent challenges to the universality of binding constraints, there is a second empirical issue deserving scrutiny. It concerns reflexives which can be bound only by a subject, although they must be locally bound in accordance to principle A. Examples of languages with such items are Malayalam and Hindi, from India, Lango from Uganda, Bahasa from Indonesia, Japanese, Korean and Russian (vd. Palmer 1994, p. 100ff, and Manning and Sag, 1995). Example (33) is taken from Lango (Palmer, 1994, p. 101).

- (33) òkélò_i òkwàò àlábâj p̀rÈ kÉnÊ_{i/#j}.
 Okelo asked Alaba about self
 Okelo_i asked Alaba_j about himself_{i/#j}.

The solution put forward in Manning and Sag 1995, ex.(6), to account for this particular sort of reflexives is to formulate a new binding principle, the so called a-subject principle, where an a-subject is defined as the "entity that is first in some ARG-S list":

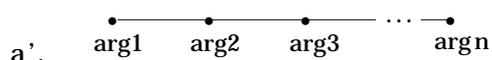
- (34) A-Subject Principle
 Reflexives must be a-subject-bound (in some languages).

Deciding whether the binding theory should include principle A or a-subject principle would depend thus on the language in which it is being considered.

Non-linear obliqueness

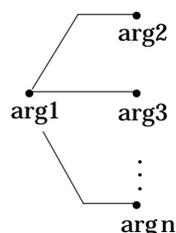
The alternative solution we propose for subject-orientedness does not involve different formulations for binding principles or additional principles. In this solution, binding theory is kept invariant. One simply has to state that, for those languages, such as Lango, which have subject-oriented reflexives, the obliqueness hierarchy is not as it is usually assumed to be, as sketched in (35)a./a', but rather as it is depicted in (35)b./b'. (for simplifying the discussion, we concentrate on ARG-S, and the proposals on this feature are assumed to extend to SUBCAT). The diagrams in a'. and b'. (Hasse diagrams for obliqueness displayed with a rotation of 90° right) are illustrative representations of the orders coded in the feature structures in a. and b., respectively:

- (35) a. [ARG-S $\langle arg1, arg2, arg3, \dots, argn \rangle$]



- (i) [ARG-S $\langle arg1, \{arg2, arg3, \dots, argn\} \rangle$]

b'.



Chinese long-distance reflexive

Let us thus discuss in some detail empirical motivation for this proposal, beginning with the Chinese long-distance reflexive, widely discussed in the literature on long-distance reflexives.

Chinese *ziji* has been taken until recently as a subject-oriented reflexive pronoun which does not obey either principle B or principle A. As illustrated in (36), *ziji* may be bound by an antecedent from outside or inside its local clause, but apparently it cannot be bound by an antecedent which is not a subject (Xue et al. 1994, ex.(2) – but see Cole and Wang, 1996, and the discussion in Section 4.1 below).

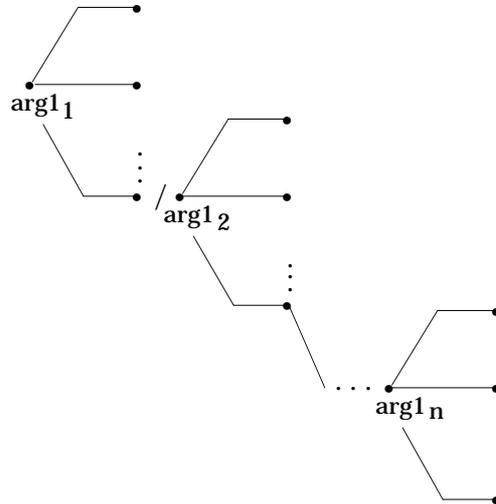
- (36) Zhangsan_i cong Lisi_j chu tingshuo Wangwu_k bu xihuan ziji_{i/#j/k}
 Zhangsan from Lisi place hear Wangwu not like self
 Zhangsan_i heard from Lisi_j [Wangwu_k does not like him_{i/#j}/himself_k].

Xue, Pollard and Sag (1994) discussed at length the properties of this anaphor. These authors elucidated its particulars, namely that *ziji* is inherently animate, and ambiguous between a discourse pronoun and what the authors called a (syntactic) z-pronoun. As a z-pronoun, it is suggested to obey a new specific constraint (Xue et al. 1994, (38)):

- (37) Principle Z
 Z-pronouns must be o-bound.

Nevertheless, these authors offered no solution to account for the fact that syntactic *ziji* is subject-oriented. This solution now follows naturally and immediately from the assumption that in Chinese the elements of obliqueness hierarchy receive the non-linear order of (35)b.. Principle Z alone – to which we will return in much more detail in the next chapter – is now enough to make the correct predictions about *ziji* as soon as the o-command relations are established over the non-linear binding obliqueness hierarchy of multi-clausal sentences displayed in (38), typical of languages with subject-oriented reflexives.

(38)



Any node in the o-command hierarchy is preceded only by subjects because in each (clausal) obliqueness hierarchy, only subjects can be less oblique than any other argument.

Norwegian anaphors

Empirical data that lend further empirical support to the hypothesis on the non-linearity of the obliqueness hierarchy is not confined to Chinese *ziji*, whose subject-orientedness, however, has been challenged by data presented in recent studies, such as Cole and Wuang, 1996. The contrasting behavior of reflexive and non-reflexive pronouns in the context of Norwegian verbs taking an object and an oblique is also highly illuminating in this respect.

Considering the examples provided by Dalrymple (1993, exs.(1.65),(1.75)), taken from Hellan, 1988, it is possible to illustrate the usual symmetry between reflexives and non-reflexives:

- (39) a. Jon_i fortalte meg om seg selv_i.
 Jon told me about self
 Jon told me about himself.

- b. # Jon_i snakker om ham selv_i.
Jon talks about self
Jon talks about him.

The next contrast (Dalrymple, 1993, exs.(1.66),(1.1.74)) is thus an important piece of evidence to support the idea that obliqueness hierarchy branches:

- (40) a. # vi fortalte Jon_i om seg selv_i.
we told Jon about self
We told Jon_i about himself_i.
b. vi fortalte Jon_i om ham selv_i.
we told Jon about self
We told Jon_i about him_i.

Interestingly, the contrast in (39) is reverted in (40). Both the reflexive and non-reflexive pronouns are in oblique position. However, in (40) while the reflexive cannot take an antecedent in object position, that option is possible for the non-reflexive.

What would be a counterexample for the generalizations encoded in principles A and B receives a neat account under the hypothesis that the obliqueness hierarchy may be non-linear. Given this hypothesis, the data above can be understood as resulting from the fact that the subject is less oblique than the object and the oblique complement, but the obliqueness relation branches at that point and the object and the oblique complement are not comparable under that relation. Accordingly, the nominals in oblique position are not o-commanded, and a fortiori not possibly o-bound, by nominals in object position. That makes it possible for the pronoun in oblique position to take an antecedent in the object position, in compliance with principle B, and for the reflexive in that same position not to take an antecedent in object position, in accordance to principle A.

Reflexives in Portuguese oblique complements

Another related problem for the current binding theory comes from Portuguese as the theory fails to deliver the correct predictions for binding patterns involving reflexives in the context of verbs with two oblique complements. One such verb is *falar com A acerca de B* (English: talk to A about B):

- (41) a. A Maria falou com o Pedro acerca do novo Director.
the Maria talked with the Pedro about of_i the new Director
Mary talked to Pedro about the new Director.
- a'. A Maria falou acerca do novo Director com o Pedro.
the Maria talked about of_i the new Director with the Pedro
Mary talked about the new Director to Pedro.

Given the linear order for the ARG-S value typically assumed for binding theory, it is predicted that if a reflexive occurring as the oblique complement Y is grammatically bound by an antecedent occurring as the oblique complement X, then X is less oblique than Y. Moreover, it is also predicted that the reversed binding configuration, where the reflexive would occur as the oblique complement X, will be ungrammatical. These predictions are sketched in the following contrast schemata, where the Portuguese *si próprio* is a reflexive ruled by principle A (for more details on *si próprio* see the next chapter, 4 Symmetries and Duality):

- (42) a. A Maria falou [PREP-X o Pedro_i]_{OBL-X} [PREP-Y *si próprio*_i]_{OBL-Y}.
b. # A Maria falou [PREP-X *si próprio*_i]_{OBL-X} [PREP-Y o Pedro_i]_{OBL-Y}.

The failure of these predictions is presented in (43), which depicts the instantiation of schemata (42). In (43)a./a', PREP-X is made equal to *com* (English: to) and PREP-Y to *acerca de* (English: about); in (43)b./b' it is the opposite. The pairs a./a' and b./b' simply exhibit different surface orders for the oblique complements in the sentence, a grammatical possibility illustrated in (41)a./a'. In all examples the binding of the reflexive is ungrammatical.

- (43) a. # A Maria falou com o Pedro_i acerca de *si próprio*_i.
Maria talked to Pedro_i about himself_i
- a'. # A Maria falou acerca de *si próprio*_i com o Pedro_i.
Maria talked about himself_i to Pedro_i
- b. # A Maria falou consigo *próprio*_i acerca do Pedro_i.
Maria talked to himself_i about Pedro_i
- b'. # A Maria falou acerca do Pedro_i consigo *próprio*_i.
Maria talked about Pedro_i to himself_i

This is another puzzle for the current binding theory which can be neatly solved with a branching hierarchy for the ARG-S value. In particular, the data presented above can be adequately accounted for, if the ARG-S feature of verbs like *falar_com_acerca* is as follows, where the two PP complements do not precede each

other, and a reflexive occurring in one of them cannot be bound by an expression occurring in the other:

- (44) a. [ARG - S ⟨NP, {PP[com]: *npro* PP[com] *cerca de*: *ana*}⟩]
 b. [ARG - S ⟨NP, {PP[com]: *ana* PP[com] *cerca de*: *npro*}⟩]

In this connection it is worth noting that effects of branching obliqueness seem not to be visible only in terms of binding facts but also, at least, in terms of linear order constraints. In Portuguese, and putting aside so called “heavy” objects, the linear order of a post verbal object followed by an indirect object seem to be stable (vd. contrast a./b.), unless it is affected by specific prosody (vd. c.):

- (45) a. O Pedro apresentou o João à Maria.
 Pedro presented João to Mary.
 b. # O Pedro apresentou à Maria o João.
 Pedro presented to Mary João.
 O Pedro apresentou, à Maria, o João.
 Pedro presented, to_the Mary, the João

In contrast, the linear order among two post verbal oblique complements seem not to be fixed:

- (46) a. A Maria falou com o Pedro acerca do João.
 Maria talked with Pedro about João.
 b. A Maria falou acerca do João com o Pedro.
 Maria talked about João with Pedro.

Although it is out of the scope of the present dissertation to uncover all the implications of branching obliqueness, in particular with respect to linear precedence constraints, it is nevertheless interesting to note that the contrast between (45)/(46) may be due to the difference between *apresentar* and *falar* in terms of the two associated obliqueness hierarchies. While the first is linear, the second is non-linear.

Identity *be*

The fact that some anaphors such as the Norwegian *seg selv* are systematically subject-oriented may lead one to hypothesize that verbs having linear or non-linear

obliqueness hierarchy associated with them is a language parametric issue. However, Portuguese constructions involving verbs with two oblique complements point in a different direction. Given that the Portuguese anaphoric nominals are not systematically subject-oriented, and that this kind of behavior only residually shows up in the referred type of verbs, this leads to think that linearity, or non-linearity, of obliqueness hierarchies may be specific to particular verbs, or classes of verbs.

Striking examples giving empirical support to this idea can be found in English constructions – and presumably in other languages – involving the verb to be in an identity reading (the examples below are taken from Safir, 1992, p.10):

- (47) a. John_i is him_i.
b. He_i is John_i.

Under the assumption that the obliqueness hierarchy associated with the verb to be is non-linear, the puzzling data above –systematically referred to in the literature as a residual problem for the generalizations captured in the binding theory – can be naturally explained. As neither of the two complements o-commands the other, neither principle B, in a. above, nor principle C, in b., is violated.

An interesting prediction stemming from the assumption that the identity to be has a non-linear obliqueness associated with it is that a reflexive in the accusative position would enter an exemption position (the example below is taken from Safir, 1992, ex.(16)):

- (48) John is (not) himself (again).

Accordingly, as discussed above with respect to the “point of view” factor, the effects of that special occurrence are expected to be somehow noticed in interpretive terms. This is in fact confirmed by the comments of Safir (1992, p.10): “Note, however, that while [the above construction] has a grammatical reading it is not understood as an identity statement, i.e., that John is in fact John [...] Rather [it] is understood to have the idiomatic meaning that John is (or is not) healthy or is (or is not) acting once again as he usually does.”

3.4 Summary

After a general introductory presentation of the so-called binding theory, we discussed two major challenges to the universality of binding constraints. First, we discussed some apparent counterexamples relating to lexically derived constructions (e.g. passives, causatives, objectives) where binding principles appeared to support incorrect predictions. Second, we discussed some other problematic cases where the anaphoric nominals presented what has been termed in the literature as subject-oriented behavior, roughly speaking a behavior such that the binding principles seem to unfold their predictions correctly only when the relevant antecedent is found in the subject position.

In both of these cases, we argued that the problematic issues are but the effect of the interference of other linguistic phenomena or constraints that are collateral to the binding constraints proper. We lend empirical support to the claim that after these collateral constraints have been disentangled from the binding constraints and their effect has been abstracted away from the problematic data, it is still possible to consider the generalizations captured in binding constraints as valid.

In the first case, we argued that, with respect to lexically derived constructions, the locus for validation of binding constraints is subject to parametric setup across languages. Building on an initial proposal by Manning and Sag (1995), we assumed that obliqueness hierarchy is split into two obliqueness relations in a lexically derived construction. One of the hierarchies corresponds to the derived construction. The other corresponds to its lexically underived counterpart. Under this assumption, we have argued for the hypothesis that, depending on the language at stake, binding constraints hold at the underived obliqueness hierarchy (e.g. Toba Batak) at the derived one (e.g. Balinese), or at least at one of them (e.g. Japanese).

In the second case, we argued that the problematic issues involving subject-orientedness result from the fact that the obliqueness hierarchy is not necessarily linear, and that it may branch. In some languages, such as Chinese or Norwegian, this branching seems to be systematic, and all predicators are associated with branching obliqueness hierarchies, the bottom of which is the subject. In other languages, such as Portuguese or English, predicators may differ with respect to the branching of the respective obliqueness relation.

4 Symmetries and Duality

4.1 Long-distance Reflexives	122
The Portuguese long-distance reflexive	124
World reflexivity	132
4.2 New Symmetries with Four Constraints	135
Long-distance exemption	135
Principle Z	136
The binding square of opposition	138
4.3 The Quantificational Structure of Binding Constraints	139
Natural language quantification	140
Phase quantification ingredients in binding	143
Binding phase quantifiers	144
4.4 New Insights into Binding	146
Exemption	146
Gaps	146
Implied epistemic twists	146
Integration in grammar and in reference processing	146
4.5 Summary	146

Long-distance reflexives have been a major focus of inquiry for theories of binding based on the existence of three binding constraints. These reflexives are expressions that have to get their interpretation from suitable antecedents occurring in the sentence, either inside or outside the typical local domain of (short-distance) reflexives. Such constraint is not accounted for by any of the three "classic" binding principles. Notice that principle A requires o-binding by a suitable antecedent occurring in the relevant local domain, and principles B and C have to do with o-freeness.

In what follows, we develop a study of long-distance reflexives where we argue for a formal account of the binding constraint associated with this fourth class of anaphoric nominals. We will discuss how this constraint consistently fits into the existing binding theory based on three principles and comment on its far reaching consequences both for a radically new conception of binding and for seemingly important new insights into the formal semantics of nominal expressions.

4.1 Long-distance Reflexives

The continued insistence on taking distributional symmetry between admissible antecedents of (short-distance) reflexives and pronouns as the central fact that Binding Theory should mostly account for has had considerable impact – both theoretically and methodologically – on the shape of the mainstream attempts to deal with long-distance reflexives. In accordance with the unique central role

assigned to that symmetry, the phenomenon of long-distance reflexives has somehow been given a sort of marginal or derived status.

In Government and Binding (GB) framework (Chomsky, 1981, 1986), for instance, the anaphoric link between a long-distance reflexive and its antecedent has been taken as being but a successive-cyclic association of short-distance anaphoric links. Examples of this type of approach can be found in Cole and Sung, 1994 (as well as in the many references cited therein), according to which the long-distance reflexives undergo head movement to Infl (a hypothesized item corresponding to inflection features) at the level of LF (Logical Form); in Huang and Tang, 1991, according to which there is no head movement but adjunction to IP (GB construct corresponding to an Inflection Phrase); in Progovac, 1993, according to which the movement is replaced by coindexation with the Agr node. All these short-distance links support a long-distance anaphoric relation by means of recursive concatenation.

This research perspective has had the non-negligible side effect of funneling attention to a specific set of empirical correlations which, in turn, became the standard empirical touchstone for the research on long-distance reflexives. However different the several alternative proposals may be in their details, they all share the common stance that the central facts to be accounted for in studying this issue gravitate around the correlation between so-called morphologically simple anaphors, long-distance binding, subject-orientedness and binding blocking by intervening subjects. In their essential aspects, the different accounts run like this: the morphologically simplex anaphors have some sort of morphological deficit that must be supplemented by some kind of local "link" (e.g. movement, coindexing, etc.) to the Inflection of the local subject (this explains subject-orientedness); links of the same sort across different clauses successively subordinated may be connected (this explains long-distance); in some languages, such as Chinese, the concatenation of these links is interrupted when there is an upwards subject which does not support the relevant kind of link (this explains the blocking of long-distance binding by intervening subjects).

Taking aside the blocking effect, which has been thoroughly discussed only for the Chinese language, the relevant correlations, assumed to be universal, may thus be conspicuously stated in the following table:

	Subject-oriented	Not Subject-oriented	Morphological simplicity	Morphological complexity
Short-distance reflexives		X		X
Long-distance reflexives	X		X	

The Portuguese long-distance reflexive

In this connection an interesting discovery our research has led us to is that Portuguese has a long-distance reflexive which does not pattern as expected from the correlations above.

In Portuguese *si próprio* is the third person short-distance reflexive and *ele* is the third person pronoun. It is well known that their behavior as anaphoric expressions follows closely the behavior of English *himself* and *he*, respectively. As we are going to show, what is new is that, on a par with these expressions, Portuguese has also the long-distance reflexive *ele próprio* (lit.: *he own*).

Other readings of *próprio*

Before, however, empirically testing out the anaphoric properties of this phrase, it should be noted that it includes the expression *próprio*. In particular, it is important to note the specific semantic properties of this subexpression, which as we will discuss below, sometimes have to be disentangled from the properties of *ele próprio*.

As made evident by the translation into English of the sentence below, *próprio* may have a reading close to the meaning of English *even*:

- (1) O próprio presidente morreu no acidente de aviação.
the PRÓPRIO president died in_the crash of aviation
Even the president died in the plane crash.

Interestingly, for reasons which would be beyond the scope of this dissertation to examine, this reading seems to observe restrictions with respect to positions other than subject, as illustrated by the following ungrammatical construction:

- (2) * A Maria gosta do próprio presidente.
the Maria likes of_the PRÓPRIO president

Apparently, this restriction concerns only the “even” reading of *próprio*. In fact, when *próprio* enters a context where it can express a possessive reading, an NP which includes *próprio* is acceptable in object position:

- (3) A Maria matou o próprio pai.
the Maria killed the PRÓPRIO father
Maria even killed her own father.

The possessive reading of *próprio* was studied by Brito (1983), who built on similar research by Giorgi (1983) on the Italian *proprio*. That study was meant to show the long-distance ability of *próprio*. However, as the examples considered involved only *próprio* in possessive positions such as the one illustrated above, they were mostly restricted to exempt occurrences of this expression. Accordingly, no general conclusion as regards its binding properties may be safely drawn from these data, and more importantly for our present concerns, no conclusions can be reached as regards the binding properties of *ele próprio*.

In the following, we will likewise not focus on any non-reflexive readings of *ele próprio*. In particular, we will disregard the use of *ele próprio* which seems to be identical to the use of the English expression *himself* that is characterized by Hintikka (1970) as filtering eventual attributive interpretations of the NPs it happens to be associated with. This reading of *ele próprio* is exemplified below with the contrast between the possible interpretations of (4)a. — known as referential and attributive uses in Donnellan's (1966) terminology, or extensional and intensional readings in other terminologies —, and the preferential extensional interpretation of (4)b., where *ele próprio* occurs:

- (4) a. O presidente eliminou o líder dos dissidentes.
the president eliminated the leader of_ the dissidents
 The president eliminated the leader of dissidents.
- b. O presidente eliminou *ele próprio* o líder dos dissidentes.
the president eliminated he PRÓPRIO the leader of_ the dissidents
 The president himself eliminated the leader of dissidents.

In syntactic terms this non-reflexive, “disintensionalizer” use of *ele próprio* seems to be restricted to its occurrences in non-argumental positions, as in the example above. In this respect, it is interesting to observe that this expression under such reading cannot be grammatically replaced by any other type of nominals, or that no other type of nominal expression seems to be able to carry over its capacity of filtering intensional interpretations of other clause mate NPs:

- (5) a. O presidente eliminou *ele próprio* o líder dos dissidentes.
the president eliminated he PRÓPRIO the leader of_ the dissidents
- b. * O presidente eliminou *ele* o líder dos dissidentes.
the president eliminated he the leader of_ the dissidents

- c. * O presidente eliminou si próprio o líder dos dissidentes.
the president eliminated SI PRÓPRIO the leader of _the dissidents
- d. * O presidente eliminou o presidente o líder dos dissidentes.
the president eliminated the president the leader of _the dissidents
- e. * O presidente eliminou Bokassa o líder dos dissidentes.
(in the intended reading, where Bokassa is the president, not the leader of dissidents)
the president eliminated Mobutu the leader of _the dissidents

Analyzing the reflexive *ele próprio*

Focusing on the anaphoric value of *ele próprio*, the contrast in (6) illustrates that, as with short-distance anaphors, this nominal is an anaphoric expression which requires an antecedent:

- (6) O Carlos_i gosta dele próprio_{#j/i}.
the Carlos likes of _he PRÓPRIO
Carlos likes himself.

Contrast (7)a./(7)b., in turn, shows that *ele próprio* must be o-bound. (7)a./b. illustrates the parallelism between *ele próprio* and the short-distance reflexive, and (7)b./c. the difference between *ele próprio* and the pronoun, with respect to the requirement of o-binding by the antecedent.

- (7) a. # [As pessoas que falaram com a Ana_i] gostam de si própria_i.
the people who talked with the Ana like of SI PRÓPRIO
[People who talked with Ana_i] like herself_i.
- b. # [As pessoas que falaram com a Ana_i] gostam dela própria_i.
the people who talked with the Ana like of _she PRÓPRIO
[People who talked with Ana_i] like herself_i.
- c. [As pessoas que falaram com a Ana_i] gostam dela_i.
the people who talked with the Ana like of _she
[People who talked with Ana_i] like her_i.

Examples from (8), in turn, illustrate the long-distance ability of *ele próprio*:

- (8) a. O Pedro_i convenceu a Ana de [que o Carlos_j gosta dele próprio_{i/j}].
the Pedro convinced the Ana of that the Carlos likes of _he PRÓPRIO
Pedro_i convinced Ana [that Carlos_j likes him_i/himself_j].

- b. O João_i disse-me [que tu achas [que o Carlos_j gosta dele próprio_{i/j}]].
 the João told me that you think that the Carlos likes of_he PRÓPRIO
 João_i told me [that you think [Carlos_j likes him_i/himself_j]].

And the contrast in (9) is meant to reinforce the evidence for the long-distance anaphoric behavior of *ele próprio*, as it shows that, also when coindexed with an antecedent outside the local domain, *ele próprio* must be o-bound, contrarily to what happens as regards pronouns.

- (9) a. # [O apartamento [que o Carlos ofereceu à Ana_i]] revela que ele pensa nela
própria_i.
 the apartment that the Carlos offered to_the Ana reveals that he thinks in_she PRÓPRIO
 [The apartment [that Carlos offered to Ana_i]] reveals that he cares about
 herself_i.
- b. [O apartamento [que o Carlos ofereceu à Ana_i]] revela que ele pensa nela_i.
 the apartment that the Carlos offered to_the Ana reveals that he thinks in_she
 [The apartment [that Carlos offered to Ana_i]] reveals that he cares about her_i.

In this respect, it is interesting to note that interference of the other readings of the subexpression *próprio* may occur. In fact, when we consider contexts where the “even” or the possessive readings are excluded, as with the ones above and the construction (10)a. below, *ele próprio* strictly behaves like a long-distance reflexive.

- (10) a. # [A rapariga [que viu o Pedro_i na festa]] disse-me que a Maria gosta dele
próprio_i.
 the girl who saw the Pedro in_the party told-me that the Maria likes of_ELE PRÓPRIO
- b. [A rapariga [que viu o Pedro_i na festa]] disse-me que ele_i próprio gosta da
 Maria.
 the girl who saw the Pedro in_the party told-me that he PRÓPRIO likes of_Maria
 The girl who saw Pedro in the party told me that he likes Maria.

However, when considering constructions such as (10)b., where *ele próprio* appears in subject position of an embedded clause and the “even” reading is not disallowed, this reading seems to rescue the otherwise ungrammatical occurrence of *ele próprio* as long-distance reflexive, once it is not o-commanded (cf. (9)a. and (10)a.).

Clitic doubling

Data collected in (6)-(9) support standard tests for checking the binding properties of long-distance reflexives. As far as Portuguese is concerned, there is another test that can be done. As this language has direct object clitic doubling, one should check which kind of clitic (anaphoric or pronominal) the phrase containing *ele próprio* can double:

- (11) a. O Pedro_i viu-se a si próprio_i no espelho.
the Pedro saw-CLIT.REFLEX to SI PRÓPRIO in_the mirror
Pedro_i saw himself_i in the mirror.
- b. O Pedro_i viu-se a ele próprio_i no espelho.
the Pedro saw-CLIT.REFLEX to he PRÓPRIO in_the mirror
Pedro_i saw himself_i in the mirror.
- c. # [A mãe do Pedro_i] viu-se a ele próprio_i no espelho.
the mother of_the Pedro saw-CLIT.REFLEX to he PRÓPRIO in_the mirror
[Pedro's_i mother] saw himself_i in the mirror.
- (12) a. [O pai da Ana_i] viu-a a ela_i no espelho.
the father of_the Ana saw-CLIT.PRON to she in_the mirror
[Ana's_i father] saw her_i in the mirror.
- b. [O pai da Ana_i] viu-a a ela própria_i no espelho.
the father of_the Ana saw-CLIT.PRON to she PRÓPRIO in_the mirror
[Ana's_i father] saw her_i in the mirror.

The above data show that *ele próprio* can double both kind of clitics, reflexive or pronominal. Interestingly, it is apparent that *ele próprio* assumes reflexive behavior when it doubles reflexive clitics (cf. contrasts in (11)), and that it assumes a pronominal behavior when it doubles pronominal clitics (cf. (12)), which shows that the properties of the clitics somehow prevail over the properties of *ele próprio*.

Interestingly, a similar phenomenon seems to be observed in Sinhala, an Indo-Aryan language from Sri Lanka. In Sinhala, instead of a clitic, there is a verbal prefix for reflexivity, *gannāwa*, whose free form corresponds to the English verb to take. As in Portuguese cliticization, reflexive objects (with the Sinhala reflexive nominal *taman*) are allowed only when they double reflexivized verbs (cf. Henadeerage, 1998, exs.(5),(6)):

- (13) a. Sryia Gune-wā tuwaalā kāla.
Sryia Gune-ACC injury_make.Past
Sryia hurt Gune.

- a'. # Sryia_i taman-wà_i tuwaalà_kàla.
Srya self-ACC injury_make.Past
Srya hurt herself.
- b. # Sryia Gune-wà tuwaalà_kàrà_gatta.
Srya Gune-ACC injury_make. Past.PresPart_take
Srya hurt Gune.
- b'. Sryia_i taman-wà_i tuwaalà_kàrà_gatta.
Srya self-ACC injury_make.Past.PresPart_take
Srya hurt herself.

The interesting point to note is that, according to the data provided in Henadeerage, 1998, Sinhala has the pronoun *eyaa* (vd. (14)a.). However, when doubling the reflexive verbal prefix (vd. (14)b.), this pronoun can act only as a reflexive:

- (14) a. Lamàya_i eyaa-wà_i#/j tuwaalà_kàla.
child 3Sing-Acc injury_make.Past
The child_i injured #himself_i/him_j.
- b. Lamàya_i eyaa-wà_i#/j tuwaalà_kàrà_gatta.
child 3Sing-Acc injury_make.Past.PresPart_take
The child_i injured himself_i/#him_j.

Coming back to Portuguese, although the overriding phenomenon now reported does not help supporting any claim about the reflexive or pronominal nature of *ele próprio*, it is worth noting that even when doubling pronominal clitics, if o-commanded, *ele próprio* keeps its inability to support deictic reference:

- (15) a. O Pedro viu-a a ela no espelho.
the Pedro saw-CLIT.PRON to she in_the mirror
Pedro saw her in the mirror.
- b. # O Pedro viu-a a ela própria no espelho.
the Pedro saw-CLIT.PRON to she PRÓPRIO in_the mirror
Pedro saw herself in the mirror.

Non-subject-orientedness

Having argued that *ele próprio* is a long-distance reflexive, we will now turn to its distinctive feature of not being subject-oriented. Examples in (16) illustrate that this expression may have antecedents which are not subjects:

- (16) a. O Pedro descreveu a Maria_i a ela própria_i.
the Pedro described the Maria to she PRÓPRIO
Pedro described Maria to herself.
- b. O Pedro convenceu a Ana_i de que o Carlos gosta dela própria_i.
the Pedro convinced the Ana of that the Carlos likes of _she PRÓPRIO
Pedro convinced Ana_i that Carlos likes her_i.
- c. O Pedro disse à Ana_i que o Carlos gosta dela própria_i.
the Pedro said to _the Ana that the Carlos likes of _she PRÓPRIO
Pedro said to Ana_i that Carlos likes her_i.

The data collected until now show thus that there is a long-distance reflexive in Portuguese which, contrarily to the correlation assumed to be universal by GB accounts of long-distance reflexives, is morphologically complex (with overt full inflection paradigm) and is not subject-oriented.

No blocking effect

Another correlation that has been assumed to hold with respect to long-distance reflexives is that they conform to so-called blocking effects in what concerns the possible upwards antecedents. This correlation has been hypothesized mainly in view of data involving the long-distance reflexive *ziji*, from Chinese.

Since Tang, 1989, it has been noticed in the literature that, with respect to *ziji*, a higher subject cannot be its antecedent if an intervening subject differs in person from *ziji*. This is illustrated by the following contrast:

- (17) a. Zhangsan_i zhidao [Lisi_j renwei [Wangwu_k zui xihuan zij_{i/j/k}]]
Zhangsan know [Lisi think [Wangwu most like self]]
Zhangsan_i knows that Lisi_j thinks that Wangwu_k likes him_{i/j}/himself_k
most.
- b. Zhangsan_i zhidao [ni_j renwei [Wangwu_k zui xihuan zij_{i/#j/k}]]
Zhangsan know [you think [Wangwu most like self]]
Zhangsan_i knows that you_j think that Wangwu_k likes him_{#i/#j}/himself_k
most.

Building similar pairs of constructions involving *ele próprio* shows, however, that it is a long-distance reflexive which is not affected by such kind of blocking effects due to intervening subjects with different morphological features:

- (18) a. O Pedro_i sabe que o João_j disse que o Carlos_k não gosta dele próprio_{i/j/k}.
 the Pedro knows that the João said that the Carlos not likes of _him PRÓPRIO
 Pedro knows João said Carlos doesn't like him_{i/j}/himself_k.
- b. O Pedro_i sabe que tu disseste que o Carlos_k não gosta dele próprio_{i/k}.
 the Pedro knows that you said that the Carlos not likes of _him PRÓPRIO
 Pedro knows you said Carlos doesn't like him_i/himself_k.

This lends empirical support to the view that, also as regards blocking effects, there is no essential correlation between this phenomenon and long-distance reflexivity, arguably orthogonal to it.

This dissociation may be indirectly supported by other data where “inverted” blocking effects can be observed as regards, not long-distance reflexives, but fully inflected short-distance reflexives in exemption positions. That is the case illustrated in Pollard and Sag, 1994, exs.(86)a.,(87)d., concerning the English *himself*, where the “intervention” of a subject with identical features of person, number and gender block the anaphoric link with an upwards antecedent:

- (19) a. # Bill_i remembered that Tom saw [a picture of himself_i] in the post office.
 b. Bill_i suspected that the silence meant that [a picture of himself_i] would soon be on the post office wall.

Therefore, we again find another contribution for dissociating long-distance reflexives from phenomena that have been seen as essentially correlated to it, from blocking effects, in the present case.

Long distance reflexives in other Romance languages

The existence of a long-distance reflexive seems not to be unique among languages of the same language family of Portuguese. Taking the data presented by ZribiHertz (1980) in another Romance language, in particular about the French *lui-même*, it is very likely that this expression is a long-distance reflexive as well. Although the examples provided do not allow exhaustive testing, they nevertheless seem to suggest some properties typical of long distance reflexives.

For instance, as with reflexives, *lui-même* cannot be used deictically (p.152):

- (20) a. J'ai honte d'elle.
 I have shame of her
 I am ashamed of her.

- b. # J'ai honte d'elle-même.
I have shame of ELLE-MÊME

But, similarly with reflexives, in exempt position *lui-même* can be used deictically (p.152):

- (21) Lui-même m'avait dit qu'il viendrait.
LUI-MÊME me had told that he would_come
He had told me that he would come

However, unlike short-distance reflexives, and like Portuguese *ele próprio*, it can be bound by a non-local antecedent (p. 150):

- (22) Elle_i veut [que je parle d'elle-même_i].
she want that I talk of ELLE-MÊME
She_i wants that I talk of her_i.

And the anaphoric link with its antecedent seems to observe an o-binding requirement since, in the ungrammatical construction (23) *toi* does not o-command *toi-même* (p.159):

- (23) # J'ai analysé pour toi_i [une description détaillée de toi-même_i].
I have analyzed for you a description detailed of TOI-MÊME

Besides the French *lui-même*, also with respect to the Italian *sè*, it has been suggested in studies such as Giorgi, 1983, Manzini and Wexler, 1987, or Reuland, 1990 that it may be a long-distance reflexive. An exhaustive testing on its binding properties, however, was not systematically entertained, enabling one to establish solid results on this particular point.

World reflexivity

The question which follows naturally now and deserves subsequent scrutiny is whether the Portuguese *ele próprio*, and most probably the French *lui-même*, are isolated exceptions to the "standard" correlation assumed in mainstream accounts of long-distance anaphora between morphological complexity and subject-orientedness.

Long-distance reflexives

The long-distance reflexive *sig* of Icelandic provides a good example of the "standard" correlation: it is morphologically simple and it is subject-oriented (Cole and Sung, 1994, ex.(11)):

- (24) Jón_i sagdi Maríu_j að þú elskadir sig_{i/#j}.
 Jon told Maria that you loved-SUB self
 Jon_i told Maria_j that you loved him_{i/#j}.

Browsing the available literature on long-distance reflexives, it is not hard to find several counterexamples to this correlation in line with the Portuguese one.

In Finnish the long-distance reflexive *hän itse* is subject-oriented but it is morphologically complex (van Steenberghe, 1991, ex.(11) – note that, as opposed to Portuguese, in Finnish the long-distance ability seems to be possible only across non-tensed clauses successively subordinated):

- (25) Pekka_i sanoi Jusille_j Mattin katsovan häntä itseään_{i/#j}
 Pekka said Jussi Mattin-GEN watch-PTC-GEN he self-POSS
 Pekka_i said to Jussi_j that Matti watched him_{i/#j}.

Chinese *ziji* is morphologically simple but, taking into account data recently elicited, it has not always to have a subject as antecedent (Cole and Wang, 1996, ex.(4)):

- (26) Zhangsan_i yiwei Lisi_j hui ba n_i ling hui ziji_{i/j/k} de jia.
 Zhangsan think Lisi will BA you lead back self DE home
 Zhangsan_i thought Lisi_j would take you_k back to his/your home_{i/j/k}.

Together with *ele próprio*, the above long-distance reflexives actually exhibit all the possible correlations between morphological complexity/simplicity and subject/non-subject-orientedness:

	Subject-oriented	Not Subject-oriented	Morphological simplicity	Morphological complexity
Icelandic	X		X	
Finnish	X			X
Chinese (?)		X	X	
Portuguese		X		X
French (?)				X

Short-distance reflexives

Coming now to short-distance reflexives, English himself is the classic illustration of the "standard" correlation: It is morphologically complex and is not subject-oriented. But also for this type of reflexives, a search on the available literature reveals that the possibilities are not confined to that setup.

	Subject-oriented	Not Subject-oriented	Morphological simplicity	Morphological complexity
English		X		X
Norwegian Icelandic	X			X
Hungarian		X	X	
Czech	X		X	

Norwegian *seg selv* and Icelandic *sjálfur sig* are complex but subject-oriented (Koster and Reuland, 1991, pp.12-13). Hungarian *maga* is not subject-oriented, but is not complex (Marác, 1989, referred to in Koster and Reuland, 1991, p.19). Finally, Czech *se* is subject-oriented and morphologically simple (Toman, 1991).

Irreducibility of long-distance reflexivity

A few important conclusions follow from the results of this section. First, it is apparent that the GB accounts of long-distance reflexives are inaccurate. On the one hand, they exclude three quarters of the possible correlations between morphological complexity and subject-orientedness, both for long and short-distance reflexives. On the other hand, given the explanatory machinery it uses, that approach can hardly be said to be on the right track or to admit subsequent improvement. The idea of reducing the binding constraint of long-distance reflexives to a recursive effect of some sort of short-distance constraints appears thus to be excessively theory driven to the detriment of empirical adequacy.

Second and more importantly, as it becomes clear from crosslinguistic examination that subject-orientedness is not correlated either to long or short-distance anaphoric binding, we are taught that there is no reason to bring the eventual solution for subject-orientedness into the formulation of binding

principles. This conclusion is confluent with the analysis of subject-orientedness put forward in the previous chapter, 2 Empirical Universality.

4.2 New Symmetries with Four Constraints

Building on Pollard and Sag (1992a) proposal, Xue, Pollard and Sag (1994) sketched an account of binding properties of long-distance reflexives whose major innovative feature is the departure from the desideratum of reducing long-distance anaphoric constraints to a recursive effect of short-distance constraints. They take into account data involving Chinese *ziji*, which was idiosyncratically classified as a "z-pronoun", and they observe that "z-pronouns must be o-bound" (principle Z).

What we argue for in this section is that this is a suitable basis to set up a generalized account of long-distance reflexives provided that, on the one hand, a more empirically adequate formulation is given to principle Z, and on the other hand, as suggested from the results above, separate treatments of binding, subject-orientedness and blocking effects requirements are assumed.

Long-distance exemption

As addressed in previous chapters, one of the most interesting features of Pollard and Sag's research on binding is the discovery that there are contexts where short-distance reflexives turn out to be exempt from the usual locality requirement on their dependent interpretation. This gave the form of a conditional statement to principle A, "a (short-distance) reflexive must be locally o-bound if it is locally o-commanded", which defines the exemption contexts as those where the short-distance reflexive is not locally o-commanded.

In this connection, the experiment which naturally presents itself is to check whether there are also exemption contexts for long-distance reflexives and whether they are the same as those for short-distance anaphors.

As for Portuguese, it is easy to verify that, closely parallel with short-distance reflexives, *ele próprio* is exempt from the binding requirement if it is not o-commanded.

(27) a. ? [O retrato dele próprio] foi pintado pela Maria.
the picture of he PRÓPRIO was painted by the Maria
His picture was painted by Maria.

b. * A Rita destruiu o retrato dele próprio.
the Rita destroyed the picture of he PRÓPRIO

Data in (27)a.-b. show that when the reflexive is o-commanded, it is required to be o-bound. In (27)a., *ele próprio* is not o-commanded, in which case it is able to support, quite marginally in the context of (27)a., a deictic use as it has no possible antecedent in the sentence. In (27)b., the reflexive is (non-locally) o-commanded, but has no (o-commanding) antecedent in the sentence, in which case the construction is ungrammatical. These should be put on a par also with (7)b. or (9)a., where (locally) o-commanded but not o-bound occurrences of *ele próprio* are also not grammatical.

This therefore strongly suggests that principle Z should be given the following definition:

(28) Principle Z

An o-commanded long-distance reflexive must be o-bound.

Principle Z

This formulation of principle Z, which applies to Portuguese, is very likely to have a general character. Let us reappraise the data available in the literature about a language of such an unrelated language family as Chinese.

Consider contrast (29), taken from Xue et al. 1994, exs.(11),(21).

(29) a. [Zhangsan_i de xin] biaoming Lisi_j hai-le ziji_{#i/j}.
Zhangsan DE letter indicate Lisi harm-ASP self
[Zhangsan_i's letter] indicates that Lisi_j harmed #him_i/himself_j.

- b. [Zhangsan_i de hua] anshi [Lisi_j de xin] zai yingshe ziji_{?i/j}.
 Zhangsan DE speech imply Lisi DE letter ASP allude-to self
 [Zhangsan's_i words] implied that [Lisi's_j letter] was alluding to
 him?_i/himself_j.

Xue et al. explain this contrast on the basis of an analogy with the unlike-person blocking effect assumed to hold in Chinese for discourse anaphora. The impossibility of *ziji* being bound by Zhangsan in (29)a., but not in (29)b., is said to be, on a par with "the unlike-person blocking, [...] a pragmatic or discourse processing effect of animate blocking".

We propose however a different tentative explanation in the light of the new principle Z. We take the contrast of (29) as possible evidence showing that also in Chinese the requirement of o-binding for long-distance reflexives only holds if the reflexive is o-commanded. The point here is to understand that, just as *ziji* is [+animate] and requires a [+animate] antecedent, its eventual o-commanders must also be [+animate] in order to qualify as o-commanders for the application of principle Z. Therefore, in (29)a. the coindexing Zhangsan/*ziji* is ruled out because, as *ziji* is o-commanded by Lisi, the long-distance anaphor is required to be o-bound and this constraint is not satisfied under that coindexing. In (29)b., in turn, the coindexing with Zhangsan or Lisi would be acceptable due to the fact that *ziji* is not o-commanded (by a [+animate] o-commander) and it is therefore exempt from binding requirements, allowing it to logophorically pick antecedents which do not o-command it.

An important consequence of this solution seems to be that we can dispense with Xue et al.'s assumption that, on a par with "syntactic *ziji*", ruled by principle Z, there is a "discourse *ziji*" whose apparent distinctive feature would be its ability to allow subcommanding antecedents.

It will be interesting to check the adequacy of our hypothesis against further empirical evidence designed by linguists speaking Chinese and other languages which have long-distance anaphors. Xue et al., 1994, ex.(26) seems, though, to provide yet further evidence supporting our analysis. It repeats an example taken from Wang (1990) where *ziji* does not need to have a (commanding or subcommanding) antecedent:

- (30) Mama de shu ye bei ziji de pengyou touzoule.
 mother DE book also BEI self DE friend steal-ASP
 Mother_i's book was also stolen by his_k friend.

In (30) *ziji* is not o-commanded (by a [+animate] o-commander) and, apparently, as with *ele próprio* in (27)a., it seems to be able to support a deictic use in the absence of overtly available antecedents in the sentence.

The binding square of opposition

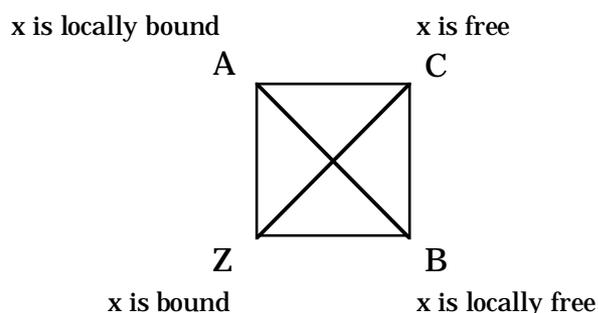
We come then to a point where it is possible and important to note that the merits of the new principle Z can be assessed not only in terms of its empirical adequacy, but also in terms of its impact on the whole set-up of binding theory. Principle Z naturally appears now not as a mere extra binding constraint, but as the fourth principle of the theory, on an equal footing with the three "classic" principles A, B and C, given the striking symmetry between the four principles.

- (31) A: A locally o-commanded sd-reflexive must be locally o-bound.
 Z: An o-commanded ld-reflexive must be o-bound.
 B: A personal pronoun must be locally o-free.
 C: A nonpronoun must be o-free.

Both reflexive (A; Z) and non-reflexive expressions (B, C) now have two binding principles ruling them, and the different senses in which the opposition local vs. non-local should be taken seems to enable a more fine-tuned clarification. Principle Z shows up as the non-local "extended" variant of principle A, in the same sense that principle C could have been taken as the non-local "extended" variant of principle B; but also principle B may be taken as the non-local "complementary" variant of principle A in the same sense that principle C may be taken as the non-local "complementary" variant of principle Z.

The elegance and heuristic value of the cross symmetries now suggested can be fully uncovered if the exact correlations between the four principles are made evident in a more formally precise way. Should the constraints expressed by the binding principles be stripped away from the reference to the class of items they apply to and cleaned up from their procedural mood, it turns out that the Binding Theory lends itself to be arranged under the form of a classical square of oppositions:

(32)



There are two pairs of contradictory constraints (one is true iff the other is false), which are formed by the corners related across diagonals, (A, B) and (C, Z); one pair of contrary constraints (they can both be false but they cannot both be true) made up from the corners related by the upper horizontal edge, (A, C); one pair of subcontrary constraints (they can both be true but cannot both be false), including the corners related by the lower horizontal edge, (Z, B); and two pairs of subaltern constraints (the subaltern is true if its superaltern is true, and the superaltern is false if its subaltern is false), formed by the corners of the vertical edges, (A, Z) and (C, B), where A and C are the superalterns — for a detailed presentation and the twenty five century history of the square of opposition vd. Parsons, 1999.

Consequently, by enlarging our sample of data, we obtain a more general and empirically adequate account of long-distance anaphora. But notably, we would like to stress that it leads to an unexpectedly more integrated binding theory. It is not unlikely that the formal oppositions between the four principles now made evident may just be the starting point for an inquiry into unsuspected properties of binding phenomena, as we will discuss below.

4.3 The Quantificational Structure of Binding Constraints

Given the results from previous sections, questions such as – Does (32) signal a subjacent quantificational structure to the Binding Theory? Does the universal nature of Binding Theory stem from its possible quantificational structure? – appear to highlight the need to pursue new promising paths of research on the nature of anaphoric nominal expressions. We will endeavor to do so in the remainder of the present chapter and in the next chapters.

Natural language quantification

As a preliminary step, criteria to uncover quantificational natural language expressions are briefly reviewed below.

Duality

Barwise and Cooper's (1981) seminal work gave rise to a fruitful research tradition where generalized quantifier theory has been applied to the analysis of natural language quantification. These authors suggested that a universal characterization of natural language nominal quantification could be formally given by means of formal properties defined in that theory. The property "to live on" – Q lives on A iff_{def} for any X , $Q(X)$ iff $Q(X \cap A)$ (Barwise and Cooper, 1981, p.178 – was postulated as being the most prominent one, admittedly being the common specific nature of all nominal quantifiers.

Later, Loebner (1987) suggested a criterion to ascertain the quantificational nature not only of nominal expressions, but also of natural language expressions in general. That is the property that, for a one place second order operator Q expressed by a given expression, there is a corresponding dual operator $\sim Q \sim$ (the outer negation of the inner negation of Q) which can also be expressed.

This duality-based perspective on the essence of natural language quantification made it possible to extend quantification beyond the well known cases of nominal quantification supported by the "classic" determiners all, some, most, many, etc., namely by covering the realms of temporality and possibility as well. Moreover, items such as still/already, and others (enough/too, scaling adjectives, many/few, etc.) although not lending themselves to straightforward analysis in terms of set quantification, can also be arranged in a square of duality. The formalization of the semantics of these aspectual items conforming to the duality criterion by Loebner led to the enlarging of the notion of quantification through the introduction of the new concept of phase quantification.

Phase quantification

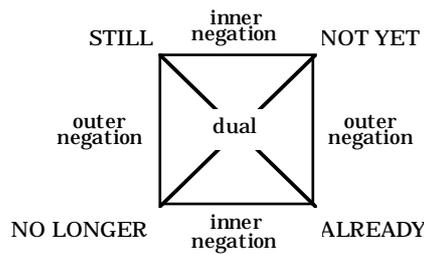
Loebner showed that still and already express second order dual operators and that they are corners of a square of duality.

Let P be "she is asleep" and $\sim P$ "she is awake", durative propositions which are the arguments of the semantic operators corresponding to already and still. We then have:

- (33) She is already asleep iff it is not the case that she is still awake.
 $\text{ALREADY } P \text{ iff } \sim \text{STILL } \sim P$

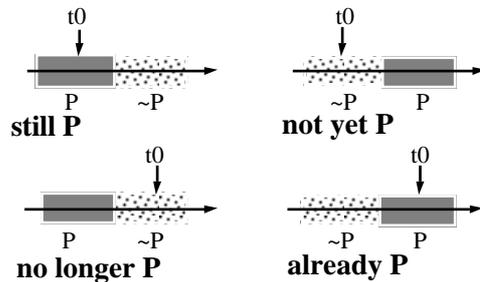
Further similar tests can be made in order to show that these aspectual items enter the following square of duality:

- (34)



In order to get a formalization of (34), Loebner noted that already should be taken as conveying the information that there is a phase not-P which has started before a given reference time t_0 and might be followed by at most one phase P which lasts until t_0 . This can be displayed in a time axis by means of the diagram in (35).

- (35)



Similar diagrams for the meaning of the other aspectual phase quantifiers of this square of duality are easily interpretable. Inner negation results then in exchanging the positive and the negative semiphases, while outer negation concerns the decision as to whether the parameter falls into the first or the second semiphase.

Phase quantifiers in general (already, scaling adjectives, etc.) were thus characterized as requiring two ingredients: (i) a property P, which defines a positive phase in a sequence of two opposite phases; (ii) a parameter point. The four types of quantifiers only differ in presupposing that either the positive or the negative semiphase comes first and in stating that the parameter point falls into the first or into the second semiphase.

Next Loebner showed that the semantics of such phase quantifiers sketched in the diagrams above can be formalized in such a way that a square of duality formed by the generalized quantifiers \exists some'(D,X)/ \exists very'(D,X) turns out to be subjacent to the square of duality of ALREADY/STILL. In order to do this, he just needed the auxiliary notion of the starting point of the relevant semiphase. This is rendered as the infimum of the set of the closest predecessors of the parameter point pt with property P, or \sim P, which form an uninterrupted linear sequence (termed GSI(R,pt) by Loebner, with R standing for P or \sim P and pt for parameter point):

(36)
$$\text{GSI}(R,pt) =_{df} \inf\{x \mid x < pt \wedge R(x) \wedge \forall y (x < y \leq pt \rightarrow \neg R(y)) \rightarrow \forall z (x < z < y \rightarrow R(z))\}$$

The semantics of the four phase quantifiers sketched above can then be rendered in the following way – in what Loebner does is to enter \exists for phase quantification –, making $pt=t0$ for the parameter point. For $R=\sim$ P:

(37) STILL: $\exists x (\text{GSI}(\sim P, pt) < x \leq pt \wedge \sim R(x))$
 If every x such that $\text{GSI}(\sim P, pt) < x \leq pt$, $\sim R(x)$

ALREADY: $\exists x (\text{GSI}(P, pt) < x \leq pt \wedge R(x))$
 If some x such that $\text{GSI}(P, pt) < x \leq pt$, $R(x)$

NOT YET: $\exists x (\text{GSI}(P, pt) < x \leq pt \wedge \sim R(x))$
 If no x such that $\text{GSI}(P, pt) < x \leq pt$, $\sim R(x)$

NO LONGER: $\exists x (\text{GSI}(\sim P, pt) < x \leq pt \wedge R(x))$
 If not every x such that $\text{GSI}(\sim P, pt) < x \leq pt$, $R(x)$

Phase quantification ingredients in binding

Taking the duality criterion for natural language quantification, our goal is thus to make apparent the possible quantificational structure subjacent to binding constraints. Given that squares of opposition and squares of duality are logically independent (vd. Loebner, 1987, p.55), we will show that binding principles are also associated with a square of duality, on a par with the square of opposition of (32).

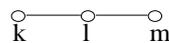
We will therefore argue that binding principles are but the reflex of the phase quantificational nature of corresponding nominal expressions: Short-distance reflexives, pronouns, long-distance reflexives and non-pronouns will be shown to express phase quantifiers acting on the grammatical obliqueness axis.

In order to show this the relevant components involved in phase quantification should be identified.

Scale

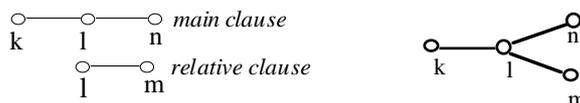
The relevant scale here is not the continuous linear order of moments of time, as for STILL/ALREADY, but a discrete partial order made of reference markers (cf. Discourse Representation Theory – for more details on this notion see next chapter 5 Formal Semantics) arranged according to the relative obliqueness of corresponding grammatical functions. Note that in multiclausal constructions there is the subordination of different clausal obliqueness hierarchies (for the sake of comparability with diagrams (35) involving time arrow, Hasse diagrams for obliqueness are displayed with a rotation of 90° right):

(38) Kim said Lee saw Max.



Note also that the relation "less oblique than" may not be linear:

(39) Kim said Lee, who saw Max, hit Norma.



Opposite semiphases

The sequence of two opposite semiphases is defined by a property P . Contrary to what happens with *ALREADY*, where operator (aspectual quantifier) and operand (durative proposition) are rendered by different expressions, in binding phase quantification the operand P is also contributed by the nominal expressing the operator, i.e. expressing the binding phase quantifier.

For a given nominal N , P is determined by the relative position of N in the obliqueness scale. We assume that for a reference marker r corresponding to N , semiphase P is a linear stretch containing only elements that are less than or equal to r in the obliqueness order, i.e. reference markers corresponding to nominals o -commanding N . Moreover, if semiphase P is presupposed to precede semiphase $\sim P$, P is such that the last successor in it is local wrt to r ; and if semiphase $\sim P$ is presupposed to precede semiphase P , P is such that the first predecessor in it is local wrt to r , however locality for binding may be parameterized for each language. In both cases the closest P neighbor of semiphase $\sim P$ has to be local wrt r :

$$(40) \quad P_r(x) \text{ iff} \\ x \leq r \wedge \forall y [(\sim P_r(y) \wedge (x < y \vee y < x)) \rightarrow (x \text{ is local wrt } r \wedge y \text{ is not local wrt } r)]$$

Parameter point

As to the parameter point, in binding phase quantification, it is the reference marker a which is taken as the antecedent of r .

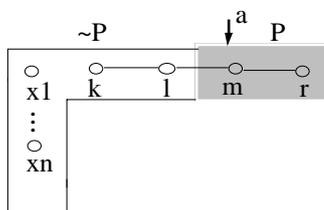
Binding phase quantifiers

We can now formalize phase quantification subjacent to anaphoric nominals. Let us start with a reflexive expression N such as *himself*.

Reflexives

Consider the following constructions, definition and schema:

- (41) a. Kim said Lee thinks j hit himself $_j$.
 b. # Kim said Lee $_j$ thinks j hit himself $_j$.
 c. Q_A : $\exists x$ some'(x , GSI($\sim P, a$), $k \leq x \leq a$), P
 d.



Here N can be interpreted as presupposing that a semiphase $\sim P$ precedes a semiphase P, requiring the parameter point to occur in the latter, i.e. the antecedent a of r is to be found in semiphase P among the reference markers corresponding to the local o-commanders of r.

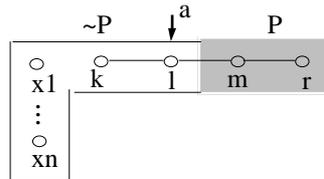
This is captured by the definition above of the phase quantifier Q_A . Satisfaction of $Q_A(P)$ obtains iff between the bottom of the uninterrupted linear sequence $\sim P$ most close to the (parameter point) antecedent a and a inclusive there is at least one reference marker in P. Given $\sim P.P$, this amounts to requiring that a be in P, and consequently that a be a local o-commander of r.

Pronouns

Following similar analysis, it is easy to see how the phase quantificational force of a pronominal expression N should be formalized:

Symmetries and Duality

- (42) a. # Kim said Lee thinks Max_i hit Kim_i.
 b. Kim said Lee_i thinks Max hit Kim_i.
 c. $Q_B: \exists x(GSI(\sim P, a) \leq a), P$
 d.



Here the parameter point *a* occurs in semiphase $\sim P$, which amounts to the antecedent being picked outside the set of local *o*-commanders of *r*. $Q_B(P)$ is satisfied iff no reference marker between the bottom of the uninterrupted linear sequence $\sim P$ closest to the (parameter point) antecedent *a* and *a* inclusive is in *P*. Given $\sim P.P$, this amounts to requiring *a* to be in semiphase $\sim P$, and *a* to be a non-local *o*-commander of *r*.

As in diagram (41)d., $\sim P$ is taken here as the complement set of *P*. All reference markers which are not local *o*-commanders of *r* are in it, either *o*-commanding *r* or not. Notice that set $\sim P$ also includes reference markers *x1...xn* introduced by previous sentences or the extra-linguistic context, which in constructions similar to (42)b. accounts for possible deictic readings of the pronoun.

Below, when studying non-pronouns, we will see why the possible non-linearity of the obliqueness order will lead us to consider that $\sim P$ is slightly more complex than just the complement set of *P*.

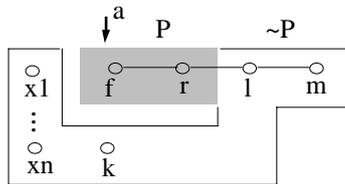
Long-distance reflexives

Coming now to long-distance reflexives, ruled by the fourth binding principle, the following formalization can be made:

- (43) a. [O amigo de Kim]_i disse que ele próprio_i acha que Lee viu Max.
 [Kim's friend]_i said LD-REFL_i thinks Lee saw Max.
 b. # [O amigo de Kim]_i disse que ele próprio_i acha que Lee viu Max.
 [Kim's_i friend] said LD-REFL_i thinks Lee saw Max.

c. $Q_Z: \lambda x \lambda a \text{GSI}(P, a) \langle x \leq a \rangle, P$

d.



Here, as with short-distance reflexives in (41), a is required to occur in P though the presupposition now is that semiphase P is followed by semiphase $\sim P$. Taking into account the definition of P in (40), the antecedent of N is thus required to be an o -commander (local or not) of N . The semantics of phase quantifier Q_Z is such that, for $Q_Z(P)$ to be satisfied, between the bottom of the uninterrupted linear sequence P closest to the (parameter point) antecedent a and a inclusive, every reference marker is in P . This amounts to requiring a to be in semiphase P , and a to be an o -commander of r .

Non-pronouns

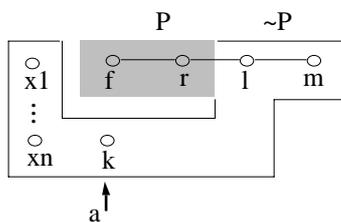
Finally non-pronouns need to be formalized as the last phase quantifier of (37):

(44) a. [Kim's friend]_i said Kim_j thinks Lee saw Max

b. # [Kim's friend]_i said Kim_j thinks Lee saw Max.

c. $Q_C: \lambda x \lambda a \text{not_every}'(x \text{GSI}(P, a) \langle x \leq a \rangle, P)$

d.



The (parameter point) antecedent a is required to occur in $\sim P$, which means that a cannot be an o -commander (local or not) of r . This renders the same condition as expressed by principle C, that non-pronouns are free, although it also encodes a not so common assumption about the referential autonomy of non-pronouns. Here, as

with other more obvious anaphoric nominals, the interpretation of non-pronouns is taken as being dependent on the interpretation of other expressions or on the salience of reference markers made available by the communicative context. Taking an extreme example in order to support the plausibility of this view and somehow awkwardly abbreviate a deep philosophical discussion, as referred to in chapter 1 Anaphora and Reference Processing, one should notice that even a proper name is not a unique label of a given individual, given that knowing who is the person called John (out of those we know that are named John) depends on the context.

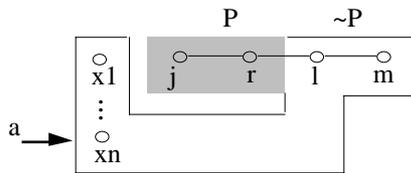
Note that as in previous diagrams, $\sim P$ is taken in (44) merely as the complement set of P . However, Q_C requires this to be seriously explored and a more accurate definition of $\sim P$ to be given for phase quantification in non-linear orders, where it is possible for not all elements to be comparable.

For $Q_C(P)$ to be satisfied, between the bottom of P and the (parameter point) antecedent a inclusive, not every reference marker is in P . Since we have here the presupposition $P.\sim P$, and given P is an uninterrupted linear sequence, this would amount to requiring a to be in $\sim P$.

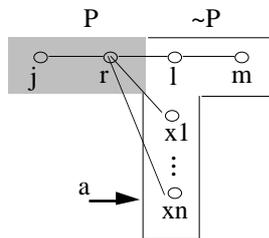
It is worth noting then that if we keep $\sim P$ simply as the complement set of P , the interpretation of non-pronouns is not adequately predicted by $Q_C(P)$. Consider the following sentence:

(45) a. John_i said Kim_j thinks Lee_k saw Max_l.

b.



c.



Let D be $\{x: GSI(P,a) < x \leq a\}$, which has the role of domain restrictor in Q_C . Taking schema (45)b., it is easy to check that in constructions and interpretations such as

(45)a., D is always empty. In fact, it is not the case that $GSI(P,a) \leq a$ as $a=xi$ is not comparable to any element of P , and a fortiori it is not comparable to the bottom of P . Consequently, $every'(D,P)$ is trivially true whatever reference marker xi we take as antecedent for r , and Q_C , i.e. $not_every'(D,P)$, is trivially false. The interpretation of (45)a. sketched in (45)b. would thus be incorrectly ruled out.

What these considerations seem then to suggest is that, when phase quantification operates on non-linear orders, negation of the operand P is slightly more complex than simple Boolean negation rendering the complement set. We are thus taught that negation of P also involves the lifting of the complement set of P , \bar{P}_\perp , with \perp equal for r , the top of P , when $P.\sim P$ (though it is empirically not necessary, for the sake of formal uniformity, when $\sim P.P$, the order-theoretic dual of this specification of $\sim P$ can be assumed). It is easy to check with diagram (45)c. that this specification of $\sim P$ makes it possible to satisfy $Q_C(P)$ in exactly the correct constructions.

Split antecedency

The above statement that this definition of $\sim P$ makes it possible to satisfy $Q_C(P)$ in exactly the correct constructions is to be understood as addressing the constructions involving singular antecedents.

As discussed in the previous chapter, reflexives and non-reflexives (including pronouns and non-pronouns) differ, among other aspects, with respect to split antecedents. While a plural reflexive cannot take several antecedents with different degrees of obliqueness, a non-reflexive can take antecedents split over different degrees of obliqueness, eventually bearing different grammatical functions.

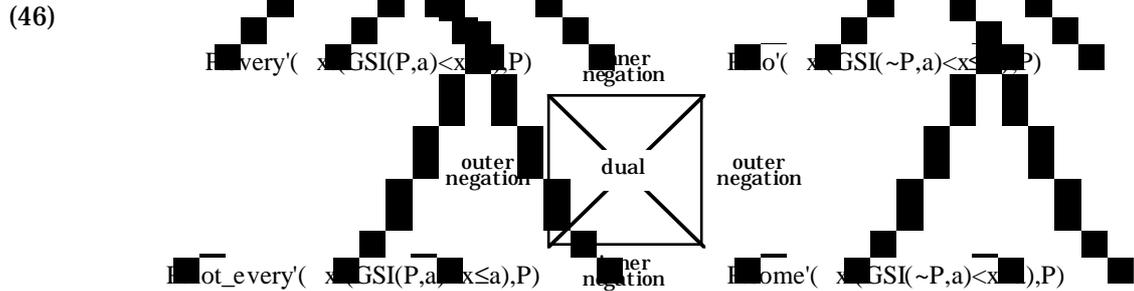
The behavior of reflexives with respect to unsplit antecedency is correctly accommodated in the definition of phase quantifiers associated with reflexives, as only one possible antecedent is admitted in the definition of Q_A as well as Q_Z . The same is not true, however, for non-reflexives, as only one possible antecedent is admitted in the definition of Q_B and Q_Z .

This specific difference between reflexives and non-reflexives can be formally accounted for simply by means of a more accurate definition of $\sim P$, which also covers the case of plural non-reflexives. In particular, besides being defined as the lifting of the complement set of P , \bar{P}_\perp , $\sim P$ can also be required to include, among the antecedents $x_1\dots x_n$, the order-theoretic "individual sums" a la Link of the elements of the complement of P . Accordingly, and taking the plural operator '*',

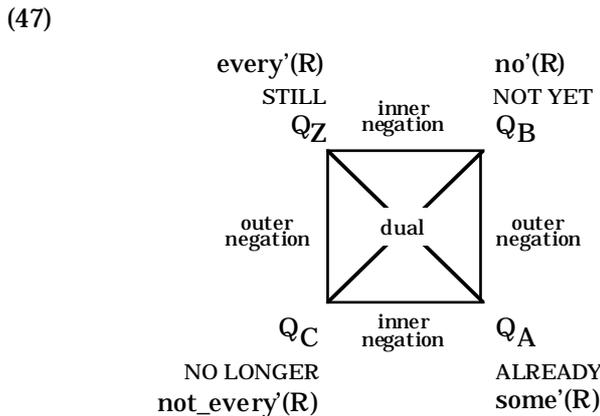
which when applied to P generates all the individual sums from the extension of P (cf. Link, 1983, p.306), $\sim P$ can then be defined – now covering both singular and plural cases – as the lifting of the “pluralization” of the complement set, $(\ast \bar{P})_{\perp}$.

The binding square of duality

Following Loebner’s claim that second order duality is the cardinal property to recognize the quantificational character of natural language expressions, we are thus led to the view that the interpretation of anaphoric nominals is ruled by their phase quantificational force over the obliqueness order. Since the defining formulae of binding quantifiers result from (37) merely by assigning P the definition in (40) and taking the parameter point pt to be the eventual antecedent a, it is with no surprise that the following square of duality is obtained for binding phase quantifiers:



Which compares to other squares of duality as follows:



4.4 New Insights into Binding

This radically new conception of binding constraints is very likely to bear important consequences not only in terms of the understanding of anaphoric mechanisms captured by binding theory, but also in terms of our conception of generalized quantification in natural language and of the semantics of nominal expressions. While we aim to address most of these implications in the following chapters, in this section we mostly analyze how a few central issues more closely associated with binding are handled under this new viewpoint.

Exemption

One of such issues is the so called exemption positions. Under this new quantificational perspective, an exempt position can be characterized as a borderline case where the maximum shrink of semiphase P occurs, i.e. when P is the singleton whose sole element is r , the reference marker whose interpretation is to be anchored by finding an antecedent for it.

Given the definition of binding phase quantifiers, the maximum shrink of P into a singleton significantly affects only the quantifiers where the (parameter point) antecedent a is found in P , namely Q_A and Q_Z . In these cases, for a to be in P and the quantification to be satisfied, a can only be r , r being thus its own antecedent. Consequently, although the quantification is satisfied, a "meaningful" anchoring of the reference marker r has yet to be accomplished, as r is anchored to itself by the sole result of quantification being satisfied. Admittedly, an overarching interpretability requirement is imposed for the significant anchoring of nominals to be consummated, which will induce the so-called logophoric effect in the present case: For the reflexive (short or long-distance) to be interpreted, and given that

satisfaction of its binding constraint is ensured, it should thus find an antecedent outside any specific restriction from binding theory.

Logophoricity appears then to be a derivative property of reflexives which emerge only when these items are forced, in order to satisfy binding phase quantification, to enter into a reference anchoring loop. The breaking of the loop enforced by an interpretability requirement induces the exceptional logophoric effect.

This therefore explains the exemption restrictions in the definitions of principles A and Z and so-called logophoric effects associated with exempt anaphors. Restrictions which hitherto appeared to be mere stipulations can now be justified in a more principled way thanks to this new approach.

Gaps

It is also well known that although the four binding principles are hypothesized to be universal, there are languages which do not have all the corresponding four types of nominals. One of such cases typically pointed out is the English language, which is not known to have long-distance reflexives.

A justification for the circumstance that some languages exhibit this kind of “gaps” in terms of binding constraints may follow now in simple terms.

In other squares of duality, it is possible that not every corner of the square is lexicalized, an issue that Loebner (1987) discusses at length. With respect to English, for instance, he noted that the square of duality concerning deontic possibility involving right happens to have only two lexicalized corners, right and duty, while in the duality square of classical logic only the three corners, some, every and no are lexicalized.

Likewise, with respect to binding quantifiers, it is then natural to admit that in some languages it may happen that some corners of the corresponding duality square may not be expressed by any type of anaphoric nominal.

This is thus a hypothesis for explaining binding “gaps”, which appears on a par with the hypothesis put forward in Section 3.1. We will refer back to this topic in the last chapter 8 Conclusions: Summary and Outlook for a more detailed examination.

Implied epistemic twists

Besides the advances in accounting for these specific issues associated with binding, there seems to be some other significant implications arising from the new quantificational perspective.

Quantification in natural language

Russell is credited as having been the first to stress the fact that there is no correspondence between surface form and logical form of quantificational expressions of natural languages. Loebner has emphasized this non-correspondence as he pointed out that, while domain restrictor and quantified predicate are rendered by two different expressions in nominal quantification, only the quantified predicate may be superficially available in phase quantification.

Now, with binding phase quantification, this gulf between surface and logical form of natural language quantification has widened further: There is no surface expression either for the domain restrictor of quantification or for the quantified predicate.

But other even more important implications for our understanding of the semantics of natural languages may have been uncovered as well. Present results seem to advocate new and important extensions with respect to the realm of quantification in natural language. On the one hand, phase quantification is extended to partially-ordered universes which are possibly non-linearly ordered. On the other hand, and most important, to universes whose elements are not individuals or events of the "extra-grammatical" world, but entities of the "intra-grammatical" world itself. Actually, the models against which binding phase quantification is interpreted are not "representations of the world" but "grammatical representations", where the basic entities are reference markers.

Given this, satisfaction of a formula made out of a binding phase quantifier turns out thus to be a well-formedness constraint on the syntax and semantics of the sentence where the quantifier occurs. For "classic" quantification to be interpretable, in turn, one has to know what the world/model has to be for it to be true; for binding phase quantification to be interpretable one has to know what the grammatical representation/model has to be for it to be true, i.e. one has to know what the well-formedness restrictions are on the grammatical representation induced by the requirement for quantification to be true.

Nominal semantics

Focusing further on the issue of natural language quantification, it is also worth noticing that the present results uncover an unsuspected quantificational nature in nominals which are typically classified as non-quantificational in the literature. It is interesting to note that, in opposite direction, Evans (1980) had already signaled the referential nature of quantificational NPs when involved in anaphoric chains, as in the following example:

- (48) Most of the senators_i voted against. They_i were not satisfied with the formulation given to the proposal.

The following therefore seems to be taking shape: All NPs seem to have both referential and quantificational force. While those which have “genuine” quantificational force (every man, most students,...) have a somewhat “deferred” referential force, conversely, those which have “genuine” referential force (John, the book, that cat,...) have a somewhat “deferred” quantificational force. Therefore, except possibly for indefinites – which appear to be an isolated class –, the dividing line may not be between quantificational and non-quantificational NPs (or referential and non-referential), but for want of better terms, between extra-grammatical and intra-grammatical quantificational NPs, respectively, extra-grammatical and intra-grammatical referential NPs.

This issue will be elaborated on in further detail in the next chapter 5 Formal Semantics.

Universality

It is also worth discussing how this new quantificational view on binding may contribute to understanding the universality of binding constraints across languages.

In this respect a first crucial observation has to do with the highly non-trivial nature of the quantificational account to binding. Its non-triviality results from two concomitant facets. On the one hand, the quantificational devices at work replicate analogous devices which underlie other linguistic phenomena and which were independently designed. This enormously reduces the chances that we have given an ad hoc solution for this research topic, given its basic pattern is not an arbitrary construct invented from scratch – such as the informal format of

syntactic binding principles – but one that is known to consistently emerge in other areas of grammar.

On the other hand, the quantificational devices at use here are not generic analytical tools that can be easily used and extended at will to the analysis of most linguistic phenomena – e.g. syntactic trees plus a movement rule, or feature structures plus some unification mechanism. This also reduces the chances that we have given an inconspicuous solution to our problem not related to eventual requirements bearing on its specificity.

A second observation concerns the high degree of abstraction and generalization achieved with the quantificational approach. On the one hand, obliqueness quantifiers, though building in a number of auxiliary constructs and notions – e.g. reference markers, locality, obliqueness, etc. –, impose constraints on the anaphoric capacity of nominals in a notoriously simple and elegant way, v.z. as a well known kind of relations on sets (taking a generalized viewpoint on quantifiers). On the other hand, the quantificational approach provides for a unique and notoriously elegant way of relating the different binding constraints with each other in a extremely compact “theory”.

All these observations make it convincingly manifest that there is a reduced chance of the quantificational shape of binding not being rooted in essential aspects of binding, despite appearing quite surprisingly innovative. This self-suggesting imagery of disparate pieces neatly falling into places to make a natural whole seems thus to militate against the possible remark that the quantificational approach to binding constraints merely provides us with a notational variant of the usual binding principles. The history of scientific inquiry may help us to agree that, even though two approaches or theories form identical empirical predictions, there may be reasons to opt for one of them however eccentric it may initially appear. Kepler’s geocentric approach of the solar system, for instance, was able to deliver roughly the same empirical predictions concerning the movements of the planets as those permitted by Copernic’s heliocentric approach. It was less elegant, however, than the latter because it involved *inter alia* the implied assumption that the planets could stop and invert their motion, an assumption arising from the way the movements of planets could be “immediately”, or “naturally”, observed from earth.

Given thus the universality of quantification and quantificational devices in natural languages, if binding constraints are the visible effect of binding quantifiers then it is not at all surprising that they are universally operative across natural languages.

We will return to this issue of the universality of binding constraints again in the last chapter 8 Conclusions: Summary and Discussion.

Integration in grammar and in reference processing

Finally, a last issue worth examining in this chapter concerns the difference between binding constraints in their mainstream definition and in their new quantificational formulation as regards the object they are constraining.

Notice that binding constraints as phase quantifiers are constraints on the reference markers that may be taken by an anaphoric expression as its antecedents. This means that binding constraints are to be understood as circumscribing the intrinsic anaphoric potential of each anaphoric expression with which they are associated, with no exogenous concerns, such as the anaphoric potential of other co-occurring expressions, or how these other expressions happen to be resolved. In this sense, binding constraints are independent from the subsequent procedure of reference processing, in general, and anaphor resolution, in particular, that may eventually exclude some of the markers accepted as being in the anaphoric potential of the relevant anaphor due to the effect of other filters (and will certainly pick one of the markers in that potential as the preferred antecedent).

Binding constraints as they have basically been taken in their mainstream formulation since the early eighties, in turn, are constraints on coindexation between NPs. They are thus a set of sieves on tentative solutions for the anaphoric relations between cooccurring nominal expressions. This supposes that binding constraints are somehow active a posteriori, or at least as a last step, as regards the procedure of anaphor resolution.

This difference has a significant impact on how binding constraints can be integrated into grammar, on the one hand, and into the global procedure of reference processing and anaphor resolution, on the other. As we will discuss at length in the following chapters, setting binding constraints as filters at the outcome of the anaphor resolution process raises huge difficulties either in terms of their neat integration into grammar and reference processing procedure, or in terms of the complexity of the algorithm required to validate them.

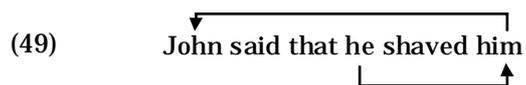
These are difficulties that the quantificational conception of binding constraints can help to solve. It is worth noting, however, that this improvement is achieved not only due to this new conception of binding constraints, but also due to an accurate understanding of the functions that the mainstream binding constraints play in the overall procedure of reference resolution. In particular, following the unique insights of Giorgi, Pianesi and Satta (1990, p.124), which have been duly appreciated in subsequent literature only by Stuckardt (1996b), it is relevant to recognize that mainstream binding constraints are actually not confined to circumscribe the intrinsic anaphoric potential of anaphoric expressions. As we are going to discuss

below, given their setup, they concomitantly ensure some extra-grammatical functions that belong to the process of anaphor resolution.

Reciprocal validation

In this connection, a first point worth bringing to light is that, contrary to quantificational binding constraints, the validation of mainstream binding constraints involves, inasmuch as their objects are coindexations, a covert reciprocal validation of anaphoric links. Let us try to make this issue clearer by means of a concrete example.

Consider the following construction.



For the sake of the argument, suppose that he takes him as its antecedent, and him in turn takes John as its antecedent, and finally John is interpreted against a contextually salient reference marker. Without further elaboration, this resolution of the anaphoric expressions involved would be deemed to comply with binding constraints. In fact, both he and him take antecedents that do not locally o-command them, and the antecedent of John does not o-command it. However, under this resolution of reference the construction is not interpretable apparently because he and him would end up referring to the same entity.

When looking for what might have failed in the application of binding constraints, we are led to realize that nothing failed. The point is that each anaphor solution was considered in isolation. As relevant binding constraints were applied taking into account a specific solution for each anaphor in isolation, the reciprocal application of binding constraints over anaphoric links which is implicit in the use of coindexation was not observed.

Repeating the example above with the coindexation corresponding to the resolution considered, as in (50) below, it comes to light that, under the mainstream binding theory, the use of coindexation implicitly imposes for any anaphoric link a reciprocal validation of that link. This validation is reciprocal in the sense that, not only the anaphoric link between anaphor and its antecedent is verified as complying with the binding constraint adequate for the anaphor at stake, but also the reverse link between antecedent and anaphor is submitted to the binding constraint adequate for the antecedent.

(50) # John_i said that he_i shaved him_i.

In fact, while taking him as antecedent, the pronoun he establishes an anaphoric link that, under that “perspective”, complies with principle B, but which is however an anaphoric link which is not reciprocally validated, in as much as him cannot be locally o-commanded.

The issue of covert reciprocal validation is not confined to examples as the one above, with pronouns. Constructions such as those below are just a few examples of other instances where this covert reciprocal validation of anaphoric links by mainstream binding constraints can be detected (link i should be taken here just from left to right; link j just from right to left):

- (51) a. # The girl who saw John's friend_j said that the boy_i shaved the boy_{i/j}.
b. # The girl who saw John's friend_j said that he_i shaved the boy_{i/j}.
c. # The girl who saw John's friend_j said that the boy_i shaved him_{i/j}.

Under the new quantificational conception of binding constraints, the role of these constraints is tuned to circumscribe the intrinsic anaphoric potential of anaphors by taking into account lexically defined anaphoric properties (and the contextual surrounding grammatical geometry, over which they may be active). How this potential may be realized and the anaphor eventually resolved (especially taking into account, among all the many relevant factors, the way other cooccurring anaphors happen to be resolved) is an issue to be handled in the process of anaphor resolution, and not in the process of binding constraints verification, under a formal setup which will be discussed in detail in the following chapters.

Mediated antecedency

A second related issue is referred to, for want of a better term, as selective mediated antecedency.

As can be seen from the contrasts below in (52), when it comes to selecting a specific antecedent for a given anaphor out of the set of its grammatically admissible antecedent candidates, that resolution is constrained by the way other cooccurring anaphors happen to be resolved. Notice that the antecedent John that him takes in a., cannot be taken in a'.. Contrary to a., in a' the subject of the clause where him occurs also takes John as antecedent.

A similar contrast can be observed in b./b'..

- (52) a. Mary convinced John_i that Jane shaved him_i.
 a'.# Mary convinced John_{j/i} that he_j shaved him_i.
 b. The girl who knows John_i said that Jane shaved the boy_i.
 b'.# The girl who knows John_{j/i} said that he_j shaved the boy_i

As with reciprocal validation, this is an attribution of the process of anaphor resolution that is, however, accommodated inside the mechanics of mainstream binding constraints. Notice that in circumstances as those depicted above the two indexes *i* and *j* (taken as distinct here just for the purpose of signaling two different anaphoric links) are in fact to be identical, as one and the same index, which means that he and him are coindexed, and him may be locally o-bound, thus infringing principle B.

What is thus interesting to note is that this sensitivity to mediated, transitive antecedency, holding for pronouns and non-pronouns, seems to be selective in the sense that it is not active for reflexives. That is what the contrast between the data in (52) and in (53) seems to suggest.

- (53) a. The girl who saw John_i shaving himself_i loves Peter.
 a'. The girl who saw John_{j/i} shaving himself_i loves him_j.
 b. A rapariga que sabe que o Pedro_i gosta dele próprio_i foi com o João ao cinema.
 the girl who knows that the Pedro likes of _ele PRÓPRIO went with the João to _the movies
 The girl who knows that Pedro likes himself went with João to the movies.
 b'. A rapariga que sabe que o Pedro_{j/i} gosta dele próprio_i foi com ele_j ao cinema.
 The girl who knows that Pedro likes himself went with him to the movies.

Though in a'. and b'. the reflexives happen to be (transitively) coindexed with an NP that does not locally o-command them, in violation of principle A, the corresponding constructions do not turn out to be ungrammatical.

This different sensitivity of different classes of anaphors to transitive antecedency has its subtle reflex in the mainstream setup of binding constraints. For instance, while short-distance reflexives are required to be locally coindexed, pronouns are not required to be non-locally coindexed with some antecedent. Rather, pronouns are required not to be locally o-bound. This handy solution allows in fact to accommodate the specific sensitivity of pronouns to transitive antecedency in binding theory. However, contrarily to empirical evidence and intuition about

its anaphoricity, it awkwardly eliminates from mainstream binding constraint on pronouns any requirement that these expressions be anaphorically resolved, that is, that they have an effective anaphoric nature and take antecedents.

Again, under the new quantificational conception of binding constraints, the role of these constraints is tuned merely to circumscribe the anaphoric potential of anaphoric expressions. This approach for binding constraints clearly separates what should count as defining the intrinsic anaphoric potential of anaphors from other factors involved in the concrete actualization of that potential. Under this perspective, as with reciprocal validation, the selective mediated antecedency is to be handled in the course of the extra-grammatical process of anaphor resolution.

4.5 Summary

In this chapter we began by establishing the result that Portuguese has a long-distance reflexive which, contrarily to widely assumed correlations about this class of nominals, is morphologically complex and non-subject-oriented. This indicated a research path which, taking into account data available in the literature on several other languages with long-distance reflexives, enabling us to show that, even without this result from the Portuguese language, these putative correlations and the associated assumption that anaphoric links between long-distance reflexives and their antecedents are reducible to a concatenation of short-distance links are empirically ungrounded.

These first results on long-distance reflexives would show to be highly serendipitous. They lend empirical support to the claim that there is a fourth binding principle belonging to the binding theory, with very significant consequences both for binding and for the overall semantics of nominals.

We showed that the now four binding principles can be arranged into a square of logical oppositions. This stimulated research on the eventual quantificational character of binding constraints. Adopting Loebner's duality criterion for quantification in natural language, and the formal tools he developed for the analysis of phase quantification, we showed that binding constraints can be seen as

the result of phase quantifiers associated with each class of anaphoric nominals acting on the obliqueness order axis.

This lent support to a radically new understanding of the formal nature of binding constraints, and to interesting explanations of some associated issues such as the exemption occurrences and logophoric behavior of reflexives.

Finally we hinted at possible implications of this new understanding of binding constraints for the overall conception of the semantics of nominal expressions. It appears now that quantificational force is also associated with nominals typically identified as strictly referential, as opposed to others deemed to be quantificational.

5 Formal Semantics

5.1	Obliqueness Quantification in Grammar	146
	Component vs. composition of meaning	146
	A framework for semantics	146
5.2	Obliqueness Quantification as a Component of Meaning	146
	Representing intra-grammatical quantification	146
	Underspecification of anaphoric links	146
	Toward a uniform dichotomic semantics for nominals	146
5.3	Binding Constraints as Determinants of Meaning Composition	146
	Content of representations and content of rules	146
	Principled constrained construction rules	146
	Between two approaches	146
5.4	From Anaphor Resolution to Reference Processing	146
	Coreference	146
	Bridging, indirect or associative	146
	E-type anaphora	146
	Split antecedents	146
	“Donkey” anaphora	146
	E-type and bridging anaphora	146
	Bound anaphora	146
5.5	Summary	146

Far from drawing to a close the inquiry into grammatical conditions on anaphoric interpretation, the new quantificational perspective on the nature of binding constraints sheds light on several new questions. It appears as the first step of a number of new lines of inquiry. On which level of grammar should one accommodate obliqueness quantification? How is obliqueness quantification possibly coupled with the classical occurrences of quantification? What is its relevance for the processing of anaphora? These are just some of the many questions that offer themselves now for scrutiny.

While we aim to address most of these questions arising from the new perspective on binding in the remainder of this dissertation, we begin by querying how to accommodate obliqueness quantification in grammar. In this chapter we also discuss how the grammatical representation of binding constraints may be coordinated with the processing of reference, in general, and the resolution of anaphors, in particular.

5.1 Obliqueness Quantification in Grammar

It may not be straightforward to ascertain whether binding constraints, in quantificational format, belong to the realm of Syntax or to the realm of Semantics. It is true that important syntactic notions, e.g. grammatical function or obliqueness

hierarchy, are involved in the make up of binding constraints. However, the singular mechanisms involved in the formulation of obliqueness quantification have no parallel with any other devices at work in syntactic explanation and representation.

On the other hand, it is also true that core semantic constructs, e.g. reference markers or phase quantification, are involved in the new formulation of binding constraints. But also on the semantic side, and from a standard perspective, these constructs exhibit a considerable degree of idiosyncrasy. In formal semantics, representations of natural language expressions are interpreted alongside models whose elements are extra-grammatical entities, such as individuals, events, relations, instants of time, etc.. Obliqueness quantifiers, however, seem to presuppose for their interpretation a model that includes intra-grammatical entities, such as reference markers and obliqueness relations.

Component vs. composition of meaning

Despite eventual *prima facie* perplexity, there seems to be reasonable justification to think that the grammar level of Semantics is better suited to shelter binding constraints.

From a practical perspective, unlike Syntax, Semantics provides essential devices needed to express quantification, even if obliqueness quantifiers happen to require an unheard-of intra-grammatical format given the kind of entities its restrictor and scope range over. Roughly, one just has to allow the model against which semantic representations are interpreted to include entities belonging also to the intra-grammatical universe (such as reference markers, obliqueness relations, etc.), in parallel to the more usual extra-grammatical universe of events, times and individuals (such as man's first landing on the moon, Plato, this dissertation, and so on). Accordingly, in such setup, binding constraints will be additional components to the meaning of the nominals they happen to be associated with.

Another line of reasoning supporting the view that binding constraints should be sheltered in Semantics is based on a somewhat more distanced conceptual analysis of their role in grammar.

The fact that binding constraints have to date been addressed mostly by syntacticians should not presuppose that binding constraints belong to the realm of Syntax, even less that "binding theory captures the distribution of pronouns and anaphors" (cf. Reinhart and Reuland, 1993, p.657). Taking Syntax as being

concerned with the well-formedness of expressions, it appears that binding constraints should not in fact be counted on the side of Syntax. Although they impose conditions on the interpretation of nominal anaphors based on syntactic geometry, these constraints do not express well-formedness requirements concerning the grammaticality of expressions. By replacing, e.g. a pronoun by a reflexive, one is not turning a grammatical construction into an ungrammatical one, even if one assigns to the reflexive the antecedent adequately selected for the pronoun. In that case, one is just asking the hearer to try to assign to that sentence a meaning that it cannot express, in the same way as what would happen if one asked someone whether he could interpret *The red book is on the white table* as describing a situation where the white book is on the red table. In the latter case, given how they happen to be syntactically related, the semantic values of *red* and *table* cannot be composed in a way that their sentence could be used to describe a situation concerning a red table, rather than a red book. Likewise, if we take *John thinks Peter shaved him*, given how they happen to be syntactically related, the semantic values of *him* and *John* cannot be composed in a way that this sentence could be used to describe a situation where John thinks that he shaved himself. The basic difference between the two cases is that, while in the first, the composition of the semantic contributions of *red* and *table* is constrained by local syntactic geometry, in the latter, the composition of the semantic contributions of *Peter* and *him* is constrained by non-local syntactic geometry.

Accordingly, there seems to be principled justification to count binding constraints on the side of Semantics, as semantic composition is typically one of the concerns of this discipline. Note however that, while providing further support for a semantic account of binding constraints, this argument points towards a role of obliqueness quantification in grammar rather than being a component of meaning, as previously suggested. Obliqueness quantifiers appear now rather as determinants of the composition of meaning.

In what follows we try to set up a semantic account of binding by exploring these two possible views on obliqueness quantification: As a component of meaning and as a determinant of meaning composition. In order to have a formal ground to discuss this account, we adopt the framework of Discourse Representation Theory (DRT) under its standard shape, as presented in Kamp and Reyle, 1993, and van Eijck and Kamp, 1997.

A framework for semantics

DRT originated from the studies of Kamp (1981) and Heim (1982), which incorporated widespread basic insights on the semantic representation of discourse and anaphora that had been maturing in the literature, such as inter alia Karttunen, 1976, Webber, 1982, or Sag and Hankamer, 1984. The key idea of DRT, as can be found in a recent formulation by van Eijck and Kamp (1997, p.181), is that

“... each new sentence *S* of a discourse is interpreted in the context provided by the sentences preceding it. The result of this interpretation is that the context is updated with the contribution made by *S*; often an important part of this process is that anaphoric elements of *S* are hooked up to elements that are present in the context.... An important aspect of this kind of updating of contexts is the introduction of elements – so called reference markers or discourse referents – that can serve as antecedents to anaphoric expressions in subsequent discourse.”

In its rudiments, DRT appeared primarily as an attempt to cope with a certain type of problematic anaphoric links reported since the Middle Ages and introduced in the modern semantic literature by Geach (1962). This problem became known as “donkey” anaphora due to apparently unavoidable preference of researchers for examples illustrating it which systematically report on a situation where a donkey is beaten, like the following one:

- (1) If Pedro owns a donkey, he beats it.

The problem with this type of anaphoric link is that traditional approaches, which tend to interpret pronouns either as bound variables or as constants and existential NPs as expressing existential quantifiers, are at odds to capture the reading where Pedro beats the donkeys he owns. The option for interpreting the pronoun as a constant has to be discarded since a donkey does not refer to any individual that can be the denotation of a constant. And the option for interpreting it as a variable bound by the quantifier expressed by a donkey is not feasible as quantifiers’ scope typically does not extend beyond the sentence where they occur (for a detailed discussion of this issue, vd. Reinhart, 1987). But even if that were possible, the resulting reading – a reading where there is at least one donkey beaten by Pedro – is other than the intended one – a reading where any one of the donkeys owned by Pedro is beaten.

The solution proposed in DRT involves the adoption of a level of semantic representation, Discourse Representation Structure (DRS), consisting of two parts. One is a set of reference markers, known as the universe of the DRS. The other is a set of conditions, which loosely speaking may be said to specify properties of the

elements of that universe. These two parts are entered in the DRS respectively in its upper and lower sections as illustrated below.

(2) a. A man entered. He was whistling.

b.

m h
man (m) enter (m)
h = m whistle (h)

For an elaboration on the status of this level of semantic representation – as “mental representation”, “partial model”, or “logical form” – see Asher, 1993, pp.63ff.

This format for semantic representation does not involve a formal language with an expressive power greater than the one of first order logic. A translation procedure from DRT to first order logic can be found in Section 1.5 of Kamp and Reyle, 1993, or more concisely in Definition 4.5 of van Eijck and Kamp, 1997, p.193, where the basic idea is that reference markers are variables to be bound by first order existential quantifiers and conjunction is the connective relating the conditions inside the boxes of DRSs. Nevertheless, given the specific format and interpretation procedure of DRS, the difference between first order logic and DRT is claimed “...to reside entirely in the different way in which DRT handles context...” and in the fact that “...reference markers behave more like variables in programming languages than like variables in ordinary first order logic.” (van Eijck and Kamp (1997, pp.194,211)).

Bearing on this separation between reference markers and conditions, rules for constructing DRSs from syntactic representations can be conceived in a way that the representation of the “donkey” sentences can be interpreted as conveying the problematic, intended reading. The indefinites induce representations that correspond to open sentences. The quantificational determiner of the subject, if it is a relative “donkey” sentence, or the quantificational adverb, as always or usually, if it is a conditional “donkey” sentence, are interpreted as unselective quantifiers binding the variables in the open sentence obtained from the indefinite which is the antecedent of the donkey pronoun – vd. Chierchia, 1995, pp.39ff and Lappin and Francez, 1994 for details and alternative approaches.

In DRT both indefinites and definite descriptions are viewed as introducing a reference marker in the universe of the relevant DRS. The difference is that unlike a reference marker corresponding to an indefinite, the reference marker of a definite description must be anaphorically linked to another reference marker. This captures the presupposition associated with definite descriptions that the speaker takes the referents he is evoking as being in some sense familiar to the hearer – cf. Kamp and Reyle, 1993, pp.248, and also pp.246ff concerning proper names.

Intensive research since its inception have suggested that DRT can be easily extended to account for phenomena other than “donkey” anaphora. It has been suggested also – see, for instance, chapter 4 on anaphora of the survey *The State of the Art in Computational Semantics* by Cooper et al. (1994) – that DRT provides a highly flexible and conspicuous framework for the study of anaphora in general.

5.2 Obliqueness Quantification as a Component of Meaning

Having introduced the framework we will be using, we can now focus on binding constraints under the conception which envisages their quantificational import as a component of the meaning of the nominals with which they happen to be associated. As discussed above, this implies providing a semantic representation for the intra-grammatical quantificational force of binding constraints. In what follows we pursue this goal by using pronouns as a case study.

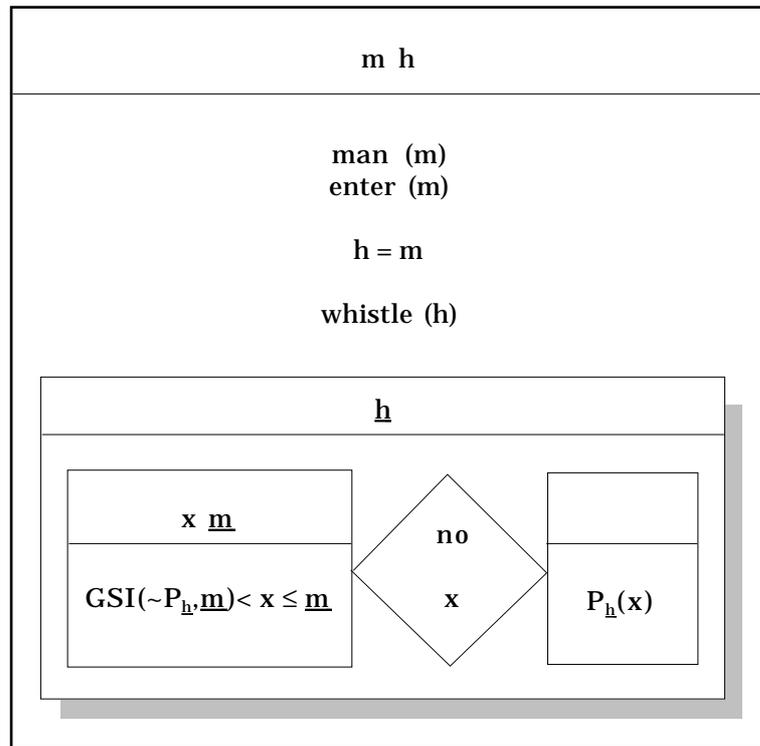
Representing intra-grammatical quantification

In order to represent obliqueness quantification in DRT as intra-grammatical quantification, we need a slight extension of the DRS language. This extension is

meant to allow us to make available reference markers and conditions that are to be specifically interpreted against intra-grammatical entities. We then use shaded DRSs to indicate that conditions and markers in them should be interpreted against a specific interpretive model with such intra-grammatical entities. Additionally, given that in shaded DRSs we want to talk about a given discourse referent r possibly present in non-shaded DRSs, we adopt the convention that an underlined letter in shaded DRSs \underline{r} is to be interpreted “intra-grammatically” as referring to the reference marker with the same but not underlined letter r (which should not be confused with the “extra-grammatical” referent of r).

With this simple extension in place, we can improve (2)b. and obtain (3) as the semantic representation of (2)a., where the binding constraint associated with the pronoun is now included.

(3)



It may be objected that this extension of the DRS in (2)b. is in a certain sense innocuous or irrelevant as the new shaded DRS corresponding to the binding quantifier does not add any real constraint to the meaning already represented by the DRS conditions in (2)b. In fact, the conditions in (2)b. about h are somewhat stricter than the conditions expressed by the shaded DRS in (3).

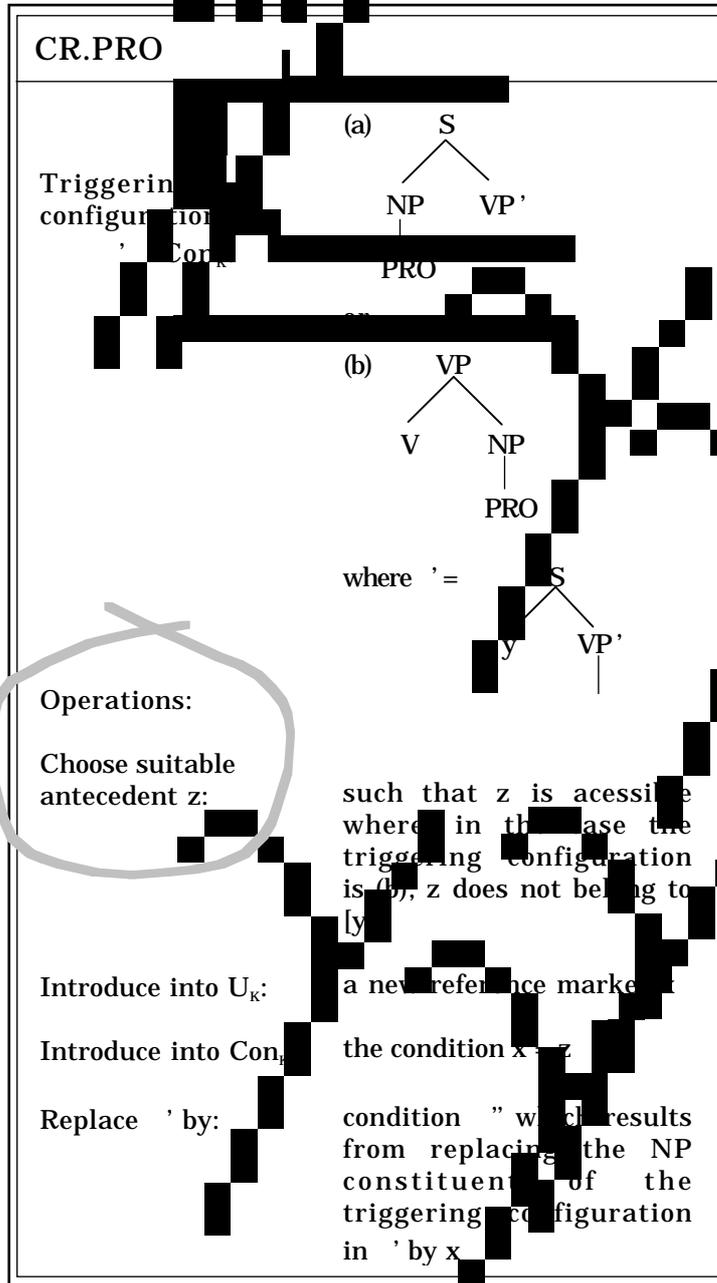
The shaded DRS states that no reference marker less oblique than nor as oblique as the antecedent m is local with respect to h in the relevant obliqueness relation, i.e. m is not a local o-commander of h . The DRS of (2)b. in turn states the more strict condition that m , a reference marker introduced during the parsing of a previous sentence, is the antecedent of h .

This objection, however, makes sense only if one ignores the fact that, in the framework of DRT, unlike other DRS conditions, the inclusion of conditions accounting for anaphoric links, such as $h=m$, does not result from strict semantic analysis. As it was extensively discussed in chapter 2 Heuristics and Constraints and is acknowledged in Kamp and Reyle, 1993, p.70, the inclusion of this type of conditions results from “all sorts of considerations, non-linguistic as well as linguistic, that makes a particular choice of the antecedent suitable”. This is further clarified in van Eijck and Kamp, 1997, pp.221-222:

“DRT [...] is not a theory of anaphora resolution: the theory itself tells us little about how to select the intended antecedent for a given anaphoric expression [...] Arguably, this is as it should be. It would be unreasonable to demand of a theory of linguistic semantics [...] that it incorporate a detailed account of anaphora resolution, which would have to rely on a host of pragmatic principles as well as on an indefinite number of world knowledge. It seems not unreasonable, however, to demand of such a theory that it offer a suitable interface to other components of a comprehensive theory of meaning which are designed to deal with anaphora resolution [...] and to allow these other components to come into action at those points when the information needed for anaphora resolution has become available and the resolution is necessary for the interpretation to proceed.”

Consequently, although they are well-formed expressions of the DRS language and contribute to the representation of the meaning of the phrase at stake once written down, anaphoric conditions such as $h=m$ are ad hoc from the strict point of view of semantic analysis. Unlike other DRS conditions, their inclusion in DRSs is totally stipulatory as they do not result from any systematic mapping from syntactic representation into semantic representation. They are there purely for the sake of permitting the interpretation of DRSs, which without them would remain open, uninterpretable formulae.

(4)



The ultimate root of this circumstance, with respect to pronouns, can be spotted in the Construction Rule that yields semantic representations for pronominal NPs. In

that rule, stated in Kamp and Reyle, 1993, p.238, and repeated above, one is merely told to “choose a suitable antecedent” that can enter in the anaphoric condition – actually, a simple wording for pushing back the complex issue of anaphor resolution.

Consequently, the inclusion of the shaded DRS in (3) is not an innocuous contribution for the semantic representation of the pronoun. It contributes to reduce the stipulatory character of the anaphoric condition $h=m$ inasmuch as m is constrained not to be one of the local o-commanders of h .

Underspecification of anaphoric links

These considerations, while partially answering the objection on the eventual irrelevance of the shaded DRS in (3), also suggest how we should move towards a more thorough representation of (2)a. where the semantic contribution of binding constraints is fully unfolded.

First, for the sake of permitting that the importance of representing obliqueness quantification comes fully to light, we opt for an attitude of strict parsimony in the construction of semantic representations. In the process of constructing DRSs from corresponding syntactic representations, we only include semantic conditions that can be non-arbitrarily obtained from the syntactic input for the construction rules of DRSs.

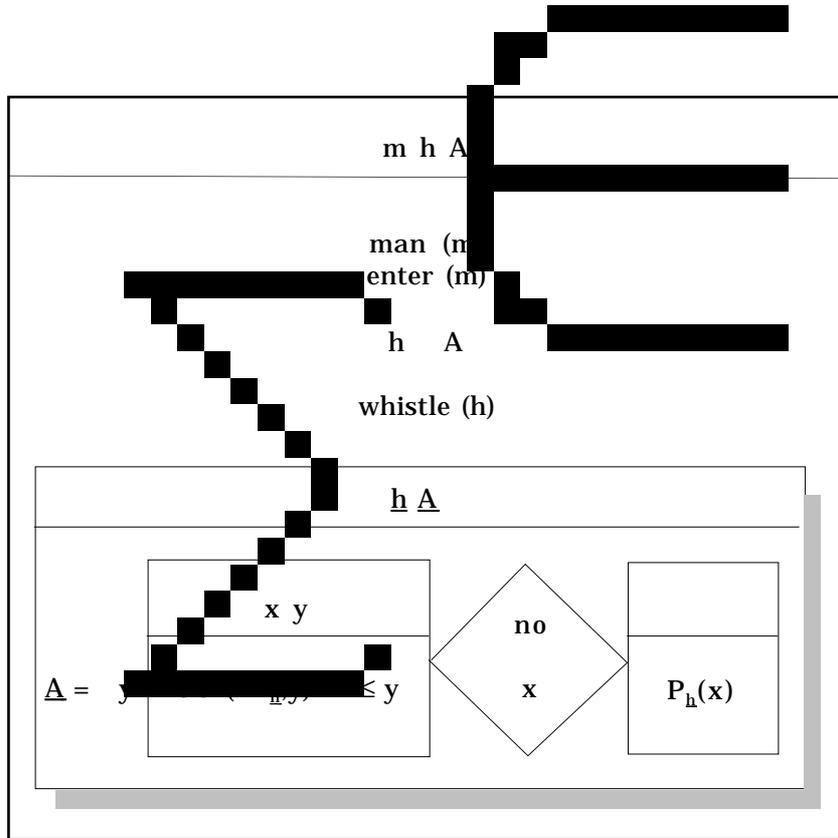
Second, we explore the true potential of binding constraints for the construction of semantic representations that can be used as interfaces with other language processing components, namely anaphor resolvers. This is accomplished by means of two concomitant moves.

On the one hand, we opt for an underspecified semantics (vd. Reyle, 1993) of the anaphoric capacity of the pronoun. In this light, the stipulatory condition $h=m$ is replaced with the condition $h \in A$, where A is the set of antecedent candidates complying with the binding constraint at stake.

On the other hand, we use the semantic representation of binding constraints to define the set A . We use the Abstraction operator \underline{A} (Kamp and Reyle, 1993, pp.311ff) to obtain the reference marker \underline{A} that stands for the set of markers that satisfy the duplex condition of the shaded DRS as y . Consequently, \underline{A} may be conceived as including all reference markers that, when used as antecedents of h , comply with principle B, i.e. markers which are not local o-commanders of h .

Following these guidelines, (3) receives this reformulation:

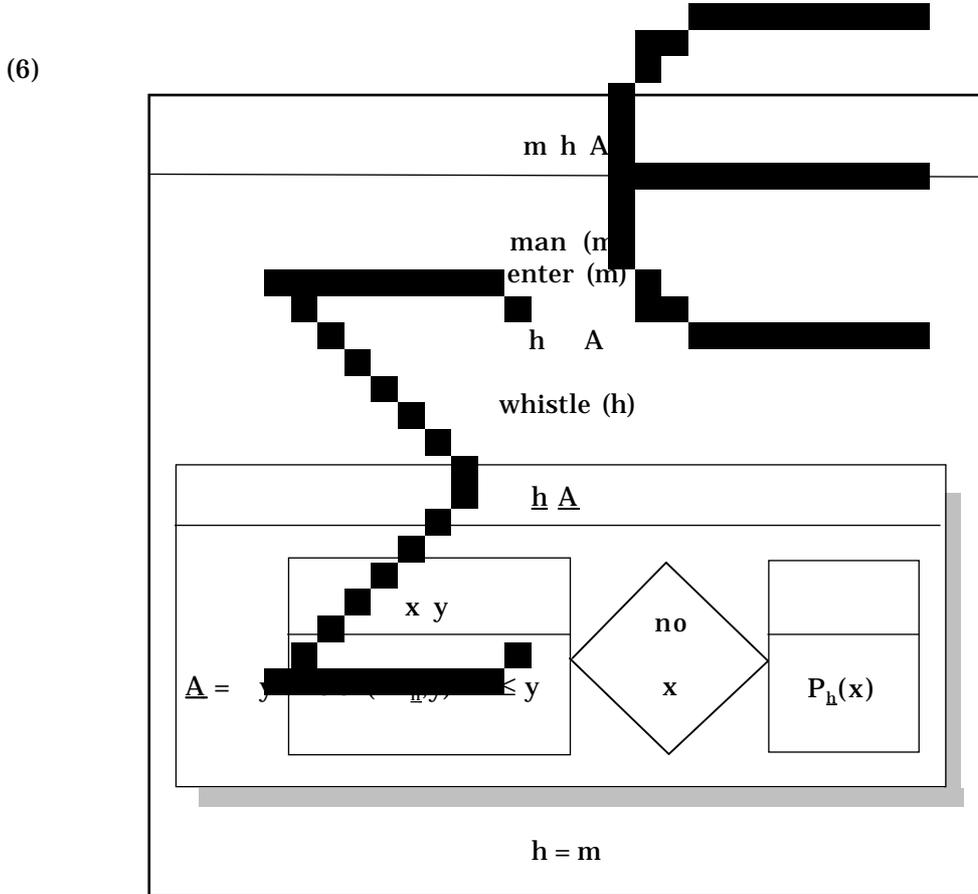
(5)



This DRS is then the proposed semantic representation of (2)a.. We find in it a semantic representation of the intra-grammatical quantificational force of the pronoun.

Note that this representation not only ensures a statement of that significant piece of linguistic knowledge associated with the pronoun, but appears also as instrumental in other important related purposes. This representation avoids the stipulatory encoding of the anaphoric capacity of pronouns. Additionally, it contributes also to setting up suitable interface points between grammatical representations and anaphor resolvers.

Underspecified representations such as (5) are an adequate basis for monotonic improvements by means of further specification of the relationships between anaphoric expressions and their antecedents. It follows quite naturally that a module for anaphor resolution should output (6) taking (5) as input, where the condition expressed in $h \ A$ is restricted to conjunction with $h=m$, whereby the anaphoric resolution of *he* in (2)a. is achieved (following the basic idea of Harper (1992)).



Toward a uniform dichotomic semantics for nominals

From the discussion of our case study (2)a., involving a pronoun, it becomes evident how the representation of other anaphoric NPs – be they nonpronouns, short-distance reflexives or long-distance ones – should be handled. Each anaphoric nominal should thus receive a semantic representation similar to the representation of the pronoun in (5) with an appropriate duplex condition expressing the corresponding obliqueness quantifier.

Having devised how to accommodate obliqueness quantification in grammar as a component of the meaning of anaphoric nominals, we can turn now to some more speculative issues, arising from this type of account of binding constraints.

We are aware of the highly intriguing semantics of nominals, as illustrated for instance – and taking only the case of the so called definite descriptions – in Ostertag's (1998) reader. Cross-cutting distinctions between mass and non-mass terms, generic vs. non-generic readings, attributive vs. referential, and many others, form an intricate corpus of research themes that intersects the issue of the anaphoric behavior of nominals. But let us for a moment make the distinction, generally assumed in the linguistics literature, between quantificational and non-quantificational NPs as being a major one that partitions the set of NPs into two disjoint classes. Also, given that the interpretation of different sorts of non-quantificational NPs exhibit, in some degree, a dependency from antecedents, it would not be unreasonable to count non-quantificational NPs as bearing some kind of anaphoric capacity in a framework like DRT, as suggested by van Eijck and Kamp (1997, p.189).

There is a very interesting point here that our exercise of viewing obliqueness quantification as a component of the meaning of anaphoric NPs brings into play.

What tends to be seen as quite a sharp dividing line between quantificational and non-quantificational NPs becomes blurred, because non-quantificational anaphoric NPs also appear to have a quantificational component in their meaning. On its own, this is a facet of anaphoric nominals clearly underlined above in a previous subsection. However, it acquires a considerable new dimension when it is put on a par with the eccentric anaphoric capacity of quantificational NPs exhibited in examples of e-type anaphora.

- (7) Between eleven and nineteen senators voted against the proposal. They were afraid of riots in the streets.

In examples such as the one above, the contribution of between eleven and nineteen senators to the meaning of the discourse where it occurs is not confined to the quantificational force it expresses. Under the framework of DRT, it can be seen as contributing also with a reference marker that serves as the antecedent of they (cf. Kamp and Reyle, 1993, p.309). Contrary, however, to what happens with non-quantificational NPs introducing reference markers in DRSs, this reference marker of between eleven and nineteen senators is introduced exclusively to serve as an antecedent to anaphors. This implies that the eventual referring force of quantificational NPs, when it is elicited, is somehow restricted to an intra-grammatical role.

These observations are meant to suggest that a semantic account of binding constraints may turn out to be instrumental in justifying a uniform view of

5.2 Obliqueness Quantification as a Component of Meaning

nominals as always having two twin semantic facets: Quantificational and referential. Thus, what would partition the set of nominals would not be their being quantificational or not (or referential or not), but the dimension, intra-grammatical or extra-grammatical, at which the nominal at stake expresses its respective quantificational and referential forces.

In this scenario, every nominal is envisaged as always making contributions in both of the two semantic dimensions.

However, a kind of symmetry effect exists inasmuch as nominals expressing extra-grammatical quantification have intra-grammatical referential capacity. Vice versa, nominals expressing intra-grammatical quantification have extra-grammatical referential capacity:

(8)

		REFERENCE	
		INTRA-GRAMMATICAL	EXTRA-GRAMMATICAL
QUANTIFICATION	INTRA-GRAMMATICAL	every student most senators three donkeys ...	him John the student ...
	EXTRA-GRAMMATICAL		

As a hypothesis for a new overall understanding of the semantic abilities of nominals, this is an appealing research path to be further pursued. However, a thorough examination of all of its implications lies beyond the scope of the present dissertation.

5.3 Binding Constraints as Determinants of Meaning Composition

Having examined how binding constraints can be accommodated in Semantics as components of the meaning of anaphoric expressions, we turn now to the other possible semantic account referred to above in Section 5.1. Instead of focusing on their deep quantificational structure, this approach focuses on the more immediate role of binding constraints in grammar, envisaging them as determinants of the (non-local) composition of the meaning of anaphoric nominals.

Content of representations and content of rules

Although this view of binding constraints as determinants of meaning composition appears as a quite natural one taking into account a *prima facie* analysis of their contribution to the interpretation of anaphors, it has received a very meager recognition in the literature.

In DRT, the Construction Rule for pronominals, presented above in (4), somehow tries to accommodate the binding constraint for this type of anaphors by means of the condition that the reference marker z (arbitrarily) taken as antecedent of the pronoun “does not belong to $[y]_k$ ” (where y is the reference marker corresponding to the subject of the sentence being reduced, and $[y]_k$ is the set of reference markers identified with y). However, this is more of a rough and ready account of the binding restrictions on pronouns than a full-fledged accommodation of binding constraints in the Construction Rules of DRT.

To the best of our knowledge, while making their argument against the syntactic approach to binding constraints, Gawron and Peters (1990a, p.139) are the only ones

arguing for an eventual conception of these constraints as conditions on the composition of semantic representations:

“Our basic strategy is to recast the sort of Binding Theory presented in Chomsky, 1981 as a set of constraints on the relationship between structure and meaning-descriptions”

The interest of this passage is not confined, however, to the expression of a point of view which we find justification to share. Interestingly enough, it ends with a mark to a footnote. In this footnote, the authors elaborate further on the argument against a syntactic reduction of binding constraints. They point towards another possible semantic account of these constraints which would consist in viewing them alternatively as components of the meaning of nominals. That is exactly the spirit (and the letter) of the approach we developed above and to which the newly uncovered quantificational nature of binding constraints possibly gives an enhanced empirical support that Gawron and Peters did not have at hand at the time they wrote that “a rather natural option [...] is to attempt to state binding conditions on the contents themselves [...] [A possibility] we have not developed in this work.” (Gawron and Peters, 1990a, p.141, footnote 3).

Principled constrained construction rules

In the standard formulation of DRT presented in Kamp and Reyle, 1993, syntactic and semantic representations are interfaced via construction rules such as the one in (4). These rules are responsible for the construction of DRSs and hence, for the composition of meanings. They come under the basic format of conditionals and are functions from (reducible) DRSs to (reduced) DRSs.

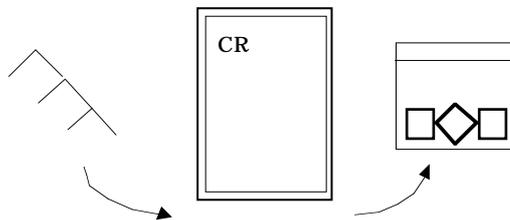
The antecedent of the conditional states the application conditions of the rule. It describes a piece of grammatical representation termed as Triggering Configuration.

The consequent states the contribution of the rule for the construction of the relevant DRS. This contribution appears under the heading of Operations. These operations involve destructive moves – requiring the elimination of at least parts of the triggering configuration –, and constructive moves – involving the introduction of reference markers and conditions into the DRS. Construction rules are successively applied in the process of building up a semantic representation from a syntactic one until an unreducible DRS is reached, that is until no piece of syntactic representation from the initial DRS is left to be eliminated.

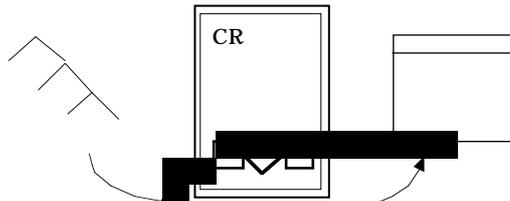
Coming back to our case study, it is easy then to envisage a way to account for the binding constraint on pronouns as a determinant in the composition of its meaning. Roughly, what one needs is just: (i) to associate with each NP a unique reference marker before the application of the Construction Rules; (ii) to displace the quantification in the semantic representation (5) to the construction rule in (4), as pictorially suggested below:

(9)

Component of meaning approach:

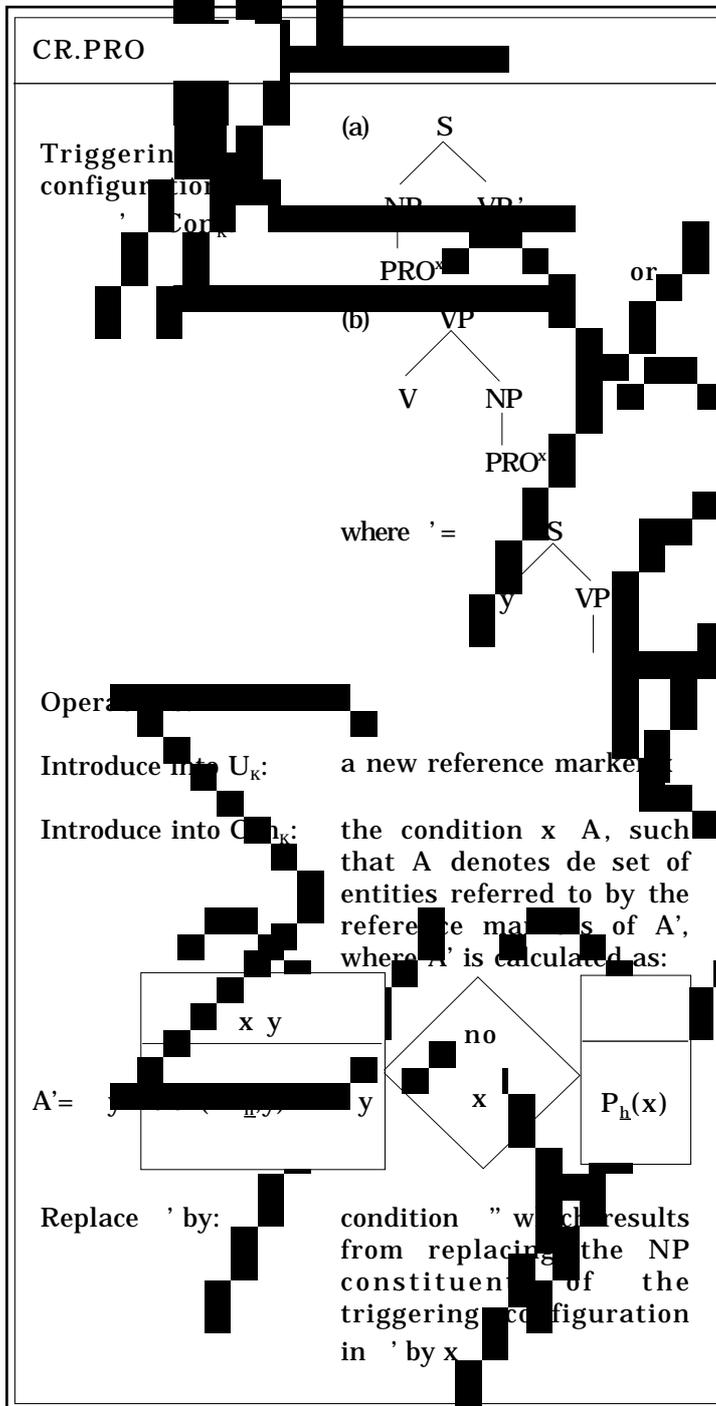


Composition of meaning approach:



This can be done with a simple reshuffling of the construction rule for pronouns. In the Operations section, the reference marker x to be introduced in the (reduced) DRS is the one which happens to have been previously associated with the pronoun at stake (and appears in the input/triggering configuration as a superscript of the relevant NP). Also, the instruction for the ad hoc introduction of the anaphoric link condition $x=z$ is overridden by the instruction to introduce the underspecified requirement that $x \in A$, and λx the basis of abstraction over the obliqueness quantification associated with pronominals. This gives rise to the following reformulation of the construction rule for pronouns:

(10)



It follows quite naturally from these considerations about our case study that the construction rules for other anaphoric nominals can be submitted to similar reshuffling. This set of adjustments in the construction rules involving anaphoric nominals is thus what is needed for a semantic account of binding constraints which envisages them as determinants of meaning composition.

Between two approaches

Having designed an approach to binding constraints as conditions on the relationship between syntactic structure and meaning representation, we can now contrast it with the first approach for their accommodation in grammar we designed.

A first point to note is that these two approaches – meaning-component vs. meaning-composition – share the feature that they are both semantic accounts of binding constraints. This results from the fact that, in a standard view of the architecture of grammar, not only the representation of meaning but also the composition of meanings at the interface between Syntax and Semantics count as matters belonging to the realm of Semantics.

Two other points worth noting stem from the fact that the meaning-composition approach provides the same welcome side-effects as the meaning-component approach. On the one hand, it removes stipulativeness from the construction rules for anaphors regarding the ad hoc introduction of anaphoric linking conditions. On the other hand, it also contributes to setting up the same kind of underspecified interface points with anaphor resolvers.

The major point where the two approaches seem to diverge is at the hypothesis on the uniform dichotomic semantic nature of nominals. While the meaning-component approach stimulates a thorough inquiry into the possible double bipolar nature of nominals – quantificational/referential vs. intra-/extra-grammatical –, the meaning-composition approach definitely demotes this issue from the research agenda.

This being said, we think that it turns out to be manifest that the results obtained above on binding constraints do not offer compelling arguments for opting for one of the approaches in detriment of the other. We believe that possible advantages of one of the approaches over the other may become clear when brought into the context of the discussion of a wide range of issues which, even though related, are accessory to the core facts of binding.

It may happen that a more theoretical and philosophically informed discussion on the semantics of nominals will bring to light justified reasons to support the view that nominals have a double bipolar semantic nature. In such circumstances, the meaning-component approach to binding constraints would appear as preferable over the meaning-composition one. Or, to give another example, it may happen that a more practical discussion on the computational implementability of binding constraints (cf. chapter 6 Constraint-based Specification and chapter 7 Computational Implementation) leads to engineering requirements that compel us to the inverse position, i.e. to adopt the meaning-composition approach instead of the meaning-component one.

5.4 From Anaphor Resolution to Reference Processing

It follows from the discussion of the semantic account of binding constraints, be it envisaged under the meaning-composition or the meaning-component perspective, enhances the setting up of new interface points between the representation of meaning and the modules for anaphor resolution. What was identified above as set A , resulting from abstraction and quantification, can be further restricted down by filters other than binding constraints. In an integrative, filter and rank approach to anaphor resolution, after A having been appropriately sieved, the elements of the resulting set are ordered with the help of preferences. In light of this, we could perhaps slightly reformulate and (10) above so that instead of $h \in A$ we would have a more correct $h \in A_s$, A_s being the scaling down of A into A_s by virtue of the filtering process in anaphor resolution.

Note that this scaling down may be due to some of the several factors we discussed in chapter 2 Heuristics and Constraints and in Section 4.4, in particular to what was termed there as reciprocal validation and mediated antecedency.

With the above type of representation of the anaphoric potential of a given anaphor h , it is possible to handle reciprocal validation by requiring that A_h be scaled down to A_{hs} by removing a reference marker x from A_h if h is not in A_x .

Also the effect of mediated antecedency referred to in the last chapter and holding for non-reflexives can be neatly expressed in the present setup. Given a nonreflexive discourse referent h and the antecedent x it happens to be related with by the introduction of a condition $h=x$, for every y such that $x=y$, $y \in A_h$. Not given the reflexivity of the relation '=', this formulation will account both for the effect of mediated antecedency and reciprocal validation of the non-reflexives.

But anaphor resolution is not all that matters in the processing of anaphora. Besides the resolution of anaphors, the processing of reference comprises also being able to find the adequate semantic relation between anaphors and antecedents.

In what follows, we argue that the semantic account of binding constraints designed above not only provides suitable interface points for anaphor resolution, it also constitutes an appropriate basis for the wider purpose of reference processing. For the sake of making it clearer, in the discussion to be developed below we assume that the process of anaphor resolution has been completed. We thus assume that the anaphoric reference markers in DRSs have already been assigned their antecedents, so instead of dealing with a set of antecedent candidates we will deal with the elected antecedent.

Coreference

The simplest form of semantic relation between anaphor and antecedent is the one where anaphor and antecedent receive the same interpretation. This is exemplified below (superscripts indicate reference marker contributed by the phrase; subscripts indicate reference markers taken as antecedents):

- (11) John/a man/your friend^x entered the room at noon. He/the man^y_x was whistling.

As discussed above, this type of anaphora is accounted for by introducing two reference markers in the DRS universe, standing for the antecedent and for the anaphor, respectively, and an anaphoric condition which imposes that the two markers refer to the same entity. Concentrating only on the details that are essential for reference processing, the semantic representation of this kind of anaphoric link can be said to pattern as depicted below, where x and y stand for the reference marker introduced, respectively, by the antecedent and by the anaphor:

- (12) ... $y=x$...

Bridging, indirect or associative

In a type of anaphoric dependency known as associative, indirect or bridging, although the interpretation of the anaphor depends on the interpretation of the antecedent, their semantic values are not identical. Typically, the interpretation of the anaphor is an elaboration upon the interpretation of the antecedent (for a broader conception of bridging see Asher and Lascarides, 1998). As extensively discussed and classified in Strand, 1996 and Poesio and Vieira, forth., under this kind of anaphoric link the anaphor may refer to an entity that is e.g. an element or part of the denotation of the antecedent or an entity that includes the denotation of the antecedent.

- (13) a. Peter could not use his computer^x. The monitor^{y_x} was not working.
 b. The priest canceled the marriages^x planned for tomorrow. He realized that the brides^{y_x} were not baptized.

In this type of anaphora the semantic relation between antecedent and anaphor is not as direct as in the previous case of coreference. The semantic linking needs to be mediated by a function *brid* that ensures the “bridging” between the reference marker of the antecedent and the reference marker of the anaphor. Thus, in this sort of anaphoric link, the semantic representation is set up in accordance to the following schema:

- (14) ... $y = \text{brid}(x) \dots$

Function *brid* is supposed to be sensitive to the context and to the possible semantic relations between the content of antecedents and anaphors, and to deliver the convenient intermediating relation between the corresponding reference markers. It seems not to be unreasonable, at least from the point of view of fostering a uniform semantic representation of apparently different types of anaphora, to see coreference as a special case of bridging anaphora, where *brid* would be the identity function.

A specially interesting example of the context sensitivity of function *brid* is given below by means of a sentence exhibiting a kind of anaphoric link usually known in the literature as “paycheck” anaphora:

- (15) a. The man who gave his paycheck^x to his wife is wiser than the one who gave it^{y_x} to his mistress.
 b. Susan said to her husband^x that Mary should not divorce from him^{y_x}.

In (15)a. the pronoun *it* has *his paycheck* as antecedent but the semantic link between anaphor and antecedent holding there is not coreference. Apparently, we may have here a specific case of bridging anaphora where the antecedent itself contains an anaphor. Arguably due to that fact, while taking the representation of the antecedent as argument, the bridging function *brid* may provide a new resolution to the anaphor inside it. Accordingly, the pronoun *it* can be interpreted as referring to the paycheck of the man that gave it to his mistress, not the paycheck referred to by its antecedent, the paycheck actually given by the other man to that man's wife.

In this light, it is thus interesting to contrast (15)a. with (15)b.. Although in the latter, the anaphor *him* also has an antecedent which contains an anaphor – *her* in *her husband* – the kind of “paycheck” bridging seems not to be possible here. We hypothesized that this is probably due to the fact that the syntactic structure parallelism occurring between the sentence with the antecedent and the sentence with the anaphor in (15)a. does not occur in (15)b.. If this hypothesis turns out to be correct, then this is a very eloquent example of the context sensitivity of the bridging function. On the other hand, it eloquently suggests that similarities should be explored between the possible constraints on this kind of anaphoric links, and the usual “parallelism” constraints on ellipsis, in particular the specific behavior of elided phrases whose non-elided counterpart also contains anaphors (vd. Cooper et al., 1994a):127ff and Cooper et al., 1994b, pp.47ff for an overview on these issues).

E-type anaphora

There are other cases where the relation between the interpretation of the antecedent and the interpretation of the anaphor is also not as direct as in the examples of coreference. Again, in the examples below, the semantic relation between anaphor and antecedent requires some inference:

- (16) Few senators^x voted against the proposal. They^y_x were afraid of riots in the streets.

Yet, the kind of intermediate inference at stake here seems to differ from the one discussed for bridging anaphora. Since the antecedent *few senators* is not interpreted as denoting a set of senators – but a set of sets, in line with the Generalized Quantifier approach to the semantics of nominal quantification, cf. Barwise and Cooper, 1981 –, it is not the case that the anaphor *they* could be interpreted as a part of, or as including – in any conceivable sense – the denotation

of the antecedent. Rather, when taking few senators as its antecedent, they is interpreted as referring to the set of senators who voted against the proposal, in what is known as a case of e-type anaphora.

Further interesting evidence can be found in examples such as those depicted in (17) – (17)a. is taken from Elworthy (1995, ex.(29)), and (17)b. from Cooper et al. (1994, ex.(3.54)).

- (17) a. Every doctor^x was working late last night. They^y_x/ He^y_x had too many patients.
- a'. Every doctor^x has his^y_x/ their^y_x own patients.
- b. Every meeting had a chairperson^x. He^y_x was selected from one of the participating companies.
- b'. Each department has a computer^x. They^y_x were bought to XPTO company.

The contrast (17)a./a'. and the impossibility of having a singular anaphor in (17)a. should be taken as reinforcing evidence that e-type anaphora cannot be seen as a variant of bound anaphora, a kind of anaphoric relation illustrated in (17)a', and which will be discussed below. Example (17)b'. shows, in turn, that indefinites can play the role of antecedents in e-type anaphora, and the contrast between (17)b. and (17)a. reveals that, unlike NPs, the interpretation of indefinite NPs can enter anaphoric relations of the coreferential type.

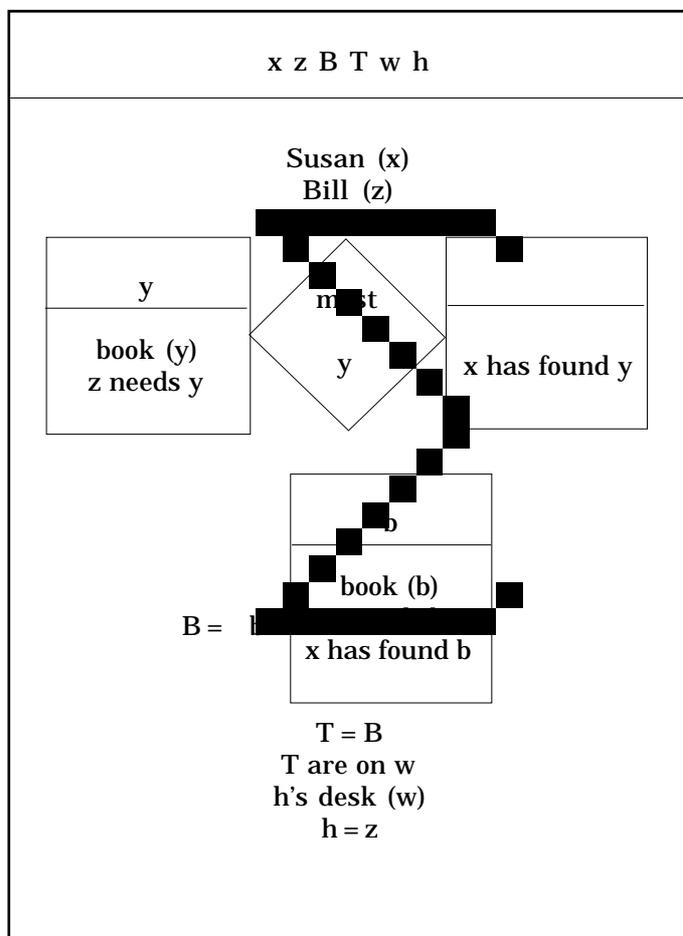
The identification of e-type anaphora, as well as the term e-type, is due to Evans (1977), (1980). Following the insights of Kamp and Reyle (1993), under this kind of anaphoric relation the anaphor refers not to the denotation of its quantificational antecedent but to the set of entities that comply with the conditions stated both in the restrictor and in the domain of the corresponding quantificational determiner. Therefore, the semantic relation between anaphor and antecedent presupposes the mediation of a function over the representation of the antecedent phrase whose outcome is the reference marker which can serve as antecedent.

As illustrated below, the duties of this function in DRT are taken by the Abstraction operation (Kamp and Reyle, 1993, pp.311ff), which provides a reference marker standing for the set of entities satisfying the condition on b at the right of the NP. This operation acts on unreducible DRSs and provides a sort of inference mechanism on semantic representations useful for the account of e-type anaphora.

For the sake of illustration, the semantic representation in DRT of the sentence below in a. is given in b. – cf. Kamp and Reyle, 1993, ex.(4.17).

(18) a. Susan has found most books which Bill needs^B. They^T_B are on his desk.

b.



Given that the mediation function should be placed now on the side of the antecedent, the general schema that a semantic representation of e-type anaphora should be given is as sketched below, where *abst* stands for the abstraction function, which applies to the semantic representation of the antecedent phrase.

(19) ... $X = \text{abst}(\dots)$... $y = X$...

This implies that the contribution of quantificational NPs to the obliqueness partial order – upon which the binding constraints are defined – requires the mediation of an abstraction function. Specifically, a quantificational NP contributes to the obliqueness relation with a reference marker that results from abstraction over the conjunction of the conditions corresponding to its restrictor and

domain (below we will consider other sort of contribution to the obliqueness relation by quantificational NPs in bound anaphora relations).

Additionally, this also explains why e-type anaphora does not hold when an anaphor occurs in expressions corresponding to the restrictor or to the domain of its quantificational NP antecedent, as in (17)a'. with respect to they. In this case, there is a sort of interpretive deadlock inasmuch as for the antecedent marker to be set up, the anaphor – occurring in the domain of the quantifier – must be interpreted, but for the anaphor marker to be interpreted, in turn, the antecedent marker must have been set up. This implies that when an anaphor and its quantificational antecedent occur in the same predication structure, an e-type anaphora does not escape this interpretive deadlock, which means that it is not possible.

Below, we discuss why, for independent reasons, bound anaphoric links between anaphors and quantificational antecedents – as in (17)a'. with respect to his – escape this deadlock and present a complementary distribution as regards e-type links.

Split antecedents

E-type anaphoric links are not the only types of link, whereby the reference marker associated with the antecedent phrase involves a certain degree of inference upon the semantic representation of the expression in which the antecedent occurs. Inference is also required in what is known as split antecedent anaphora. Here, the antecedent marker has to be set up as referring to the set of entities separately referred to by the syntactic antecedents of the relevant anaphor. This type of anaphoric link is illustrated below:

(20) John^{x1} said to Mary^{x2} they^y_x should go to the movies tonight.

They has both John and Mary as syntactic antecedents. However, its semantic antecedent has to be set as a new reference marker not introduced individually in the semantic representation of the discourse either by John or Mary. In DRT, this new reference marker is originated by the operation of Summation (Kamp and Reyle, 1993, p.306).

A semantic representation of this type of anaphoric link would thus be organized around the following pattern:

(21) ... $X=x_1+\dots+x_n$... $y=X$...

For the sake of generality, we may want to give the abstraction function *abst*, presented above, a definition such that all cases of inference on the side of the

semantic representation of the antecedent phrase – be it abstraction or summation, in DRT terminology – would be accounted for by it. Be that as it may, it is worth noting that the elements of the obliqueness relation may be “plural” entities comprising other entities in the relation (for split antecedent anaphora involving mass terms cf. Freitas and Lopes, 1996).

“Donkey” anaphora

A type of anaphora subject to intensive research is the one which became known as “donkey” anaphora, already briefly described above in Section 5.2. There are different alternative proposals for accounting for the semantics of this kind of anaphoric relation. Although for the purpose of semantic representation, we have been assuming the framework of DRT in the present chapter, we will adopt the rationale of the account proposed by Lappin and Francez (1994) and will propose a few adjustments thereto.

Under Lappin and Francez’s approach of “donkey” anaphora, the pronoun is assumed to be associated with a sort of bridging function. In a sentence such as (22) below, this is a function from farmer_i to i -sum_z of donkeys that are owned by the corresponding farmer. In the relevant semantic representation, one thus gets $y=f(z)$ as the semantic piece of structure corresponding to the anaphoric link.

(22) Every farmer_z who [_i owns a donkey_i]

This approach is reportedly successful in accounting for the various semantic features exhibited by “donkey” sentences, in particular in conveniently allowing for so-called universal and existential readings.

Lappin and Francez’s (1994) account is said to have been set up under the perspective that “donkey” anaphora is a specific case of e-type anaphora – following works by Lappin (1989), Heim (1990), and Neale (1990), which build on previous ideas of Evans (1980). Their account actually presents a considerable deviation from some of the essential details of e-type of anaphora. As discussed above, e-type anaphora requires an abstraction function, which places the burden, so to speak, of the mediation between the semantics of the anaphor and the semantics of its syntactic antecedent on the side of the antecedent. In the proposal of Lappin and Francez (1994), however, this burden lands instead on the side of the anaphor. As a consequence, one ends up with two quite different types of semantic representation – for e-type and for “donkey” anaphora – for what would be expected to count as two instances of a single type of anaphora.

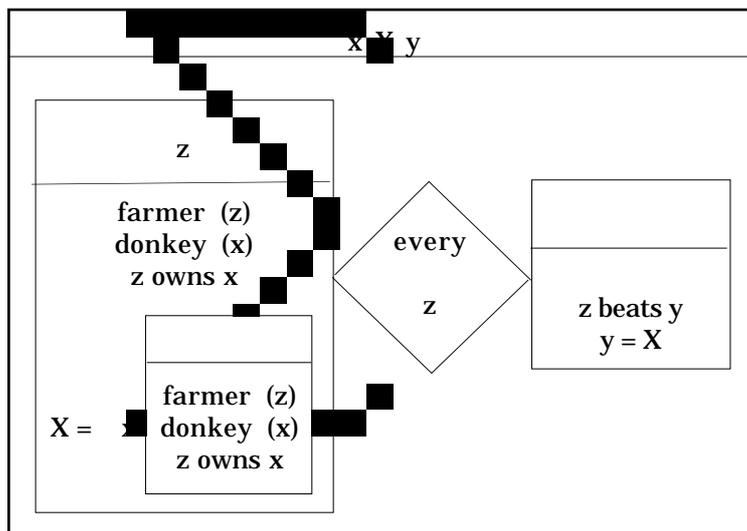
More seriously, however, the semantic value of a given “donkey” anaphor turns out not to be dependent on the semantics (or an abstraction over the semantics) of its antecedent. Surprisingly, the interpretation of the pronoun *it* in (22), inasmuch as it is assigned as semantic value, a function ranging over farmers, not donkeys, turns out to be semantically dependent on its antecedent, the NP *every farmer*, not on the NP *a donkey*.

We would like to suggest a variant to Lappin and Francez, 1994, that seeks to follow more strictly the intuition that “donkey” anaphora is an instance of e-type anaphora. In that sense, we propose that semantic mediation between the anaphor and the antecedent be made, as happens in e-type anaphora in general, via an abstraction function on the side of the antecedent (which means we are using the DRT analysis for e-type anaphora to put forward an alternative for the DRT account of “donkey” anaphora). Accordingly, we will obtain a semantic representation of “donkey” anaphora which tends to be rather organized around the following pattern:

$$(23) \quad \dots X = \text{abst}(\dots z \dots) \dots y = X \dots$$

In doing so, we are not only providing a uniform semantic representation for possibly different subtypes of e-type anaphora and removing a somewhat stipulatory function – Lappin and Francez’s function *f* from farmers to donkeys. Insofar as the abstraction function acts on the antecedent *a donkey*, we are also correctly accounting for the fact that a donkey is the antecedent of the “donkey” pronoun:

(24)



As the semantic representation of the (relative) clause in which the NP a donkey occurs contains z , a variable externally bound by every, this will induce the correct reading of the “donkey” pronoun. For every farmer z who owns a donkey, due to abstraction over the semantic representation of the relative clause in which the antecedent occurs, X will refer to the collection of donkeys owned by z , and a fortiori, given $y=X$, the pronoun can be made to refer to the donkey or donkeys owned by z . It should also be noted that the nice results of Lappin and Francez (1994) on the distinction between universal vs. existential will also be reflected in the variant of their account we are proposing.

The distinction universal vs. existential was claimed to be a side effect of the fact that the antecedent marker of the “donkey” pronoun stands for a collection of entities whose suitable part is constrained by the denotation of the verb taking the “donkey” pronoun as object. The details of this account need not be changed in the schema suggested in (23).

Furthermore, the conditional format of “donkey” anaphora – If a farmer owns a donkey, he beats it – can be accounted for without the stipulation of a bridging function f . By using abstraction function abst for the antecedents of the pronouns in conditional “donkey” anaphora, an approach inspired in the one proposed by Heim (1990) will be sufficient. A sentence like If a farmer owns a donkey, he beats it, will receive the interpretation:

- (25) For every different situation s , if a farmer owns at least one donkey in s , then s is part of a situation s' in which the man who owns at least one donkey in s beats in s' the elements (selected by the semantics of the verb in the consequent of the conditional) of the collection of donkeys that that man owns in s .

E-type and bridging anaphora

Another interesting type of anaphoric link is the one resulting from a concomitant occurrence of e-type and bridging anaphora. In these cases, the semantic processing should provide for mediating functions both on the side of the anaphor and on the side of the antecedent, giving rise to representations that conform to the following basic pattern:

- (26) ... $X=\text{abst}(\dots)$... $y=\text{brid}(X)$...

Two examples of this type of anaphoric links are shown below:

- (27) a. Mais de nove senadores^x não puderam usar o novo sistema de voto electrónico. As passwords^y_x tinham sido mudadas.
more of nine senators not could to use the new system of vote electronic. the passwords had been changed
More than nine senators could not use the new electronic system for voting. Their passwords had been changed.
- b. At least five couples^x went to the party. The men^y_x had a good time.

In the first example, the phrase *as passwords/the passwords* refers to entities associated (bridging anaphora) with the reference marker *which* which refers to the set of senators unable to use the new electronic system for voting. This reference marker resulted from abstraction (e-type anaphora) over the conjunction of the restrictor and domain of the quantification expressed by the antecedent *more than nine senators*. Also in the second example, the interplay of bridging and e-type anaphora helps to understand what otherwise would appear as a very hard pattern of anaphora.

Again, in the cases exhibiting bridging and e-type anaphora, as happened above when discussing bridging anaphora alone, there are interesting examples illuminating the context sensitiveness of the bridging function.

- (28) a. Every department has several computers^x.
The company connected them^y_x in an intranet.
- b. Every department has several computers^x.
The sales department bought them^y_x to XPTO.
The research department bought them^y_x to OTPX.

In discourse (28)b. the two occurrences of *them* have the same syntactic e-type antecedent, *several computers*. Nevertheless, the pronouns have different denotations. The first one refers to the computers bought by the sales department (not to the computers bought by all departments, as *them* happens to refer to in (28)a.), the second one refers to the computers bought by the research department. Arguably, this may be due to different outcomes of the context sensitive bridging function in each of the two cases.

Bound anaphora

It is well known, at least since Reinhart, 1983, that a certain type of anaphoric relation between anaphors and quantificational antecedents exhibit a semantic relation which differ from all the ones discussed above. This type of anaphora is illustrated below with an example from English and another from Portuguese, and it is known under the designation of bound anaphora.

- (29) a. Every sailor^x has a tattoo representing his^y_x mother.
b. Cada marinheiro^x tem uma tatuagem com o nome da mãe^y_x.
each sailor has a tattoo with the name of the mother
Each sailor has a tattoo with his mother's name.

Under this type of relation, the referential force of anaphor illustrated in the examples of previous subsections seems to be inhibited. The anaphor semantically behaves in the manner of a logical variable bound by a quantifier. In terms of semantic representation, the reference marker corresponding to the anaphor therefore seems to be equated with a reference marker introduced as a bound variable by the quantifier expressed by the antecedent, with no “autonomous” referential force:

- (30) ... Qx(... x ... y=x ...) ...

Moreover, bound anaphora exhibits additional limitations with respect to the admissible grammatical geometry between the anaphorically linked NPs. This type of anaphoric link does not comply with the relevant binding constraints associated with the anaphors at stake, as the following contrast illustrates.

- (31) Every sailor^x likes himself^y_x / #him^y_x.

Nevertheless, there are additional specific limitations apparently carried over to the anaphoric relation not by the anaphor but by the specific type of antecedent involved.

As illustrated below, the bound anaphora interpretation is possible if intra-sentential phrase boundaries are crossed – cf. (29) and (32)a. –, but it seems to be precluded in favor of an e-type reading if inter-sentential boundaries are to be crossed – cf. (32)b. (the precluding of e-type reading in (32)a. is due to the sort of interpretive deadlock concerning e-type anaphora pointed out above).

- (32) a. Every sailor^x knows that his^y_x / #their mother(s)^y_x will be always waiting for him.
- b. Every sailor^x knows many sea towns. They^y_x / #He^y_x travel(s) a lot.

Besides, even if this intra-sentential requirement is met, additional conditions should also be observed. Resorting to examples such as those below, exhibiting what is usually known in the literature as weak cross-over effect, it is possible to suggest that, for the bound anaphora relation between them to hold, anaphor and antecedent must be comparable under the obliqueness relation, i.e. the anaphor must be o-commanded by the antecedent.

- (33) a. The captain [that engaged John^x / #every sailor^x] likes him^y_x.
- b. The captain [that engaged him^y_x] likes John^x / #every sailor^x.

The double behavior of quantificational antecedents – as inducers of e-type and bound anaphora – can be easily accounted for under the quantificational approach to binding constraints we have been developing. First, reflecting the different referential status of the reference markers at stake in the two readings, we assume that each quantificational NP *W* contributes to the obliqueness relation with two reference markers. One of the markers, say *w*, results from abstraction, the other is identical to the binding marker of the relevant quantifier expressed by *W*. Notationally, they are distinguished by assigning a subscript to the latter, *w_q*.

Second, the specific extra conditioning on binding possibilities due to the quantificational antecedent under the bound anaphora reading is accommodated. Accordingly, with respect to a given anaphoric reference marker *r* and a quantificational NP *W*, we assume *w_q* is included in the obliqueness partial order of reference markers relevant for *r* iff *w_q* is comparable to *r* under the obliqueness relation. This requires just a slight reformulation of the definition of *P* and non-*P* phases, as depicted below:

- (34) a. $P'_r = P_r \setminus \left\{ x_q : \neg(x_q \parallel_{obl} r) \right\}$
- b. $\sim P'_r = \sim P_r \setminus \left\{ x_q : \neg(x_q \parallel_{obl} r) \right\}$

With respect to the symmetric exclusion of e-type anaphora when quantificational antecedents o-command the anaphors, as argued above, we believe that this should not be explicitly coded in binding constraints or auxiliary devices since, as suggested above, it simply results from an independent interpretive deadlock.

5.5 Summary

In this chapter, we discussed how to accommodate binding constraints in grammar under the obliqueness quantification format. We argued that the natural place for these constraints is at the level of Semantics, and we have shown that two suitable semantic accounts can be set up.

In one of them, as they are expressed at the level of semantic representations, binding constraints are viewed as components of the meaning of nominals.

In the other one, in so far as they are expressed at the level of the construction rules mediating between syntax and semantics, binding constraints are understood as determinants of meaning composition.

In any of the two cases, we have shown that the semantic account of binding constraints provides for neat, underspecified interface points between semantic representations and anaphora processing systems.

The preceding overview of the semantic representation of the major types of anaphora clearly suggests that the semantic approach to binding constraints ensures effective coordination between these constraints and the representation of the meaning of anaphorically linked expressions.

This overview also underlined that applying binding constraints to circumscribe the anaphoric potential of anaphors and then resolving them is not the end of the story with respect to interpreting anaphoric expressions. Once the anaphor and corresponding antecedent are identified, their reference markers have eventually to be related by a bridging function. This function is independent from binding constraints. It depends on the context, relying most likely not in grammatical principles but in world knowledge or on the interface procedures between the representation of linguistic meaning and the representation of knowledge (see e.g. Poesio et al., 1997). It is thus also likely that the mechanics and preferences for obtaining this bridging function are not completely unrelated to the mechanics and heuristics for ranking the elements of the antecedent candidates set and resolving anaphors.

On the other hand, this overview also clarified how the reference markers of antecedents happen to be set up and how they are contributed to the relevant obliqueness relation. A first major dividing line seems to exist between “quantificational” reference markers which, in contrast to other reference markers, enter the obliqueness partial order relevant for a given anaphoric reference marker iff they are comparable with it, i.e. iff they o-command the relevant anaphor. This division correlates with the distinction between bound and non-bound anaphora

As for the contribution of other types of antecedent reference markers to the obliqueness partial order, other qualifications seem to hold. Some of the reference markers result from a process of summation. Here the marker refers to a group of entities each of which were contributed by a different NP – this correlates with the class of split antecedent anaphora. Some other markers result from a process of abstraction, typically over the restrictor and domain conditions of the quantifier expressed by the antecedent – this correlates with the class of e-type anaphora.

Finally, for some other antecedents, no inference is required as its semantic contribution already involves the “direct” introduction of the corresponding antecedent marker – this correlates with anaphora to which no special term has been assigned involving referential nominals as antecedents.